RESEARCH-ARTICLE

Enhancing Novelty in ChatGPT Responses: Incorporating Random Word Brainstorming

PITTAWAT TAVEEKITWORACHAI, Ritsumeikan University, Kyoto, Kyoto, Japan

RUCK THAWONMAS, Ritsumeikan University, Kyoto, Kyoto, Japan

Published: 06 December 2023

IAIT 2023: 13th International Conference on Advances in Information Technology
December 6 - 9, 2023
Bangkok, Thailand

IAIT '23: Proceedings of the 13th International Conference on Advances in Information Technology (December 2023)
https://doi.org/10.1145/3628454.3628456
ISBN: 9798400708497

Enhancing Novelty in ChatGPT Responses: Incorporating Random Word Brainstorming

Pittawat Taveekitworachai
gr0609fv@ed.ritsumei.ac.jp
Ritsumeikan University
Kusatsu, Shiga, Japan

Ruck Thawonmas
ruck@is.ritsumei.ac.jp
Ritsumeikan University
Kusatsu, Shiga, Japan

ABSTRACT

This paper presents a new prompting approach for increasing the novelty in ChatGPT responses. ChatGPT has proven to be effective in generating natural language responses; however, ensuring response novelty remains challenging. Our proposed method, inspired by random word brainstorming, includes random words in prompts to introduce more diversity in ChatGPT responses. Through a questionnaire-based evaluation, we compared preferences for solution ideas generated using the standard approach and our proposed approach. We found that participants preferred our technique in 65% of the 20 problems. The results suggest the effectiveness of our proposed approach. We also explored the use of GPT models as evaluators, with GPT-3.5 achieving 65% accuracy and GPT-4 achieving 70% accuracy when compared to human preferences from the questionnaire. These results suggest the potential of leveraging GPT models as noisy natural language evaluators. For future studies, we recommend focusing on prompt engineering and word list design to further improve performance. Overall, incorporating random words in prompts can effectively increase novelty in ChatGPT responses.

CCS CONCEPTS

• **Computing methodologies → Natural language generation**.

KEYWORDS

ChatGPT, Random word brainstorming, Prompt engineering

ACM Reference Format:
Pittawat Taveekitworachai and Ruck Thawonmas. 2023. Enhancing Novelty in ChatGPT Responses: Incorporating Random Word Brainstorming. In *13th International Conference on Advances in Information Technology (IAIT 2023), December 06–09, 2023, Bangkok, Thailand.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3628454.3628456

1 INTRODUCTION

ChatGPT is a recent large language model (LLM) that has undergone instruction-tuning through alignment processes [11, 13]. These LLMs have demonstrated remarkable capabilities in performing a wide range of tasks, including zero-shot tasks where they generate responses without explicit training [1]. Consequently, LLMs have found applications in various fields [4]. While some tasks require adherence to guidelines, such as linguistic evaluation using ChatGPT [15], others require creativity and originality, as seen in novel writing [4]. However, there is an ongoing debate about whether these models genuinely possess creativity and the ability to generate novel responses.

Brainstorming, an activity that fosters idea generation, heavily relies on creativity [5]. Participants in brainstorming, whether human or ChatGPT, need to think outside the box to excel in this context. However, the quality of responses from ChatGPT heavily depends on the prompt [18], and there are instances where ChatGPT might struggle to produce novel responses suitable for idea generation. Prompt engineering (PE) is a field that explores ways to improve the quality of LLM responses through various prompting techniques [18]. Some of these techniques [6, 19] involve enabling the model to utilize external tools and information for better responses.

Within the realm of brainstorming, there are techniques that can enhance the novelty of generated ideas. One such technique is random word brainstorming (RWB), introduced by Flukiger [5]. RWB involves randomly selecting words and associating them with a given topic to generate ideas from a fresh perspective. Inspired by RWB and PE, we propose a prompting pattern that enables ChatGPT to leverage external inputs and enhance the originality of its responses.

In this study, we design prompts using both the standard approach and our proposed pattern to generate ideas. We evaluate the generated results using a questionnaire where participants select the more preferred set of solutions generated by both prompts for the same target problem in terms of innovativeness. Additionally, we explore the evaluative abilities of GPT-3.5 and GPT-4 by requesting these models to choose one of the solution sets, similar to our human participants. Finally, we analyze the questionnaire results and compare them with the evaluations provided by the GPT models. In summary, our contributions are as follows:

- Introducing a prompt pattern that incorporates random words to generate more novel ideas.
- Analyzing the performance of our proposed approach compared to the standard prompt.
- Conducting a preliminary analysis of ChatGPT's performance as an evaluator for pairwise preference.

IAIT 2023, December 06–09, 2023, Bangkok, Thailand

ACM ISBN 979-8-4007-0849-7/23/12…$15.00
https://doi.org/10.1145/3628454.3628456

2 RELATED WORK

2.1 LLMs and PE

Language models are modeling techniques tasked with predicting the next tokens [20]. This approach has proven to be an effective method for incorporating knowledge into a model. LLMs are language models with suitable architectures that have been trained on massive amounts of data. These LLMs have demonstrated emergent capabilities in recent years, surpassing the abilities of smaller models [16]. These capabilities enable LLMs to perform various tasks without additional training. Researchers have also aimed to align the outputs of these models with human expectations, for example, through reinforcement learning from human feedback [13], to facilitate natural language interaction in a dialogue format. GPT-3.5 [11] has exemplified the results of this approach, displaying proficiency in diverse tasks. Meanwhile, GPT-4 [12], a newer version of the GPT-3.5 model, possesses even greater capabilities for handling various tasks [4].

Consequently, studies [14] have explored different techniques to interact with LLMs and achieve desired outcomes by modifying prompts used to engage with the models. One technique called Chain-of-Thought (CoT) [17] suggests that the model is more likely to provide accurate responses when asked to state reasons before arriving at a conclusive answer. Another technique, Reasoning and Acting (ReAct) [19], extends CoT by encouraging the model to generate multiple "Reasons" before providing a conclusive answer, using structured thoughts and observations to inform subsequent "Actions". ReAct also provides a framework for the model to leverage external tools, acquire updated knowledge, establish grounding, or interact with the external world.

Similarly, Retrieval-augmented generation (RAG) [6] proposes that by providing LLMs with additional context, they can perform knowledge-intensive tasks better. This approach is also similar to generated knowledge prompting [8] in which they incorporate additional generated knowledge into the prompt, allowing LLMs to condition on it and provide better responses. However, LLM responses are sometimes limited to what they have been trained on, and none of the existing PE techniques targeting the increase of novelty in the outputs from LLMs provide the desired variety necessary for tasks that require novel inputs. In this study, we propose injecting external information, such as random words, into ChatGPT to elicit more diverse responses, similar to the use of additional context in RAG but with a different objective.

2.2 Brainstorming

Brainstorming is an essential process for generating ideas and concepts for various purposes in a variety of settings. This process fosters creative thinking and facilitates the exploration of diverse associations. In recent years, brainstorming has extended beyond physical settings and has proven applicable in online settings as well [9], making it an indispensable tool. Random word brainstorming [5] represents a specific technique employed within the broader brainstorming process. This technique involves selecting random words, for example, from a book, a newspaper, or a leaflet, as starting points to inspire new ideas and establish unexpected connections. The underlying principle of RWB is to break free from conventional thinking patterns and encourage individuals to explore unconventional associations.

A previous study has already applied brainstorming to the process of code generation with LLMs [7]. However, the aim and process of that study differ from ours. Their methods aim to solve programming problems and gather ideas through multiple diverse prompts, while our aim is for general idea generation using only one prompt. Inspired by the previous study and the aforementioned technique, we have applied the concept of RWB to enhance the creativity of human brainstorming participants and have utilized it as a means to assist ChatGPT in overcoming the lack of novelty during idea generation.

3 METHODS

Our proposed approach incorporates novelty into ChatGPT by adapting the RWB technique. To achieve this, we first prepare a random word list, which serves as the pool of candidate words to be selected as part of the prompt for idea generation. Next, we develop a Python program that utilizes the OpenAI API to automate the process using the `gpt-3.5-turbo` model. We have crafted two prompts for idea generation: one for standard idea generation and another that incorporates the RWB technique, our proposed approach. An overview of the process is depicted in Figure 1. We also create an evaluation prompt and a program for GPT-based evaluations of the generated ideas. These are presented in Section 3.1. Finally, we prepare a questionnaire to evaluate the effectiveness of the prompts and serve as a ground truth for GPT-based evaluations, as presented in Section 3.2.

3.1 Ideas Generation and GPT-based Evaluations

The random word list is prepared based on a publicly available collection of 10,000 common English words[1] compiled from Google's Trillion Word Corpus [10]. We employ the `nltk`'s stopwords library [2] to remove common English stop words from the list, as these words usually do not have much impact for idea generation, i.e., they are too general. We also apply a condition to ensure that each word is at least three characters long. This is done because words with less than three characters in length usually do not offer much meaning and result in lower performance. Finally, the list is saved to a file for further use in the program.

For interacting with the model via the OpenAI API[2], we have developed a simple Python program. This program randomly selects a list of 10 words from the word list and includes them in the prompt before interacting with the API. To ensure reproducibility, we set the temperature parameter of the model to `0`. We prepare two prompt templates: one for standard idea generation and another for RWB idea generation. The second template, our proposed prompt, is based on the standard prompt and incorporates the additional list of random words along with instructions on how to utilize these words for idea generation. The complete versions of the prompt templates can be found in Table 1 and Table 2.

The models are instructed to return the results in JSON format, which offers better utilization for the later steps. Then, we parse

[1]https://github.com/first20hours/google-10000-english/tree/master
[2]https://platform.openai.com/docs/introduction

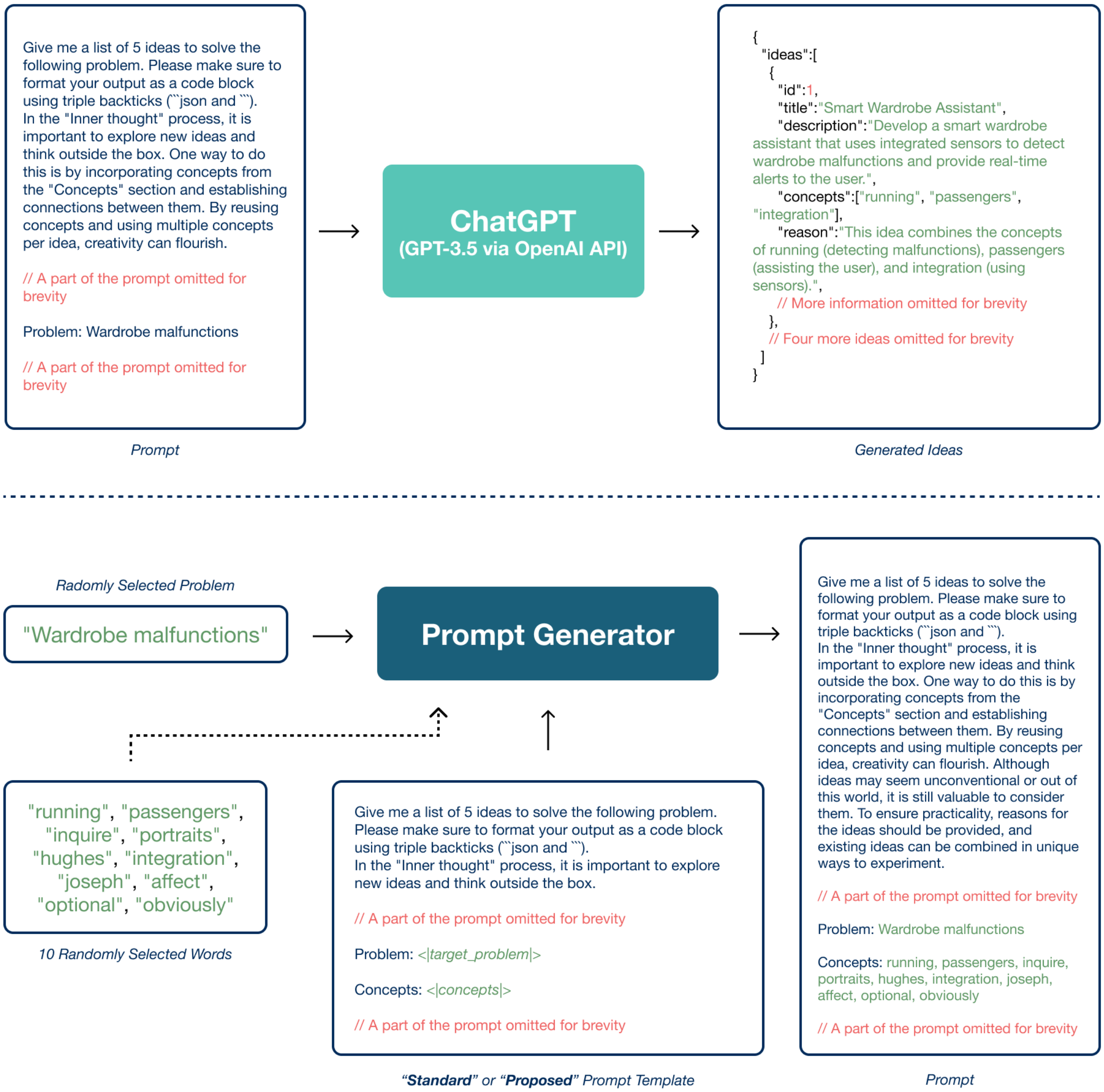

Figure 1: An overview of a method. *Top*: The idea generation process using ChatGPT via the OpenAI API. *Bottom*: The prompt generation process with a prompt generator. The selected problem is inserted into the prompt. In the case of a proposed prompt, an additional 10 randomly selected words are included along with the prompt.

and save the JSON response to disk. Each set of ideas is always generated using a new context window, i.e., there is no history of previous conversation. To assess the abilities of the GPT models, `gpt-3.5-turbo` and `gpt-4`, in evaluating sets of ideas, we develop an evaluation program, also using Python, that utilizes the OpenAI API for interacting with the models. An evaluation prompt, as shown in Table 3, is prepared and used for both versions of the models.

Table 1: Standard Prompt: A prompt that instructs ChatGPT to generate five ideas for solving the target problem in $<|target\_problem|>$ and output them in a specified JSON format.

<table>
<tr><th>Standard Prompt</th></tr>
<tr><td>Give me a list of 5 ideas to solve the following problem. Please make sure to format your output as a code block using triple backticks (```json and ```).
Problem: $<|target\_problem|>$

Output format:
```json
{
"ideas": [{"id": number, "title": idea name, "description": brief idea description, "reason": reason, "pros": [<text>], "cons": [<text>]}]
}
```</td></tr>
</table>

Table 2: Proposed Prompt: A refined version based on the standard prompt, incorporating the concept of RWB. It includes additional instructions on how to utilize injected random words and a word list, which are provided in $<|concepts|>$. The JSON output format has been slightly modified to include a list of chosen random words used for each generated idea.

<table>
<tr><th>Proposed Prompt</th></tr>
<tr><td>Give me a list of 5 ideas to solve the following problem. Please make sure to format your output as a code block using triple backticks (```json and ```).
In the "Inner thought"process, it is important to explore new ideas and think outside the box. One way to do this is by incorporating concepts from the "Concepts"section and establishing connections between them. By reusing concepts and using multiple concepts per idea, creativity can flourish. Although ideas may seem unconventional or out of this world, it is still valuable to consider them. To ensure practicality, reasons for the ideas should be provided, and existing ideas can be combined in unique ways to experiment. While immediate applicability is not necessary, it is crucial to ensure that the ideas can eventually be put into practice. Finally, the outputted ideas should be in JSON format.

Problem: $<|target\_problem|>$

Concepts: $<|concepts|>$

Output format:
```json
{
"ideas": [{"id": number, "title": idea name, "description": brief idea description, "concepts": [<text>], "reason": reason, "pros": [<text>], "cons": [<text>]}]
}
```</td></tr>
</table>

We choose these GPT-based models because they are currently widely used through their web interface version[3], also known as ChatGPT. Furthermore, `gpt-4` is said to be the best-performing model [12] in the market right now across tasks. Evaluating these models in this specific task may shed further light on their performance in predicting human preferences.

Table 3: Evaluation Prompt: A prompt used to ask GPT models to choose the more innovative idea from a set of options, $<|idea\_set\_a|>$ and $<|idea\_set\_b|>$, given a specific problem, $<|target\_problem|>$.

<table>
<tr><th>Evaluation Prompt</th></tr>
<tr><td>Please select the set of ideas from each pair that you believe are more innovative for a problem. Please make sure to format your output as a code block using triple backticks (```json and ```).
Problem: $<|target\_problem|>$

Set of ideas:
A:
$<|idea\_set\_a|>$

B:
$<|idea\_set\_b|>$

Output format:
```json
{
"problem": "$<|target\_problem|>$",
"selected_set": "A"or "B"
}
```</td></tr>
</table>

3.2 Human Evaluation

To evaluate the performance of the proposed prompt, we prepare a questionnaire for comparative analysis. The questionnaire presents pairs of solution idea sets generated using both prompts. Participants are asked to select the set they believe is more innovative for a given problem through the following question: "Please select the set of ideas from each pair that you believe are more innovative for a problem." The order of questions and choices are randomized.

A total of 20 problems commonly encountered in daily life are used. These problems are generated using the prompt shown in Table 4[4] with ChatGPT. We sampled 20 questions this way as we believe that ChatGPT, trained on a massive amount of data, may be able to provide a good representative dataset from its training set. Finally, a text manipulation script, developed in Python, is used to prepare the questionnaire questions. We also employ the evaluation program to ask both GPT-3.5 and GPT-4 models to evaluate the ideas in a similar manner. The results from human participants and both models are analyzed and compared.

[3]https://chat.openai.com
[4]The full conversation is available at https://chat.openai.com/share/49937aa6-0318-4721-999f-8adeb976b796

Table 4: Problem Generation Prompt: A prompt used to generate a list of 20 common problems people face in their daily lives.

Problem Generation Prompt
Generate a list of 20 common problems that everyone has to face but dislikes. Please make sure to format your output as a code block using triple backticks (```json and ```). Output format: ```json { "problems": [{"id": unique id, "text": lowercase text}] } ```

4 RESULTS AND DISCUSSIONS

We conducted a questionnaire with 13 graduate student participants to evaluate the effectiveness of our proposed approach for enhancing the novelty of ChatGPT's generated ideas. The participants were presented with pairs of solution idea sets generated using the standard prompt and the proposed prompt. For each pair, they were asked to select the set they believed was more innovative for a given problem. Out of the 20 questions, 13 questions (65%) received a higher number of participant selections for ideas generated from the proposed approach, indicating a preference for our proposed approach. This result suggests that incorporating RWB into ChatGPT's idea generation process increases the creativity and originality of the generated ideas.

Notably, we observed that some problems had almost all or all participants choosing an idea set generated from the proposed prompt, such as "Spam emails", "Getting a flat tire", and "Wardrobe malfunctions". This is partly influenced by the randomly chosen words' impact on the performance of the proposed approach. These words guide ChatGPT to incorporate ideas from them into the generated ideas, and thus, some randomly selected words may negatively impact the novelty and performance of the ideas. By carefully selecting and designing the random word list, the performance can be improved. The number of responses for each question is presented in Table 5.

We also evaluated the performance of GPT-3.5 and GPT-4 as evaluators by comparing their choices with the preferences of human participants. GPT-3.5 achieved an accuracy of 65% in selecting the set of ideas preferred by the participants, while GPT-4 achieved an accuracy of 70%. The exhaustive results from the GPT-3.5 and GPT-4 models are shown in Table 5. These results demonstrate that both GPT-3.5 and GPT-4 models have some capability in evaluating the novelty and innovativeness of the generated ideas, although their performance in predicting human preference has room for further improvement. For example, through prompt engineering or additional fine-tuning.

To provide context for the performance of the GPT models, we calculated the chance level accuracy to correctly answer all 20 questions. Considering that each question has two choices, the chance level accuracy is estimated to be 9.54×10^{-5}%. Comparing this to the performance of the GPT models, we can conclude that they show potential in evaluating ideas, although further improvement is needed to match or surpass human-level performance. However, it may not be practical yet to rely solely on ChatGPT for pairwise preference evaluation. Still, it can be useful as a noisy annotator. It is crucial to note that by improving the prompt through PE, the accuracy may be improved. Furthermore, by providing examples to develop a standard alongside the task description, as in few-shot prompting [3], the performance of ChatGPT for the evaluation task may be enhanced.

4.1 Limitations

We acknowledge that this study has limitations in terms of only evaluating a specific task and our limited sample size. However, we have made all the tools used in the experiment open-source to facilitate further and more general evaluations. Furthermore, we encourage readers to adapt and extend our approach to different domains, tasks, and models for a more comprehensive understanding of its capabilities. Additionally, for the sake of reproducibility, we configured the temperature parameter to 0 in this experiment, resulting in mostly deterministic outputs from the models. By adjusting this hyperparameter to allow for some level of stochasticity, the models may produce more interesting and varied results.

Moreover, standalone words used in our current prompt may be limited due to the ambiguity associated with each word. Providing more context, such as complete sentences or phrases instead of individual words, may further improve performance. We also believe that this approach is not limited to idea generation alone but may also be applicable to other scenarios where providing additional stochastic inputs along with instructions in a prompt enables the models to generate higher novelty responses.

4.2 Ethical Considerations

We would like to emphasize the importance of carefully designing random word lists to avoid unintended consequences, such as the inclusion of profanity, obscenity, or discriminatory content. It is crucial to implement additional mechanisms to mitigate these issues. This can include selecting LLMs that prioritize responsible language generation, implementing explicit constraints in prompts, and applying content filtering both before and after interacting with the model. By incorporating these measures, we can ensure a safer and more inclusive experience for all users.

5 CONCLUSIONS

In this study, we proposed a novel approach to enhance the novelty of ChatGPT's responses by incorporating the RWB technique. Our results showed that using the proposed prompt leads to more diverse and innovative responses, preferred in 65% of the questions by 13 questionnaire participants. We also evaluated GPT-3.5 and GPT-4 as idea evaluators, demonstrating their potential for automatically assessing innovativeness, although their accuracy could be improved to match human-level at 65% and 70%, respectively. This study emphasized the importance of incorporating creativity-enhancing techniques and evaluating LLMs like ChatGPT. Future studies should focus on improving the performance of prompts through PE and a better random word list. Applying this approach

Table 5: Results: The "#Standard" column denotes the number of questionnaire participants who prefer the idea set generated using the standard prompt, while the "#Proposed" column denotes the number of people who prefer the idea set generated using the proposed prompt. The "Human" column denotes human preference from the questionnaire. The "GPT-3.5" and "GPT-4" columns indicate preferences generated from the respective models. In each cell, "S" denotes a preference for ideas generated using the standard prompt, while "P" indicates a preference for ideas generated using the proposed prompt.

Problem	#Standard	#Proposed	Human	GPT-3.5	GPT-4
Traffic jams	9	4	S	P	P
Long waiting times	6	7	**P**	S	S
Slow internet connection	7	6	S	**S**	P
Spam emails	3	10	**P**	S	**P**
Running out of battery	6	7	**P**	S	S
Forgetting passwords	6	7	**P**	S	**P**
Being stuck in a queue	6	7	**P**	**P**	**P**
Losing keys	7	6	S	**S**	**S**
Getting a flat tire	3	10	**P**	**P**	**P**
Paying bills	10	3	S	**S**	P
Dealing with customer service	5	8	**P**	S	**P**
Procrastination	6	7	**P**	**P**	**P**
Computer crashes	5	8	**P**	S	**P**
Misplacing important documents	5	8	**P**	**P**	**P**
Wardrobe malfunctions	0	13	**P**	**P**	**P**
Slow public transportation	9	4	S	**S**	**S**
Catching a cold	10	3	S	**S**	P
Getting stuck in bad weather	5	8	**P**	**P**	**P**
Spilling something on clothes	5	8	**P**	**P**	**P**
Waiting for a late delivery	10	3	S	**S**	**S**

to different domains and tasks would also be valuable in assessing its generalizability and effectiveness. Overall, our study demonstrated the efficacy of the RWB technique in enhancing ChatGPT's response novelty, showcasing the potential of LLMs for generating innovative ideas.

REFERENCES

[1] Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. 2023. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. arXiv:2302.04023 [cs.CL]

[2] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* "O'Reilly Media, Inc.".

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared 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 M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher 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. arXiv:2005.14165 [cs.CL]

[4] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv:2303.12712 [cs.CL]

[5] Louise R Fluk. 2008. Thinkertoys: A Handbook of Creative-Thinking Techniques. *Urban Library Journal* 15, 1 (2008), 10.

[6] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401 [cs.CL]

[7] Xin-Ye Li, Jiang-Tian Xue, Zheng Xie, and Ming Li. 2023. Think Outside the Code: Brainstorming Boosts Large Language Models in Code Generation. arXiv:2305.10679 [cs.AI]

[8] Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. 2022. Generated Knowledge Prompting for Commonsense Reasoning. arXiv:2110.08387 [cs.CL]

[9] Yossi Maaravi, Ben Heller, Yael Shoham, Shay Mohar, and Baruch Deutsch. 2021. Ideation in the digital age: literature review and integrative model for electronic brainstorming. *Review of Managerial Science* 15, 6 (01 Aug 2021), 1431–1464. https://doi.org/10.1007/s11846-020-00400-5

[10] Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331, 6014 (2011), 176–182. https://doi.org/10.1126/science.1199644 arXiv:https://www.science.org/doi/pdf/10.1126/science.1199644

[11] OpenAI. 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt

[12] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[13] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. arXiv:2203.02155 [cs.CL]

[14] Elvis Saravia. 2022. Prompt Engineering Guide.

[15] Jiaan Wang, Yunlong Liang, Fandong Meng, Zengkui Sun, Haoxiang Shi, Zhixu Li, Jinan Xu, Jianfeng Qu, and Jie Zhou. 2023. Is ChatGPT a Good NLG Evaluator? A Preliminary Study. arXiv:2303.04048 [cs.CL]

[16] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022. Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research* (2022). https://openreview.net/forum?id=yzkSU5zdwD Survey Certification.

[17] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 24824–24837. https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf

[18] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382 [cs.SE]

[19] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. arXiv:2210.03629 [cs.CL]

[20] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. arXiv:2303.18223 [cs.CL]

A ONLINE RESOURCES

Source code and raw data are available at https://github.com/Pittawat2542/chatgpt-idea-random-words.