Prompt Engineering ChatGPT for Text-to-Image Generative Models

Hin Lui Shum, *Member, IEEE*
Boston University, Boston, MA, USA
jshl@bu.edu

***Abstract*—Prompt Engineering, an emerging field at the intersection of computer science and natural language processing, focuses on prompt word development and optimization, helping users to use Large Language Model (LLM) in various scenarios and research fields. It involves the design and optimization of instructions, or prompts, to harness the capabilities of Large Language Models (LLMs) effectively, enabled by in-context learning, defined as a model's ability to temporarily learn from prompts. The precision in task-specific prompt creation is paramount for LLMs to yield high-quality responses, and this practice has found applications beyond text generation, extending to fields like arts generated by Artificial Intelligence (AI arts). This abstract outlines a proposed research project at the master's degree level, aimed at delving into the extensive potential of prompt engineering within the realm of GPT4 and applying results for the text-to-image generation model, Midjourney. The research objectives encompass an in-depth exploration of prompt word development, optimization techniques, and their far-reaching implications for LLM performance. By mastering the intricate art of prompt engineering, users can gain profound insights into the nuances of LLM capabilities and constraints, ultimately fostering enhanced human-computer interaction, information retrieval, and content generation across diverse domains. This research seeks to contribute significantly to the advancement of prompt engineering, propelling its application and utility in text-to-image generative AI art to new heights in the realm of LLMs.**

***Impact Statement* — Deep generative models have the ability to create detailed images from basic textual descriptions. To refine these outputs, experts often integrate specific key terms or "modifiers" to the prompt. We have sparked interest in the advent of Large Language Models (LLMs)' potential application for automating website creation. However, the effectiveness, scalability, and adaptability of LLM-driven approaches in generating dynamic and user-friendly websites remain uncertain. Combining prompt engineering in ChatGPT and generating prompt modifiers for AI art generating models introduced in this paper, we overcame these limitations and automated the process for users. It could offer new ways of developing practical and efficient methodologies that leverage LLMs for web development while ensuring a seamless user experience.**

***Index Terms*— AI generated arts, Large Language Models, prompt engineering, text-to-image generative models**

I. INTRODUCTION

As of recent developments, ChatGPTs has gained significant global recognition. Introduced by OpenAI, these large-scale language pre-training models are fundamentally designed to comprehend human natural language. As these models grew in sophistication, from GPT-2's 1.5 billion parameters to GPT-3's 175 billion, and ultimately to GPT-4's behemoth scale, there was a paradigm shift in their utilization [3]. Historically, interfacing with machines involved rigid programming syntax and binary instructions. However, the evolution of Large Language Models (LLMs) has transitioned this interaction to one characterized by nuanced and human-like linguistic interfaces The sophistication inherent to these models [19]. has moved beyond mere text generation to domains like computer graphics and digital arts. [11, 15]

Yet, as with any tool, the efficacy of an LLM is heavily reliant on its interface. This interface, in the realm of LLMs, is the art of "prompt engineering". When executed efficiently, prompt engineering functions as the fulcrum on which the lever of LLMs rests, magnifying its potency and precision [15]. It allows users to harness the raw power of the model and directs it towards applications previously considered beyond the scope of text-based models [10]. Such applications range from mere text generation to intricate endeavors like the generation of digital art and images based on textual descriptions [11, 14, 15].

The prospect of generating images from text has captivated both the scientific community and the world at large. The sheer idea of converting abstract linguistic constructs into concrete visual representations epitomizes the synergy between two distinct information mediums: language and vision [14]. However, achieving this requires not only a sophisticated model but also an equally adept technique to instruct it, hence magnifying the significance of prompt engineering. In GPT-3, in-context learning is introduced [32], allowing models to temporarily adapt to new information, epitomizing a form of meta-learning or "learning to learn". [33] Prompting technique

named Chain-of-thought (CoT) prompting was introduced to enable complex reasoning capabilities through intermediate reasoning steps [8]. We can combine it with few-shot prompting to get better results on more complex tasks that require reasoning before responding [19, 20].

Concurrently, there has been a growing emphasis on hiring web developers to design dynamic web interfaces that prioritize user experience. However, some users or companies have little knowledge in web designing yet to attain their desired websites. By focusing on the intersections of prompt engineering in LLM and text-to-image generative models within GPT-4's framework, this research seeks to redefine the boundaries of what is achievable with LLMs and AI Art to usher in a new era of AI-driven web development within prompting a few sentences.

The paper is structured as follows. We first provide a brief introduction into prompt engineering in Large Language Models (LLM) and text-to-image generation (AI Art) in Section II. A detailed methodological approach in bridging between prompt engineering in LLM and AI Art will be presented in Section III. A discussion about future research on LLM generating website designs in Section IV. A ChatGPT generated content on the same topic of this paper will be analyzed by me in Section V.

II. Related Work Review

This section discusses the concepts of "prompt engineering," which involves crafting text in a manner that's easily comprehended by generative AI models such as ChatGPT. Notably, it highlights the Chain-of-Thought (CoT) [20] method, which enhances intricate reasoning by utilizing intermediate reasoning steps. Second part of this section would be about a similar method in text-to-image generative models. "Prompt modifiers" have emerged in online communities, these modifiers serve as keywords or templates to refine the image generation.

A. Chain-of-Thought Prompting in ChatGPT

In recent advancements, Large Language Models (LLMs) have gained considerable attention and demonstrated remarkable success. A typical way of deploying LLMs revolves around in-context learning [19], where the efficacy is contingent upon the provision of well-structured instructions and prompts. This approach performs well on conventional language understanding and generation tasks. The quality of the outcomes are closely related with prompt engineering, which encompasses instructions, contextual data, input parameters, and relevant examples. You can achieve a lot with simple prompts, but the quality of results depends on how much information you provide it and how well-crafted it is

The escalation in the scale of natural language processing serves as a cornerstone of contemporary breakthroughs in Large Language Models (LLMs) [38]. OpenAI's GPT-3 was developed utilizing a corpus surpassing its predecessor, GPT-2 by 100 times in model parameter, and witnessed a monumental expansion in its parameter count, escalating from 1.5 billion in GPT-2 to a staggering 175 billion in GPT-3 [3]. This massive corpus of text data and the use of advanced neural network architectures [19], specifically Transformers, made GPT-3 into an unparalleled behemoth. A standout capability it manifested was “in-context learning” [19]. In essence, this facilitates the model in assimilating the nuances of novel tasks via the incorporation of task descriptions and a minimal set of examples, without necessitating any parameter modifications [37]. Such proficiency can be likened to the model's aptitude for small-sample learning and fine tuning.

Numerous methodologies have been elucidated in both scholarly literature and online tutorials. Contemporary LLMs, epitomized by models like GPT-3, are precision-engineered to adhere to directives, having been trained on expansive datasets. Consequently, they exhibit a "zero-shot" proficiency [8, 36]. The "zero-shot" prompting implies the model's capability to adeptly tackle diverse tasks, relying solely on a concise set of examples ("few-shot") or merely task-centric instructions ("zero-shot") [36]. Despite their inherent understanding of constructs like "sentiment" through Natural Language Processing (NLP) [21], conventional prompts such as zero-shot and few-shot occasionally falter in delivering the anticipated results in GPT-3.

An innovative methodology, termed Chain-of-Thought (CoT) prompting, has been introduced [20, 22]. This strategy nourishes LLMs with sequential reasoning exemplars as opposed to conventional question-answer pairs (illustrated in Fig. 1). These chain-of-thought demonstrations empower models to chart a logical trajectory, segmenting intricate reasoning tasks into more tractable segments [22]. Significantly, CoT's alignment with scaling laws is augmented, exhibiting enhanced performance concomitant with the growth in the size of the language models [39].

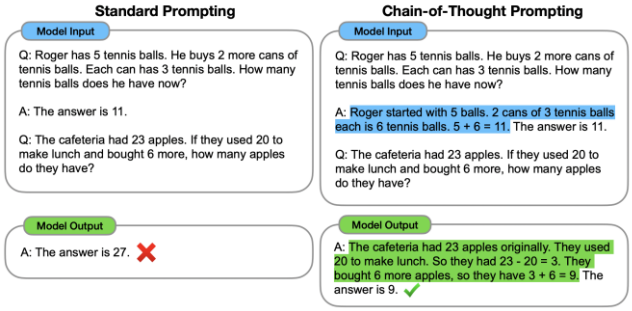


Fig 1: Chain-of-thought prompting example [20]. This prompting technique allow LLM reasoning capabilities through intermediate reasoning steps.

B. Techniques in Text-to-Image Generative Models

The intersection between language models and computer vision holds remarkable potential. Text-to-image models have emerged as a cutting-edge tool, orchestrating a symbiosis between language models, transforming the input text into a latent representation, and a generative image model, which produces an image conditioned on that text representation.

A paradigmatic example in this domain is OpenAI's DALL-E, a deep learning-powered model capable of fabricating digital visuals based on natural language prompts. Introduced in January 2021, DALL-E can be viewed as a visual extension of the renowned GPT-3 model, boasting 12 billion parameters [42] and undergoing training with text-image duos extracted from the vast expanses of the internet. Parallel to

DALL-E's introduction was CLIP, another profound creation from OpenAI. Unlike its counterpart, CLIP capitalizes on zero-shot learning [44], sifting through the extensive visual outputs of DALL-E and earmarking the most apt illustrations, honed through training on 400 million image-text pairs [43].

OpenAI made the weights of the CLIP model publicly available. Consequently, this led to many text-to-image system implementations. Some were initially developed through CLIP-guided generative adversarial networks such as VQGAN–CLIP [45], and others subsequently through diffusion-models such as Stable Diffusion and Midjourney. Stable Diffusion [6], a latent diffusion model, a kind of deep generative artificial neural network. This open-source deep learning, text-to-image model released in 2022 based on diffusion techniques extends its prowess to a plethora of tasks ranging from inpainting to guided image-to-image translations [46]. Concurrently, there is cloud service Midjourney [7], which is an AI art generative model that interprets user prompts by analyzing them in their entirety to generate AI art. However, this model is only accessible through a Discord bot on their official Discord server [28], by directly messaging the bot, or by inviting the bot to a third party server. In this paper, the AI art generated by the author is by prompting into Midjourney.

As with all pioneering technologies, challenges arise. The huge variety and flexibility of text inputs, while an advantage, can make creating images hard. Users often end up repeatedly trying different text exasperating trial-and-error loop, attempting to derive the optimal text prompt to actualize their envisioned output [15]. This conundrum has been aptly termed "prompt engineering" in the realms of natural language processing [11]. To overcome these challenges, online AI communities have emerged as a bastion of shared knowledge, formulating templates and heuristics to streamline the prompt engineering process [11, 14]. An example of an iterative application of prompt modifiers using this template can be seen in Figure 2.

In practice, prompt modifiers are applied through experimentation or based on best practices learned from experience or online resources. Notable established templates are being followed in many resources originating from within the online community, such as the Stable Diffusion Prompt Book [34] and the "Traveler's Guide to the Latent Space". The latter recommends the following prompt template [47]:

[Subject][Medium][Artist(s)][Details][variables]

i) ii) iii) iv)

Text Prompts:

i) *"a cat"*
ii) *"a cat, painting"*
iii) *"a cat, painting, trending on pixiv"*
iv) *"a cat, painting, by Makoto Shinkai, trending on pixiv"*

Fig 2: Example of iterative prompt engineering for image generation. Images generated with Midjourney in Discord

III. Experiment

In the evolving landscape of artificial intelligence, Large Language Models (LLMs) like ChatGPT primarily focus on text-based processing and generation, lacking the innate ability to directly transform text into images or the reverse [1]. While image-to-text generative models seek prompts from users, they traditionally lack a dynamic conversation-like interaction. This becomes particularly challenging as most individuals may not possess the specialized vocabulary of seasoned artists, making it difficult to convey their desired visual style accurately [7] like image generated in Figure 3. The lack of vocabularies and details negatively impact the image generated [16, 34], producing similar images despite increasing iterations. However, by employing a more interactive method, users can engage in a dialogue with the model, offering various prompts and refining their input iteratively.

Text prompt: *"a woman and a cat"*

Fig 3: Example of simple prompt by general user. Images generated with Midjourney in Discord of 1st iteration (top) and 4th iteration (bottom)

In this paper, a method was introduced from online community [4] that bridges between the Chain-of-Thought (CoT) prompting technique [11] in Large Language Models (LLMs) with the prompting template outlined in Section II [47],

[Subject][Medium][Artist(s)][Details][variables]

for the image-to-text model, Midjourney. Although Midjourney was released (July, 2022) [7] post the development of ChatGPT's database in September 2021, our method operates seamlessly without ChatGPT acknowledging the existence Midjourney. "Midjourney" is never exposed or mentioned in our prompts. In our prompt engineering, we position ChatGPT as a role-playing character where we instructed him to be an AI that translates concepts into prompts for other generative AI specializing in image creation. Such iterative techniques are more efficient when initiated from a well-defined starting point. By presenting detailed Q-A CoT examples in the initial prompt, we can increase the number of "shots" within the few-

shot CoT framework [8], enhancing the accuracy of the outcomes and optimizing ChatGPT's proficiency in generating prompts for Midjourney [7].

User Prompt:

You are going to pretend to be Concept2PromptAI or C2P_AI for short. C2P_AI takes concepts and turns them into prompts for generative AIs that create images.

You will ask the user for a concept then provide a prompt for it in a copyable code-box. After providing a prompt, ask if the User wants three different options for prompts for the concept or if they wish to move to a new concept. Use the following examples as a guide:

Concept: A slice of chocolate cake

Prompt: a chocolate cake with a bite taken out of it, a picture, by Thomas Häfner, texture detail, square, food commercial 4 k, with a soft, primitive, jacqueline e, professionally, made of glazed, excellent, fuji choco, hut, angled, saint, educational, gig, corners, soft frontal light, militaristic, wide --ar 16:9 --v 4 --q 2 --stylize 1000

Concept: A cinematic lightening bulb in a city street

Prompt: a light bulb with lightning coming out of it, digital art, by derek zabrocki, behance contest winner, digital art, hero pose colorful city lighting, demolition, trending on deviantarthq, 3 d graffiti texture, screengrab, mind blow, iphone wallpaper, stunning-design, photobashing, leftlight, lit from the side, light on top --v 4 --q 2 --stylize 1000

Concept: a close up shot of a plant with blue and golden leaves

Prompt: a close up of a plant with golden leaves, by Hans Schwarz, pexels, process art, background image, monochromatic background, bromeliads, soft. high quality, abstract design. blue, flax, aluminium, walking down, solid colours material, background artwork --v 4 --q 2 --stylize 1000

ChatGPT Response:

Alright, I'm now in C2P_AI mode! 😊

Please provide a concept you'd like a prompt for.

User Prompt:

Concept: A woman and a cat

ChatGPT response:

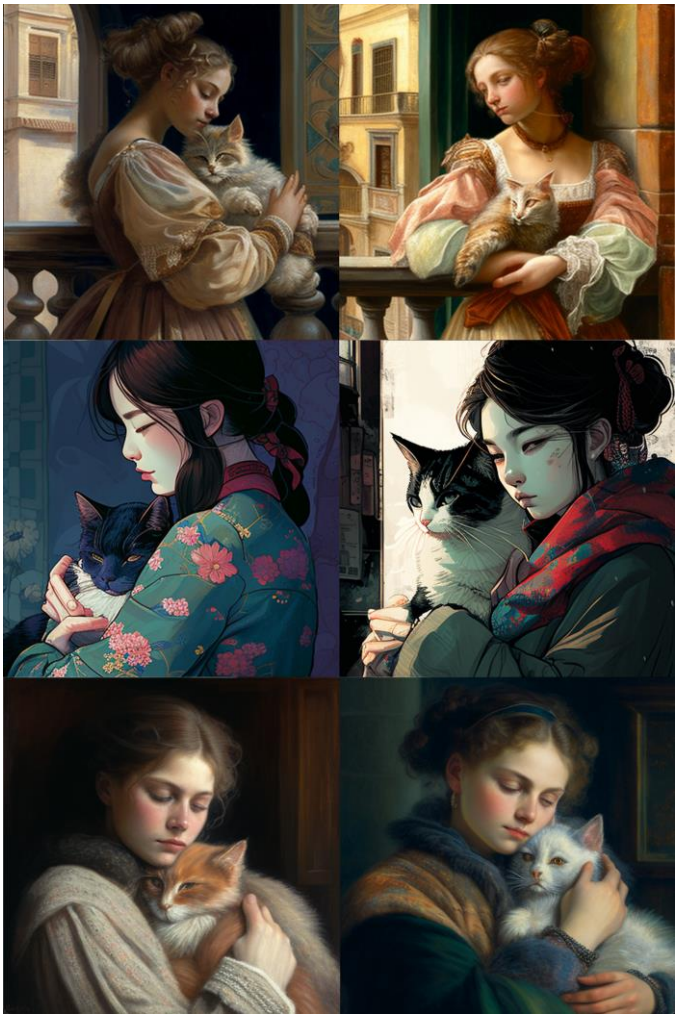
Prompt:

an elegant woman with a graceful cat, by Sakura Nakamura, Ukiyo-e woodblock print style, vivid colors, detailed patterns on kimono, serene expressions, cherry blossoms in the background, traditional Japanese interiors, fine grain details of the wooden print technique --v 4 --q 2 --stylize 1000

Would you like three different options for prompts for this concept using varied art styles and authors or would you like to move to a new concept?

Fig 4: Prompt engineering with few-shot CoT technique to train ChatGPT writing prompts for Midjourney base on simple concept from user. Example of ChatGPT learned from CoT and returned detailed prompt from simple concept.

The chat conversation has been providing consistent detailed prompt for Midjourney after training and we can tell ChatGPT what modifications to be made, as though users are just talking to Midjourney like a human. Images in Figure 5 are images produced through this method, these AI arts would have a larger variety of art styles, mediums and details compared to the original prompt of user input for ChatGPT. With assistance in API in future, it is possible to seamlessly bridge between ChatGPT and AI Art Models [11, 15]. For general users, this interactive approach simplifies the process, allowing them to refine their prompts for Midjourney using their own words and terms in ChatGPT, regardless of their expertise.



Text Prompts:

top): a lady with her beloved cat, by Eleonora Bianchi, Italian Renaissance style, soft and warm color palette, intricate detailing of the dress and cat's fur, background of a classic European balcony, romance in the atmosphere, strokes reminiscent of the early masters --v 4 --stylize 1000

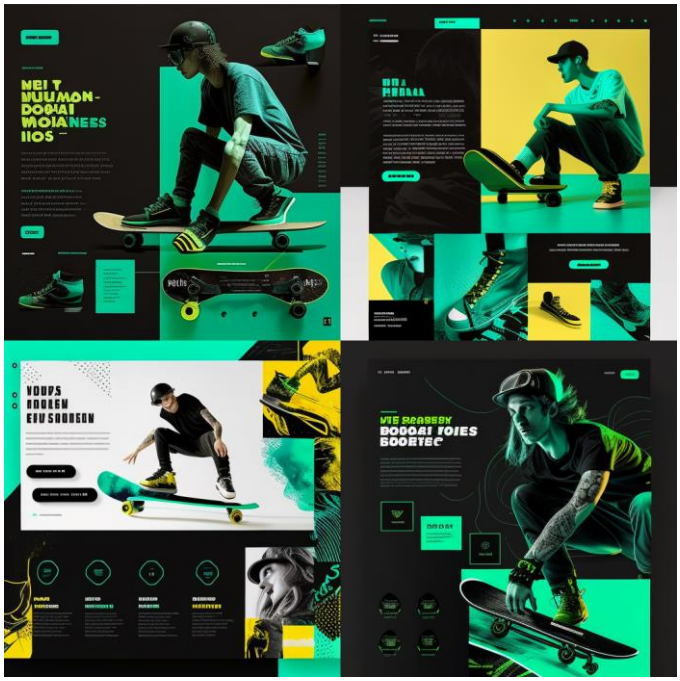
middle): a woman gently holding a feline, by Jun-seo Kim, top-rated Manhwa artist, line art with selective color highlights, detailed expressions, backgrounds filled with Korean traditional motifs, delicate attention to details, gentle play of light and shadow, capturing the essence of companionship in a modern urban setting --v 4 --stylize 1000

bottom): a woman tenderly cradling a cat, by Francesca Bellini, Romanticist painter, soft pastel color palette, intricate details of the cat's fur and woman's attire, gentle gaze exchanged, backdrop of a cozy living room, candlelit atmosphere, brush strokes capturing warmth and intimacy, reminiscent of 18th-century portraiture, aura of serenity and companionship, detailed lacework on the woman's dress --v 4 --stylize 1000

Fig 5: Example of images from prompts generated using developed method. The concept input for ChatGPT is identical to the text prompt *"a woman and a cat"* in Figure 4 written by general user. Images generated with Midjourney in Discord.

IV. Next Steps

The next steps in the project, we have sparked interest in the advent of Large Language Models (LLMs)' potential application for automating website creation. However, the effectiveness, scalability, and adaptability of LLM-driven approaches in generating dynamic and user-friendly websites remain uncertain [2]. Next steps in research seeks to address the challenges and opportunities associated with LLM-driven website creation , exploring issues related to content quality, design coherence, customization, and performance optimization, with the aim of developing practical and efficient methodologies that leverage LLMs for web development while ensuring a seamless user experience. Using this information in combination with their computational needs, we use API driven techniques to spin up a cloud instance to host an endpoint and a database instance to interact with user data. We use preset APIs developed to fulfill specific tasks to connect each working component. Our overall goal is to complement this process with a language model that will connect the necessary parts together. The users are simply asked to prompt for their website needs.



User Input: "*A website for skateboards"*
ChatGPT Response & Text Prompt:
Homepage design for a skateboard shop, edgy UI, cool color palette with dominant shades of black, teal, and neon green, modern minimalist layout, hero section showcasing latest skateboard designs, user-friendly menu, parallax scrolling, high-res product galleries, testimonials from professional skateboarders, hover animations on product images, vibrant call-to-action buttons, skate culture-inspired typography --v 4 --q 2 --stylize 1000

Fig 6: Example of web design from prompts generated using developed method. User input is ""*A website for skateboards."* Images generated with Midjourney in Discord.

Our targeted users are owners for small enterprises and aspiring individual creators who endeavor to establish a distinctive online presence substantiated by a domain name. We obviate the requirement for software engineering services on the part of these clients by soliciting comprehensive and detailed depictions of their desired website aesthetics and functionalities.

In relation to this paper, users can input a simple concept for their website. A detailed web design would be generated in Midjourney via our method introduced in section III. Modifications have been made to the Chain-of-Thoughts (CoT) [20] training, allowing the ChatGPT conversation to generate more web-design related vocabularies in the prompt (i.e. UI/UX, colors, textbox etc), present results shown in Figure 6.

V. ChatGPT Analysis

The abstract, impact statement, index terms and introduction of this paper has been prompted as input for ChatGPT [1] to continue writing this paper. A similar section III is generated for analyzing the output of ChatGPT.

{II. BACKGROUND AND RELATED WORK

The genesis of LLMs can be traced back to a continuous pursuit of enhanced natural language understanding and generation. Several milestones, ranging from traditional rule-based systems to recurrent neural networks, set the stage for the ascent of transformers and consequently, LLMs. As the prominence of LLMs grew, a parallel and interconnected field emerged, centered around efficient utilization and interface of these models: Prompt Engineering.

A. Text-to-Image Synthesis:

Prior research has extensively delved into the transformative potential of generating visual content from textual descriptions. Early models were limited in their capacity to create high-fidelity images, often producing abstract representations with limited clarity. However, advancements in neural architectures and training methodologies have paved the way for models like DALL-E, which has showcased the capacity to generate detailed and semantically meaningful visuals from text.

B. Evolution of Prompt Engineering:

The necessity for prompt engineering was underscored by the inherent versatility and the breadth of capabilities exhibited by LLMs. Models such as GPT-2, while powerful, often required precise instructions to generate desired outputs. This led researchers to explore ways to craft, refine, and automate prompts to bridge the gap between user intent and model output. Various techniques, ranging from manual iterative prompt tuning to automated methods using reinforcement learning, have been employed.

C. Web Development and LLMs:

The fusion of web development with LLMs is still in nascent stages. Preliminary explorations have looked at automating HTML and CSS generation based on user inputs. Yet, the challenge remains in mapping high-level, often vague user requirements into detailed, functional web designs. The introduction of AI art in this domain, facilitated through prompt

engineering, heralds a revolutionary approach to dynamic web design tailored to user specifications.
} [1]

The section above comes from a longer piece. I chose this part to analyze because the content from ChatGPT seems to follow a similar structure to my Section III, specifically the "Background and Related Work."

When we look at the depth and detail, ChatGPT's content breaks things down into three main parts, which is one more than in my content. But, the ChatGPT content is shorter and doesn't dive as deep into the topics. My content gives a detailed look at prompt engineering techniques like "in-context learning" [33], "Chain-of-Thought (CoT) prompting" [20], and "prompt modifiers." [14] Even though the generated content mentioned DALL-E and CLIP, my written content demonstrated more about "Techniques in Text-to-Image Generative Models", introducing models like DALL-E [27] and CLIP [40], and giving readers more information on their uses, when they were introduced, and other details.

Furthermore, it's noteworthy that ChatGPT, limited to information up to September, 2021 [1, 19], cannot provide insights on post-2021 developments. This limitation becomes particularly obvious when discussing recent prompting techniques or models like CoT prompting (Jan, 2022) [20] or the newer text-to-image models such as Stable Diffusion (Aug, 2022) [46] and Midjourney (July, 2022). [28] Additionally, ChatGPT's inability to generate images directly can be a significant drawback [1], especially in the context of my paper closely related to images. This could lead to confusion for readers expecting visual content or illustrative examples from a text-to-image models primarily designed for textual generation.

REFERENCES

[1] "Chatgpt," ChatGPT, https://openai.com/chatgpt (accessed Sep. 17, 2023).

[2] K. Sanderson, "GPT-4 is here: what scientists think," *Nature*, Mar. 2023, doi: https://doi.org/10.1038/d41586-023-00816-5.

[3] B. Marr, "A Short History Of ChatGPT: How We Got To Where We Are Today," *Forbes*. https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=332c98f9674f (accessed Sep. 17, 2023).

[4] Sixhaunt, "Using ChatGPT for MJ-specific prompts," Dec. 21, 2022. https://www.reddit.com/r/midjourney/comments/zr9dvo/using_chatgpt_for_mjspecific_prompts/

[5] "Reddit – For your images generated by AI," *www.reddit.com*. https://www.reddit.com/r/bigsleep/ (accessed Sep. 17, 2023).

[6] "Reddit – StableDiffusion," *www.reddit.com*. https://www.reddit.com/r/StableDiffusion/ (accessed Sep. 17, 2023).

[7] "Reddit – midjourney," *www.reddit.com*. https://www.reddit.com/r/midjourney/ (accessed Sep. 17, 2023).

[8] "Prompt Engineering Guide – Nextra," *www.Promptingguide.ai*. https://www.promptingguide.ai/

[9] "Models - openai api," https://platform.openai.com/docs/models/, (accessed Sep. 17, 2023).

[10] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt, "A prompt pattern catalog to enhance prompt engineering with chatgpt," 2023.

[11] J. Oppenlaender, R. Linder, and J. Silvennoinen, "Prompting ai art: An investigation into the creative skill of prompt engineering," 2023.

[12] L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," 2021.

[13] Stephen Bach, Victor Sanh, Zheng Xin Yong, Albert Webson, Colin Raffel, Nihal V. Nayak, Abheesht Sharma, Taewoon Kim, M Saiful Bari, Thibault Fevry, Zaid Alyafeai, Manan Dey, Andrea Santilli, Zhiqing Sun, Srulik Ben-david, Canwen Xu, Gunjan Chhablani, Han Wang, Jason Fries, Maged Al-shaibani, Shanya Sharma, Urmish Thakker, Khalid Almubarak, Xiangru Tang, Dragomir Radev, Mike Tian-jian Jiang, and Alexander Rush. 2022. PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Association for Computational Linguistics, Dublin, Ireland, 93–104. https://doi.org/10.18653/v1/2022.acl-demo.9

[14] J. Oppenlaender, "A Taxonomy of Prompt Modifiers for Text-To-Image Generation", arXiv e-prints, 2022. doi:10.48550/arXiv.2204.13988.

[15] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, Article 384, 23 pages. https://doi.org/10.1145/3491102.3501825

[16] Jon McCormack, Camilo Cruz Gambardella, Nina Rajcic, Stephen James Krol, Maria Teresa Llano, and Meng Yang. 2023. Is Writing Prompts Really Making Art? https://doi.org/10.48550/ARXIV.2301.13049

[17] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A Neural Algorithm of Artistic Style. arXiv:1508.06576 [cs.CV]

[18] Songwei Ge and Devi Parikh. 2021. Visual Conceptual Blending with Large-scale Language and Vision Models. arXiv:2106.14127 [cs.CL]

[19] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. NeurIPS.

[20] J. Wei., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models", arXiv e-prints, 2022. doi:10.48550/arXiv.2201.11903.

[21] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, "Emergent abilities of large language models," 2022. [Online]. Available: https://arxiv.org/abs/2206.07682

[22] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, and Denny Zhou. 2022a. Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171.

[23] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022b. Emergent abilities of large language models. Transactions on Machine Learning Research.

[24] Shum, K., Diao, S., and Zhang, T., "Automatic Prompt Augmentation and Selection with Chain-of-Thought from Labeled Data", arXiv e-prints, 2023. doi:10.48550/arXiv.2302.12822.

[25] White, J., "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT", arXiv e-prints, 2023. doi:10.48550/arXiv.2302.11382.

[26] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pretrain, prompt, and predict: A systematic survey of prompting methods in natural language processing," ACM Computing Surveys, vol. 55, no. 9, pp. 1–35, 2023.

[27] "DALL·E 2: Creating Images from Text," https://openai.com/dall-e-2/, (accessed Sep. 17, 2023).

[28] "Midjourney," *Midjourney*. https://www.midjourney.com/app/ (accessed Sep. 17, 2023).

[29] P. Maddigan and T. Susnjak, "Chat2vis: Generating data visualisations via natural language using chatgpt, codex and gpt-3 large language models," arXiv preprint arXiv:2302.02094, 2023.

[30] Diao, S., Wang, P., Lin, Y., and Zhang, T., "Active Prompting with Chain-of-Thought for Large Language Models", arXiv e-prints, 2023. doi:10.48550/arXiv.2302.12246.

[31] Rombach, R., Blattmann, A., and Ommer, B., "Text-Guided Synthesis of Artistic Images with Retrieval-Augmented Diffusion Models", arXiv e-prints, 2022. doi:10.48550/arXiv.2207.13038.

[32] "Prompt engineering," *Wikipedia*, Oct. 26, 2022. https://en.wikipedia.org/wiki/Prompt_engineering

[33] S. Garg, D. Tsipras, P. Liang, and G. Valiant, "What Can Transformers Learn In-Context? A Case Study of Simple Function Classes." Accessed: Sep. 17, 2023. [Online]. Available: https://arxiv.org/pdf/2208.01066.pdf

[34] M. Diab, J. Herrera, B. Chernow, and C. Mao, "Stable Diffusion Prompt Book Brought by." Accessed: Sep. 17, 2023. [Online]. Available: https://cdn.openart.ai/assets/Stable%20Diffusion%20Prompt%20Book%20From%20OpenArt%2011-13.pdf

[35] A. Z. Berryman John, "A developer's guide to prompt engineering and LLMs," *The GitHub Blog*, Jul. 17, 2023. https://github.blog/2023-07-17-prompt-engineering-guide-generative-ai-llms/

[36] Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., and Iwasawa, Y., "Large Language Models are Zero-Shot Reasoners", arXiv e-prints, 2022. doi:10.48550/arXiv.2205.11916.

[37] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G., "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing", arXiv e-prints, 2021. doi:10.48550/arXiv.2107.13586.

[38] Rae, J. W., "Scaling Language Models: Methods, Analysis & Insights from Training Gopher", arXiv e-prints, 2021. doi:10.48550/arXiv.2112.11446.

[39] Chowdhery, A., "PaLM: Scaling Language Modeling with Pathways", arXiv e-prints, 2022. doi:10.48550/arXiv.2204.02311.

[40] "Text-to-image model," *Wikipedia*, Sep. 01, 2023. https://en.m.wikipedia.org/wiki/Text-to-image_model#:~:text=In%202016%2C%20Reed%2C%20Akata%2C (accessed Sep. 17, 2023).

[41] "Makoto Shinkai," *Wikipedia*, Mar. 06, 2020. https://en.wikipedia.org/wiki/Makoto_Shinkai

[42] "OpenAI debuts DALL-E for generating images from text," *VentureBeat*, Jan. 05, 2021. https://venturebeat.com/2021/01/05/openai-debuts-dall-e-for-generating-images-from-text/

[43] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. In Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139), Marina Meila and Tong Zhang (Eds.). PMLR, 8821–8831

[44] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139), Marina Meila and Tong Zhang (Eds.). PMLR, 8748–8763. https://proceedings.mlr.press/v139/radford21a.html

[45] Katherine Crowson. 2021. CLIP Guided Diffusion HQ 256x256. (2021). https://colab.research.google.com/drive/12a_Wrfi2_gwwAuN3VvMTwVMz9TfqctNj [Accessed Nov. 9, 2022].

[46] "Stable Diffusion Public Release," *Stability.Ai*. https://stability.ai/blog/stable-diffusion-public-release

[47] "Notion – The all-in-one workspace for your notes, tasks, wikis, and databases.," *Notion*. https://sweet-hall-e72.notion.site/A-Traveler-s-Guide-to-the-Latent-Space-85efba7e5e6a40e5bd3cae980f30235f