Research Article

A Brief Summary of Prompting in Using GPT Models

Shuyue Wang[1], Pan Jin[2]

1. Northeast Normal University, China; 2. Independent researcher

This summary introduces the importance of prompting in the rise of GPT model applications. Firstly, the paper describes the status quo of GPT model's (mostly ChatGPT's) use in many domains that is relevant to prompting. Then, the improvement approaches that occur in concurrent studies are summarized. Finally, a methodological inference is accomplished, with the authors' expectation over the future situation of GPT models' usage. The paper is dedicated to providing a useful guide to those who is working on accelerating the merging of GPT model and human jobs for the time being.

Shuyue Wang and Pan Jin contributed equally to this work.

Corresponding author: Shuyue Wang, henri_w_91@hotmail.com

1. The rise of ChatGPT and the issue of prompts

ChatGPT is a chatbot released by OpenAI in November 2022, and it represents a significant improvement over its predecessor GPT-3[1]. It has overwhelmingly pushed forward the frontier of artificial intelligence with its quick, articulate answers across many domains of knowledge with remarkable precision and coherence. [2][3][4][5]. Users have taken its benefits with great relish for super high workflow automation [6][7][1]. ChatGPT is capable of generating text in a wide range of styles and for different purposes. It is designed with a strong focus on interactive conversations. Based on Large Language Model (LLM), it has been fine-tuned (an approach to transfer learning) using both supervised and reinforcement learning technique [8]. Hereafter, we refer to the mind-blowing ChatGPT and its likes as GPT models in this paper, unless in particular experiment description.

One major impact brought by GPT is its general-purpose characteristics [9][10]. Since LLMs become increasingly integrated into specialized applications in areas such as writing assistance, coding, and legal research, it is certain that those works have paved the way for businesses and individuals to adopt GPT's more widely [11]. GPT model's commercial use has definitely changed the look of human lives in terms of economy [8], cultures [12][13] and even policy-makings [14][15]. Eloundou et al. studied its potential impact to labor market in less than a week after GPT4.0 is released [8], which estimated that approximately 80% of the U.S. workforce could thus have at least 10% of their work tasks affected, while around 19% of workers may see at least 50% of their tasks impacted. Most of the studies (e.g. medical AI assisting with doctors' decision-making process during prostate cancer diagnosis [16], assistants compensating for human weaknesses [17], etc.) use taskspecific models, and therefore limit observations to human interaction with AI that primarily serves one function, or in one domain (e.g., writing, medicine, music, etc.); however, GPT seems to know (a little) everything.

Other than the interior mechanism of OpenAI's GPT that has attracted many studies on the generative capabilities of LLMs (there are some open-source replica of it), the information path from users to the GPT needs attention as well. For example, in the case of labor market study mentioned above where the jobs that are statistically deemed to be replaced by ChatGPT are mostly repeated and quantifiable works, the model has strengthened the importance of communication towards the GPT model from users. This is where **prompts** appear significant in GPT model [18][19][20]. This resembles to the blockchain world where the role of Oracle played as a third-party service that connects smart contracts with the outside world, primarily feeding information in from the world and also the reverse: therefore, it seems that there still needs someone to 'transfer the idea', surely via prompts.

A prompt is also referred to as 'instruction', as it is a natural language descriptions of human tasks [21] [22]. For those multi-modal GPT model, a prompt can take different forms than text only (e.g., carrying out specific tasks of generating images from captions or transcribing text from speech, etc. in generative AI modular studies). In fact, LLMs are not always as easy as to interact and collaborate with: they can be opaque and hard to debug most of the time. Also, real-world tasks can be quite complex, sometimes presenting challenges for current LLMs to solve from a single model; they may fail to capture the subtleties of many tasks that involve multiple objectives simultaneously [23], while a seamless integration of individual components for better utility, performance, and generalization has always been the goal of GPT model's applications [24]. Since LLMs can take in any natural language prompts, end users are likely to struggle to determine how to change their prompts to remedy unexpected outputs. Many tests are done to GPT models but not many are focused on the pattern that human 'speaks' (i.e. inputs) to the machine. Nonetheless, some researchers have improved the reliability of these models using different methods with advancements, so that the models' ability is enhanced to discern user intent, rendering them more user-friendly and practical [25][26]. In the rest of the paper, the authors are focused on the usage of prompt and their tactics among current studies. A technical analysis of how GPT models deals with prompts is summarized from many scholars as well. Finally, a perspective is provided by the authors that one trend of GPT model is to realize the real automation of Q&A [2], where some paths are discussed in fair details. [3]

2. Variety of prompts: as if in a conversation

Social medias across the globe have seen the carnival of inputting random or entertaining queries of prompts into GPT models, all of which comes down to the problem of proposing effective prompts. Prompts are by definition how you get GPT model to do what you want, but with plain texts (mostly English[4]).

Prompts are being discussed. In the work of societal megatrend detection using ChatGPT, Haluza et al. interacted with the AI to learn whether detailed model selection and parameters such as randomness, maximum response length, diversity, wording frequency, and text presence penalties can be applied for prompts [5]. The status quo of prompting in GPT models is also reflected in the well-known AI code assistant Copilot, where for a dataset of 166 programming problems, 60% of the remaining only natural language changes to the problem description [27]. In a study, benchmarks are prepared for a complete test [28]. The prompt there is quite simple: copy the problem description and paste into a box waiting for suggestion generation via user interface.

When applying chatGPT to medical consultation, Zuccon et al. found that that the knowledge passed in the prompt can overturn the knowledge encoded in the model to the detriment of answer correctness [29]. ChatGPT answered correctly 80% of the questions if relying solely on model knowledge. On the other hand, the evidence presented in the prompt can heavily influence the answer: it affected the correctness of the answer, reducing ChatGPT's accuracy in their task to only 63%. A wrong prompt to guide chatGPT may bring out worse result than it should.

When it comes to medical clinical diagnosis, Lyu et al. tested that ChatGPT can successfully translate radiology reports into plain language with an average score of 4.27 in the five-point system with 0.08 places of information missing and 0.07 places of misinformation [30]. In terms of the suggestions provided by ChatGPT, they were general relevant such as keeping following-up with doctors and

doi.org/10.32388/IMZI2Q

closely monitoring any symptoms. About 37% of 138 cases in total ChatGPT offered specific suggestions. ChatGPT also presented some randomness in its responses with occasionally over-simplified or neglected information, which can be mitigated using a more detailed prompt.

Baidoo-Anu et al. used ChatGPT in promoting teaching and learning, claiming that benefits of ChatGPT include but are not limited to promotion of personalized and interactive learning, generating prompts for formative assessment activities that provide ongoing feedback to inform teaching and learning etc [31]. For example, teachers can leverage ChatGPT to create prompts for open-ended questions that align with the learning goals and success criteria of the unit of instruction. Additionally, ChatGPT can be used to also generate quality rubrics that clearly and concisely explain exactly what students need to accomplish to be successful in the various required levels of proficiency.

As for translation, Jiao et al. provided a preliminary evaluation of ChatGPT for machine translation, including translation prompt, multilingual translation, and translation robustness [32]. By evaluating benchmark test sets, they found that ChatGPT performs competitively compared with commercial translation products (e.g., Google Translate) on high-resource European languages, but lagging behind significantly on low-resource. For distant languages, they explored an interesting prompt strategy 'pivot prompting' asking ChatGPT to translate the source sentence into a high-resource pivot language before into the target language, which improved the translation performance significantly. In the field of Japanese-Chinese translation linguistics, the issue of correctly translating attributive clauses has persistently proven to be challenging. Gu et al. established a prompt strategy to use ChatGPTfor translation: this prompt strategy is capable of optimizing translation input in zero-shot scenarios and has been demonstrated to improve the average translation accuracy score by over 35% [33].

Even in agriculture, Biswas et al. tested some prompts and questions to ChatGPT, where ChatGPT can be used for crop forecasting, soil analysis, crop disease and pest identification, and precision farming [34]. In supply chain management, logistics, and inventory management, ChatGPT can help optimize logistics, improve inventory management and increase supply chain efficiency. ChatGPT can be used to generate accurate and timely reports, alerts and insights which helps make informed decisions and improve customer service.

Education is a natural arena for ChatGPT, In Shue et al. research the impressive conversational and programming abilities of ChatGPT make it an attractive tool for facilitating the education of bioinformatics data analysis for beginners [35]. In another work, a bar exam is tested upon GPT-4 in comparison with ChatGPT for evaluation [6]. Katz et al. selected notoriously-challenging tests with legal languages as their prompts to the model. Instead of simply dumping the textual materials to the model, they implemented and described frameworks for multiple-choice assessment on the Bar Exam and an open-ended assessment for task-based simulation in the CPA Exam as their prior work [36]. It is quite surprising that a standardized test can leave some space for the ’prompt art’.

Another significant application of ChatGPT is in robotics. Vemprala et al. discussed design principles for creating such APIs and prompting strategies that can be used to generate code for robotics applications via ChatGPT [37]. The proposed framework allowed the generated code to be tested, verified, and validated by a user on the loop via a range of methods including simulation and manual inspection. They demonstrated how the prompts of ChatGPT can be used for multiple applications ranging from simple common-sense robotics knowledge tasks all the way to deployments in aerial robotics, manipulation and visual navigation.

In the research of Guo et al., ChatGPT can boost artistic creation [38]. They showed that ChatGPT can provide clearer and more detailed guidance on painting content, organize painting elements more reasonably, and form a clear and reasonable control of painting content. In addition, ChatGPT can understand abstract art expressions such as painting style and emotion, and connect these abstract concepts with specific painting techniques such as brushstrokes and colors through text descriptions. However, it is difficult to achieve adequate expression of emotions with existing AI art creation methods.

In design and manufacturing area, ChatGPT can play a good role in knowledge management. A case study of Hu et al. showcased the vast information and contextual awareness that ChatGPT can offer, highlighting the crucial role of suitable prompts and verification in ensuring the accuracy and relevance of generated information [39]. ChatGPT provided a novel opportunity for design knowledge acquisition through its ability to offer an integrated platform for knowledge retrieval. In such a single and centralized platform, designers can acquire sufficient knowledge that pertains to common sense knowledge, various domain-specific knowledge, as well as engineering and technique knowledge, which support design decisions throughout the design process. Consequently, it has potentials for designers to alter knowledge providers less when defining and solving problems derived from diverse stages of the design process. The reduction in the need for designers to frequently switch between different knowledge provides help streamline the design process and improve workflow efficiency. In addition, such an integrated knowledge tool can facilitate better collaboration and communication among designers by allowing them to work on the same platform and share knowledge easily. In construction projects management, Prieto et al. presented a study in which ChatGPT was manipulated via prompts to generate a construction schedule for a simple construction project [40]. The output of ChatGPT was evaluated by a panel of participants providing feedback on their overall interaction experience and output quality. Results showed that ChatGPT can generate a coherent timetable that follows a logical approach to satisfy the specified range. Participants had an overall positive interactive experience, noting the tool's potential to automate many preliminary and time-consuming tasks.It shows that ChatGPT has the potential to revolutionize the construction industry by automating repetitive and time-consuming tasks. The research by Rathore et al. on ChatGPT on textile industry helped companies to optimize the production process, provide automated customer support and generate personalized recommendations for shoppers, without any additional cost [41]. For example, ChatGPT can be trained on data from production line to detect any anomalies or changes, giving timely alerts to the personnel. It can also be used to provide personalized support and advice to shoppers and generate meaningful recommendations according to customer's preferences. Hence, with the help of this technology, companies in the textile industry can improve the customer experience and make their services more efficient, cost-effective and prompt.

When it comes to financial area, Yue et al. discussed the potential of using ChatGPT to revolutionize the transfer of financial literacy to non-financial professionals [42]. They tested ChatGPT's ability to explain complex financial concepts in a non-technical way. The results showed that ChatGPT has great potential as a tool to educate a broad target audience on complex financial concepts. They also identified limitations of ChatGPT in interpreting model predictions, gaining expected improvements in the model's immediate engineering to overcome these limitations. Ultimately, the use of ChatGPT has the potentials to enable all individuals, regardless of their financial background, to make informed investment decisions. As for marketing, Rivas et al. believed that ChatGPT can help create content faster and potentially as well as human content creators [43]. It can also help conduct more efficient research and understand consumer vocabulary, perceptions, and attitudes toward products and campaigns. ChatGPT can also help personalize emails and recommendations, provide 24/7 automated customer service, and improve call center customer service efficiency and accuracy. AI marketing tools based on ChatGPT can pose several potential risks for marketers, consumers, and other stakeholders, such as inaccuracies in the timeliness of its training data, dependency, job replacement, and societal bias that may harm transparency, bias mitigation, privacy protection, risk assessment, accountability, continuous monitoring, and ethical decision-making.

3. Improvement of prompt tactics: manners of prompting

In the world of prompts, you are not allowed to let yourself in the case of ”what to do when words don ´t work’[44]: you should try to make it work, thus the improvement of prompt tactics. Prompts are thus critical for the users because there are factual and opinionated biases produced by GPT models due to statistical lacking and unintentional institutional interests embedded in design [12][45]. A well-crafted prompt helps the model generate more accurate and relevant outputs, while a poorly crafted prompt leads to incoherent or irrelevant outputs. The art of writing useful prompts is called prompt engineering. Professional use of prompts brings better interaction thus with better result. For example, professional writers are benefited when the new text generation and editing system can be applied to gauge their responses [46]. With right prompts, it is possible to create entirely new interaction paradigms, such as having an LLM generate and give a quiz associated with a software engineering concept or tool, or even simulate a Linux terminal window [23].

A prompt can be any text: there are non-hardcoded rules that is usually followed as in the case of coding. There are some guidelines for structuring your prompt text that can be helpful in getting the best results. A **zero-shot** prompt is the simplest type of prompt, providing a beginning description of a task, waiting for GPT model to start with, e.g. a question, the start of a story, email message beginning [47]. In this case, the clearer your prompt text is, the easier it will be for GPT model to understand what should come next. The effect of engineered prompts is compared with original prompt in terms of copilot code generation failure: a total of 87 problems remained unsolved after the initial code generated by Copilot [27]. However, modifying the description for these problems led to Copilot generating a successful solution in 53 cases (60.9%). It is noted that spelling, unclear text, and the number of examples provided have an effect on the quality of the completion. Prompt size is also important: the prompt and the resulting completion must add up to fewer than 2,048 tokens in the realm of GPT models (roughly 1,500 words in the work of [23]).

Due to the fact that GPT model is trying to figure out which text should come next, including instructions and examples provides context that helps figure out the best possible completion. In practice, examples of desired outcomes in the model‘s input text prompt allow it to ’learn’ rapidly and

doi.org/10.32388/IMZI2Q

generate the required outcome on new material. That is to say, the prompt itself can be the learning target of GPT models. To demonstrate its capabilities as a keyphrase generator, Song et al. conducted a preliminary evaluation of ChatGPT in various aspects, including keyphrase generation prompts, keyphrase generation diversity, multi-domain keyphrase generation, and long document understanding [48]. Their evaluation was based on six benchmark datasets; they adopted the prompt suggested by OpenAI while extending it to six candidate prompts. ChatGPT performed exceptionally well on all six candidate prompts with only minor performance differences observed across the datasets.

Zhai et al. developed a prompt strategy of ChatGPT to solve the problems difficult to deal with in traditional ways: how to track students' learning, how to provide feedback and learning guidance, how to recommend learning materials, and how to meet the special needs of students with diverse backgrounds [49]. They used one performance expectation of the K-12 Next Generation Science Standards to develop a prompt, using which ChatGPT automatically generates a performancebased assessment task. They supplied a response and asked ChatGPT to grade and provide feedback, and then asked ChatGPT to provide learning guidance and learning materials based on the response. Lastly, they told ChatGPT that the learner was with dyslexia, and eventually ChatGPT recommended specific learning materials for the learner. Their results suggested that ChatGPT has the potential to tackle the most challenging problems of science learning through automatic assessment development, automatic grading, automatic learning guidance, and automatic recommendation of learning materials.

It has been shown that the performance of LLMs can be enhanced through incontext learning by providing few labelled examples (prompts) in addition to the test input in machine learning in machine translation [50]. One example of prompt improvement is worth illustrating: task decomposing. In AI chains that chains LLM, chaining enables users to run the same model on multiple sub-tasks, thereby granting each sub-task a higher likelihood of success (as opposed to solving the entire task in one go): the final composed paragraph turns out more comprehensive in addressing all problems, and has a more constructive tone [23]. This type of 'prompt-based learning' is a strategy that machine learning engineers can use to train large language models (LLMs) so the same model can be used for different tasks without re-training. Prompt-based learning models can autonomously tune themselves for different tasks by transferring domain knowledge introduced through prompts. The quality of the output generated by a prompt-based model is highly dependent on the quality of the prompt. Prompt-based learning makes it more convenient for engineers to use foundation models for different types of downstream uses. As opposed to “naive” prompting that requires an input to be directly followed by the output/answer, elicitive prompts encourage LLM to solve tasks by following intermediate steps before predicting the output/answer [51][52]. Some work automatically mine more effective prompts due to the sub-optimal manual prompting: however, the mined prompts tend to be less human-readable [53]. As is found in a study of Coplilot, the not-so-perfect result of prompt engineering is a potentially useful learning activity that promotes computational thinking skills [27]. This type of tactics of prompt selection has also seen its power in other works [54][55].

There is also an idea of crowdsourcing that is applied so that the prompters are able to break down complex tasks into pieces that can be performed independently [16]. The improvement of prompts is far from being limited to narrative strategy sort of thing: many assembling user interfaces are proposed so that combined experience can better utilized to application, e.g. PromptMaker in prototyping ML functionality, which is a 137-billion parameter generative language model that behaves in a similar way to GPT-3 in its ability to follow prompts [56][57].

4. Uncover the box: mind-to-mind talking

One of the key aspects of intelligence is interactivity with the chatbot from external world, which requires to verify the capability of communicating and responding to feedback via human language input, event impulse, environment change. In Microsoft’s report, the external input has been divided into two classes of ’interaction with the world’ as well as ’interaction with humans’ for research (the hidden subjective is the robot itself): the former can be exemplified by manager’s job in an enterprise virtual or in practice, calendar and email sorting, information collecting via web-browsing, etc., where tests show that too specific adaptation to prompt, such as specialized training or fine-tuning, is not necessary any more [10]. However, there are some limitations to be noted: an access to external resources are not given for granted. Sometimes, the GPT model is not able to reason about when it should use external information (e.g. most of the time, GPT-4 relies on external search engine results rather than its own reasoning ability)[10].

Giving out an explanation for the nature and mechanism of GPT models is a sudden and urgent challenge, even a succint one. The new technique also comes with new issues that has not been seen in smaller scale of models, i.e. emergence [58][59][60]. It is not that easy due to the closeness of the GPT

 doi.org/10.32388/IMZI2Q

model releaser OpenAI. One hypothesis believes that the large data with content diversity forces neural networks to learn generic ’neural circuits’, where large model size produces redundancy for the trained model to specialize and fine-tune in specific tasks [61][62]. Back to the case of prompts, taking the advantage of transfer learning, GPT models are not solely relying on pre-training. This enables it to generate responses based on the context of the conversation, even though it does not have access to external information and online data from the internet. This unusual way of retrieving information nowadays with digital devices NOT using popular search engines is the fundamental difference brought to the prompting approach. In fact, human is a different ’terminal’ of feedback for GPT models since human is the ultimate information receiver of all tasks whereas others simply intermediates.

There are many aspects that have been studied amongst researchers in terms of why and how of such a remarkable intelligence, especially the reacting mechanism of the ’mental states of characters’[10]. For example, creative reasoning, technical proficiency as well as critical reasoning are concerned. The approaches of measuring the (general) intelligence are all somewhat subjective and informal: this forms many of the elaboration on this topic. For a human who wants to make the GPT model serve their job well, the model has to understand human in theory of mind (includes the basic task of reflecting on someone else’s mental states, and the more advanced task of reflecting on someone’s reflection of someone else’s mental state, etc. [10], which is the ability to attribute mental states such as beliefs, emotions, desires, intentions, and knowledge to oneself and others, and to understand how they affect behavior and communication [63].

It is widely accepted from users’ experience that one more line of prompt is usually more than enough for a more satisfying response from GPT model. In the realm of prompting, an ’emphasis’ seems inevitable in bringing out good solution. This is probably a feature that is characteristic in GPT model, which suggests that one way to improve the explainability and debuggability of an otherwise opaque, black-box LLM is to have it do less: breaking a problem up into smaller problems, having the model solve each (smaller) problem separately, showing the intermediate results, and allowing users to edit those results [23]. People have to discover prompts that elicit e.g. step-by-step reasoning, manually providing examples when it comes to few-shot for a new task [24]. This is consistent with the necessity of ’recursive’ prompt with ’emphasis’ that is favourable for the bot to feedback with better responses.

doi.org/10.32388/IMZI2Q

There are even perspectives where LLM itself is considered as a large prompt engineer [64]. Prompts have the potential for self-adaptation, suggesting other prompts to gather additional information or generate related artifacts. These advanced capabilities of prompts highlight the importance of engineering them to provide value beyond simple text or code generation. In this spirit, prompt generation is discussed, including both differentiable tuning of soft prompts [65] and natural language prompt engineering [21][66].

As is mentioned above[5], a single large language model can support humans in a variety of sub-tasks. This is probably because of some disadvantages LLM is facing with: 1. multi-step reasoning: LLMs are designed to grasp the form of language, rather than the meaning [67]. 2. exposure bias: Since LLMs generate text sequentially in an autoregressive manner, errors or imperfections from previous runs can accumulate, LLMs thus being less likely to perform well when generating long bodies of text [68]. 3. textual sensitivity: prompts that are unnatural relative to the typical text distribution tend to be less efficient [69], while nouns and verbs are more important than adjectives and function word [70]. For example, White et al. has concluded 16 prompt patterns utilizable in GPT models, which may reflect the knowledge infrastructure of this realm [71]—

- Category of Input Semantics: Meta Language Creation.
- Category of output customization: Output Automater, Persona, Visualization Generator, Recipe, Template.
- Category of Error Identification: Fact Check List, reflection.
- Category of Prompt Improvement: Question Refinement, Alternative Approaches, Cognitive Verifier, Refusal Breaker.
- Category of Interaction: Flipped Interaction, Game Play, Infinite Generation.
- Category of Context Control: Context Manager.

In their work on improving code quality and software design, White et al. shows that the correct hinting pattern can be used to help overcome these errors and reduce errors. By default, the code generated by chat GPT is difficult to modularize, making it difficult to maintain. By continuously adding and optimizing prompts, they make the generated code more modular. By continuing adding functions to the software, the software design becomes better.

doi.org/10.32388/IMZI2Q

5. Conclusion: the floor is soon GPT's

The GPT series have brought a technological revolution with 'contextual' perspectives, which in fact explicitly the essence of the prompt problem discussed in this paper [12]. People are now confined to a conversation box when they started opening up a link to ChatGPT or ChatGPT plus: the conversation is already beginning, not later than the first release. As is believed by the authors, the only obstacle that seemingly hinders the conversation box (that you see on GPT model's webpages) to vanish is simply the lack of 'smart' input: if this barrier is lifted, a flow of input 'prompts' can continuously injected into the model. Thus, a literal 'automatic' GPT model, i.e. the perparatory work of 24/7 input information, is rather realizable [72][73]. The authors believe that a relatively deep understand of a prompt's lifecyle in the Natural Language Processing (NLP) functioning in LLM is necessary for understand as well as improving the prompt manipulation.

A static viewpoint never stands in front of this in-progress thing [13]. In this case, the discussion of prompt is no more than a quite contemporary summary. The prompt is just as temporary as the current level of GPT models' responses since the prompt-feedback is a process in evolution [74]. One thing is certain, the longevity of the job 'prompt engineer' in areas of AIGC image generation is surely impossible. A transitional semi-automatic prompt robot is certainly within reach.

For example, AutoGPT is released recently as one such kind of assistant of GPT model usage [75]: as matter of fact, this paper itself can be one day an easy job for such GPT assistant.

6. Acknowledgement

With this paper, Shuyue Wang encourages himself with Han Jiang working on things that don't scale. Pan Jin would like to express deep appreciation to Ms Yuxuan Ren for her companionship and support in the past five years. The authors both share this work joyfully with their friends in SAIC AI Lab.

Footnotes

[1]Assistance from ChatGPT is so huge that even co-authorships with the model via a formal agreement are sought by some studies that is enhanced via using ChatGPT [24]

[2] A metaphor of inputs vs. outputs.

[3] It is noted that the misuse of language about ethic issues is not discussed in the paper.

doi.org/10.32388/IMZI2Q

[4] It is speculated that in one of the ChatGPT´s rivals *Ernie* released by Baidu, the Chinese prompts are firstly translated to English so that an open-source stable diffusion big model can be utilized, i.e. using other companies' AI tools to generate results[76][77].

[5] 'task decomposing' in 'Improvement of prompt: manners of prompting'

References

1. ^*Wikipedia contributors, "Chatgpt — Wikipedia, the free encyclopedia." https://en. wikipedia.org/w/index.php?title=ChatGPT&oldid=1146153556, 2023. [Online; accessed 23March-2023].*
2. ^*C. Zhang, C. Zhang, S. Zheng, Y. Qiao, C. Li, M. Zhang, S. K. Dam, C. M. Thwal, Y. L. Tun, L. L. Huy, et al., "A complete survey on generative ai (aigc): Is chatgpt from gpt-4 to gpt-5 all you need?," arXiv preprint arXiv:2303.11717, 2023.*
3. ^*https://www.theguardian.com/profile/samantha lock, "What is AI chatbot phenomenon ChatGPT and could it replace humans? — theguardian.com." https://www.theguardian.com/technology/2022/dec/05/ what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans. [Accessed 23-Mar-2023].*
4. ^*A. Goldfarb, B. Taska, and F. Teodoridis, "Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings," Research Policy, vol. 52, no. 1, p. 104653, 2023.*
5. a, b*D. Haluza and D. Jungwirth, "Artificial intelligence and ten societal megatrends: An exploratory study using gpt-3," Systems, vol. 11, no. 3, p. 120, 2023.*
6. a, b*D. M. Katz, M. J. Bommarito, S. Gao, and P. Arredondo, "Gpt-4 passes the bar exam," Available at SSRN 4389233, 2023.*
7. ^*A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," arXiv preprint arXiv:2204.06125, 2022.*
8. a, b, c*T. Eloundou, S. Manning, P. Mishkin, and D. Rock, "Gpts are gpts: An early look at the labor market impact potential of large language models," arXiv preprint arXiv:2303.10130, 2023.*
9. ^*B. Swanson, K. Mathewson, B. Pietrzak, S. Chen, and M. Dinalescu, "Story centaur: Large language model few shot learning as a creative writing tool," in Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, pp. 244–256, 2021.*

10. a, b, c, d, e S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang, "Sparks of artificial general intelligence: Early experiments with gpt-4," 2023.
11. ^ OpenAI, "Gpt-4 technical report.," Technical report, OpenAI, 2023.
12. a, b, c A. Chan, "Gpt-3 and instructgpt: technological dystopianism, utopianism, and "contextual" perspectives in ai ethics and industry," AI and Ethics, pp. 1–12, 2022.
13. a, b C. Anton, "Probing chat gpt: A media ecology writing sampler," New Explorations: Studies in Culture and Communication, vol. 3, no. 1, 2023.
14. ^ D. Jungwirth and D. Haluza, "Feasibility study on utilization of the artificial intelligence gpt-3 in public health," 2023.
15. ^ J. Longo, "When artificial intelligence meets real public administration," Canadian Public Administration, vol. 65, no. 2, pp. 384–388, 2022.
16. a, b C. J. Cai, E. Reif, N. Hegde, J. Hipp, B. Kim, D. Smilkov, M. Wattenberg, F. Viegas, G. S. Corrado, M. C. Stumpe, et al., "Human-centered tools for coping with imperfect algorithms during medical decision-making," in Proceedings of the 2019 chi conference on human factors in computing systems, pp. 1–14, 2019.
17. ^ A. Levy, M. Agrawal, A. Satyanarayan, and D. Sontag, "Assessing the impact of automated suggestions on decision making: Domain experts mediate model errors but take less initiative," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–13, 2021.
18. ^ G. Betz, K. Richardson, and C. Voigt, "Thinking aloud: Dynamic context generation improves zero-shot reasoning performance of gpt-2," arXiv preprint arXiv:2103.13033, 2021.
19. ^ J. Liu, D. Shen, Y. Zhang, B. Dolan, L. Carin, and W. Chen, "What makes good in-context examples for gpt-3?," 2021.
20. ^ Y. Lu, M. Bartolo, A. Moore, S. Riedel, and P. Stenetorp, "Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity," 2022.
21. a, b L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," in Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–7, 2021.
22. ^ I. Trummer, "Codexdb: Synthesizing code for query processing from natural language instructions using gpt-3 codex," Proceedings of the VLDB Endowment, vol. 15, no. 11, pp. 2921– 2928, 2022.

23. ^ a, b, c, d, e T. Wu, M. Terry, and C. J. Cai, “Ai chains Transparent and controllable human-ai interaction by chaining large language model prompts,” in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–22, 2022.

24. ^ a, b, c G. Mialon, R. Dess`ı, M. Lomeli, C. Nalmpantis, R. Pasunuru, R. Raileanu, B. Rozi`ere, T. Schick, J. Dwivedi-Yu, A. Celikyilmaz, E. Grave, Y. LeCun, and T. Scialom, “Augmented language models: a survey,” 2023.

25. ^ L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, “Training language models to follow instructions with human feedback,” 2022.

26. ^ Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan, N. Joseph, S. Kadavath, J. Kernion, T. Conerly, S. El-Showk, N. Elhage, Z. Hatfield-Dodds, D. Hernandez, T. Hume, S. Johnston, S. Kravec, L. Lovitt, N. Nanda, C. Olsson, D. Amodei, T. Brown, J. Clark, S. McCandlish, C. Olah, B. Mann, and J. Kaplan, “Training a helpful and harmless assistant with reinforcement learning from human feedback,” 2022.

27. ^ a, b, c P. Denny, V. Kumar, and N. Giacaman, “Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language,” in Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, pp. 1136–1142, 2023.

28. ^ M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. de Oliveira Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, S. Gray, N. Ryder, M. Pavlov, A. Power, L. Kaiser, M. Bavarian, C. Winter, P. Tillet, F. P. Such, D. Cummings, M. Plappert, F. Chantzis, E. Barnes, A. Herbert-Voss, W. H. Guss, A. Nichol, A. Paino, N. Tezak, J. Tang, I. Babuschkin, S. Balaji, S. Jain, W. Saunders, C. Hesse, A. N. Carr, J. Leike, J. Achiam, V. Misra, E. Morikawa, A. Radford, M. Knight, M. Brundage, M. Murati, K. Mayer, P. Welinder, B. McGrew, D. Amodei, S. McCandlish, I. Sutskever, and W. Zaremba, “Evaluating large language models trained on code,” 2021.

29. ^ G. Zuccon and B. Koopman, “Dr chatgpt, tell me what i want to hear: How prompt knowledge impacts health answer correctness,” arXiv preprint arXiv:2302.13793, 2023.

30. ^ Q. Lyu, J. Tan, M. E. Zapadka, J. Ponnatapuram, C. Niu, G. Wang, and C. T. Whitlow, “Translating radiology reports into plain language using chatgpt and gpt-4 with prompt learning: Promising results, limitations, and potential,” arXiv preprint arXiv:2303.09038, 2023.

31. ^ D. Baidoo-Anu and L. Owusu Ansah, “Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning,” Available at SSRN 4

337484, 2023.

32. ^*W. Jiao, W. Wang, J.-t. Huang, X. Wang, and Z. Tu, "Is chatgpt a good translator? a preliminary study," arXiv preprint arXiv:2301.08745, 2023.*
33. ^*W. Gu, "Linguistically informed chatgpt prompts to enhance japanese-chinese machine translation: A case study on attributive clauses," arXiv preprint arXiv:2303.15587, 2023.*
34. ^*S. Biswas, "Importance of chat gpt in agriculture: According to chat gpt," Available at SSRN 4405391, 2 023.*
35. ^*E. Shue, L. Liu, B. Li, Z. Feng, X. Li, and G. Hu, "Empowering beginners in bioinformatics with chatgp t," bioRxiv, pp. 2023–03, 2023.*
36. ^*M. Bommarito II and D. M. Katz, "Gpt takes the bar exam," arXiv preprint arXiv:2212.14402, 2022.*
37. ^*S. Vemprala, R. Bonatti, A. Bucker, and A. Kapoor, "Chatgpt for robotics: Design principles and model a bilities," 2023.*
38. ^*C. Guo, Y. Lu, Y. Dou, and F.-Y. Wang, "Can chatgpt boost artistic creation: The need of imaginative int elligence for parallel art," IEEE/CAA Journal of Automatica Sinica, vol. 10, no. 4, pp. 835–838, 2023.*
39. ^*X. Hu, Y. Tian, K. Nagato, M. Nakao, and A. Liu, "Opportunities and challenges of chatgpt for design kn owledge management," arXiv preprint arXiv:2304.02796, 2023.*
40. ^*S. A. Prieto, E. T. Mengiste, and B. Garc´ıa de Soto, "Investigating the use of chatgpt for the scheduling o f construction projects," Buildings, vol. 13, no. 4, p. 857, 2023.*
41. ^*B. Rathore, "Future of textile: Sustainable manufacturing & prediction via chatgpt," Eduzone: Internat ional Peer Reviewed/Refereed Multidisciplinary Journal, vol. 12, no. 1, pp. 52– 62, 2023.*
42. ^*T. Yue, D. Au, C. C. Au, and K. Y. Iu, "Democratizing financial knowledge with chatgpt by openai: Unlea shing the power of technology," Available at SSRN 4346152, 2023.*
43. ^*P. Rivas and L. Zhao, "Marketing with chatgpt: Navigating the ethical terrain of gpt-based chatbot tec hnology," AI, vol. 4, no. 2, pp. 375–384, 2023.*
44. ^*D. Roam, Blah blah blah: What to do when words don't work. Penguin, 2011.*
45. ^*I. Van de Poel, "Embedding values in artificial intelligence (ai) systems," Minds and Machines, vol. 30, no. 3, pp. 385–409, 2020.*
46. ^*G. Davis, M. Grierson, et al., "Investigating attitudes of professional writers to gpt text generation ai ba sed creative support tools," 2022.*
47. ^*"Understanding prompts, completions, and tokens," Mar 2023.*

48. ^M. Song, H. Jiang, S. Shi, S. Yao, S. Lu, Y. Feng, H. Liu, and L. Jing, “Is chatgpt a good keyphrase generator? a preliminary study,” arXiv preprint arXiv:2303.13001, 2023.
49. ^X. Zhai, “Chatgpt for next generation science learning,” Available at SSRN 4331313, 2023.
50. ^T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” 2020.
51. ^J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, “Chain-of-thought prompting elicits reasoning in large language models,” 2023.
52. ^T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, “Large language models are zero-shot reasoners,” 2023.
53. ^T. Shin, Y. Razeghi, R. L. Logan IV, E. Wallace, and S. Singh, “Autoprompt: Eliciting knowledge from language models with automatically generated prompts,” arXiv preprint arXiv:2010.15980, 2020.
54. ^V. Raunak and A. Menezes, “Finding memo: Extractive memorization in constrained sequence generation tasks,” 2022.
55. ^S. Agrawal, C. Zhou, M. Lewis, L. Zettlemoyer, and M. Ghazvininejad, “In-context examples selection for machine translation,” 2022.
56. ^E. Collins and Z. Ghahramani, “Lamda: our breakthrough conversation technology,” Google AI Blog, 2021.
57. ^D. Adiwardana, M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, and Q. V. Le, “Towards a human-like open-domain chatbot,” 2020.
58. ^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.
59. ^B. Barak, B. L. Edelman, S. Goel, S. Kakade, E. Malach, and C. Zhang, “Hidden progress in deep learning: Sgd learns parities near the computational limit,” 2023.
60. ^K. Ahn, S. Bubeck, S. Chewi, Y. T. Lee, F. Suarez, and Y. Zhang, “Learning threshold neurons via the ”edge of stability”,” 2022.
61. ^C. Olah, N. Cammarata, L. Schubert, G. Goh, M. Petrov, and S. Carter, “Zoom in: An introduction to circuits,” Distill, vol. 5, no. 3, pp. e00024–001, 2020.

62. ^C. Olsson, N. Elhage, N. Nanda, N. Joseph, N. DasSarma, T. Henighan, B. Mann, A. Askell, Y. Bai, A. Chen, T. Conerly, D. Drain, D. Ganguli, Z. Hatfield-Dodds, D. Hernandez, S. Johnston, A. Jones, J. Kernion, L. Lovitt, K. Ndousse, D. Amodei, T. Brown, J. Clark, J. Kaplan, S. McCandlish, and C. Olah, "In-context learning and induction heads," 2022.
63. ^H. M. Wellman, The child's theory of mind. The MIT Press, 1992.
64. ^Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba, "Large language models are human-level prompt engineers," 2023.
65. ^G. Qin and J. Eisner, "Learning how to ask: Querying lms with mixtures of soft prompts," 2021.
66. ^P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," ACM Computing Surveys, vol. 55, no. 9, pp. 1–35, 2023.
67. ^E. M. Bender and A. Koller, "Climbing towards nlu: On meaning, form, and understanding in the age of data," in Proceedings of the 58th annual meeting of the association for computational linguistics, pp. 5185–5198, 2020.
68. ^B. Tan, Z. Yang, M. AI-Shedivat, E. P. Xing, and Z. Hu, "Progressive generation of long text with pretrained language models," arXiv preprint arXiv:2006.15720, 2020.
69. ^G. Branwen, "Gpt-3 creative fiction," 2020.
70. ^J. O'Connor and J. Andreas, "What context features can transformer language models use?," arXiv preprint arXiv:2106.08367, 2021.
71. ^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," arXiv preprint arXiv:2302.11382, 2023.
72. ^GPT-3, "A robot wrote this entire article. are you scared yet, human?," The Guardian, 2020.
73. ^K. LaGrandeur, "How safe is our reliance on ai, and should we regulate it?," AI and Ethics, vol. 1, pp. 93–99, 2021.
74. ^P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," 2021.
75. ^T. Saveratto, "Autogpt - your personal ai assistant in less than 30 minutes," Apr 2023.
76. ^"Baidu defends ai research capabilities after ernie bot is accused of copying," Mar 2023.

77. ^ *T. Feed, “Baidu denies claims its ernie bot tool is ”piggybacking” on other chatgpt-like programs,” Mar 2023.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.