

Aligning LLM with Human Neural Pathways: Enhancing Cognitive Synergy

Anonymous ACL submission

Abstract

Recent studies have highlighted the similarities between human neural pathways and the operational mechanisms of Large Language Models (LLMs). While the Solo Performance Prompting (SPP) exhibits cognitive synergy akin to human collaboration in LLMs like GPT-4, it faces limitations in generalizability and defining the conditions for cognitive synergy emergence. To address these issues, we introduce **Brain Performance Prompting (BPP)**, an innovative framework inspired by human neural pathways. BPP dynamically activates task-specific brain region personas to LLMs, enhancing self-collaboration and advancing cognitive synergy beyond SPP. Moreover, BPP offers new insights into the emergence of cognitive synergy by revealing its partial presence in smaller models, including GPT-4o-mini, Qwen-2.5-7B-Instruct, and Llama-3.1-8B-Instruct. Our experiments demonstrate that BPP significantly outperforms SPP and other approaches across knowledge-intensive and reasoning-intensive tasks on GPT-4o. These findings suggest that drawing inspiration from human brain information processing principles can play a crucial role in optimizing LLM performance. Our code will be made publicly available upon acceptance.

1 Introduction

Human cognitive synergy, considered an intrinsic feature of the human thought system, involves multiple cognitive processes collaborating to produce outcomes superior to those achieved by individual cognitive efforts (Mayer, 1998; Curşeu et al., 2015). Research indicates that humans begin to learn collaborative thinking through role-playing from approximately two to three years of age, which enables effective problem-solving (Piaget, 1954; Pellegrini, 2009). This collaborative approach mirrors the functionality of the human brain, which processes complex tasks based on the cooperation of various cognitive systems (Miller, 2000).

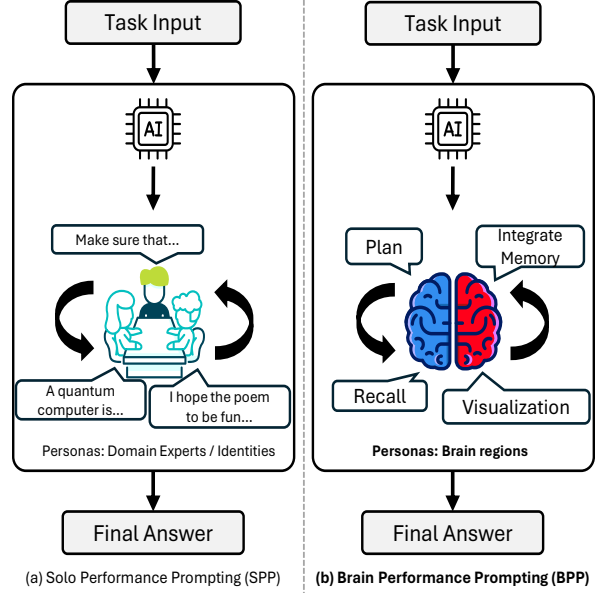


Figure 1: We compare the persona assignment strategies of SPP and BPP. While SPP uses domain experts or identity-based personas, BPP employs brain-region-based personas tailored to the characteristics of each task.

Recent studies reveal similarities between the operational mechanisms of Large Language Models (LLMs) and the neural system flows of the human brain. For example, instruction-tuning can align the internal representations of LLMs with human neural activities in language systems (Aw et al., 2024), and mechanisms extending beyond simple next-word prediction contribute to this alignment (Merlin and Toneva, 2024). These findings suggest that LLMs process information through collaborative interactions during task performance, akin to the efficient task execution observed in multiple interacting brain regions. Such similarities offer the potential to model human collaborative thinking processes within LLMs.

In this context, Solo Performance Prompting (SPP) is designed to enable LLMs like GPT-4 to

emulate human-like collaborative thinking (Wang et al., 2024b). By employing prompt engineering, SPP activates various personas within a single LLM, facilitating self-collaboration and demonstrating the emergence of cognitive synergy in these models. This cognitive synergy allows SPP to surpass previous approaches such as Chain-of-Thought (CoT) (Wei et al., 2022) and Self-Refinement (Madaan et al., 2023), significantly enhancing the problem-solving capabilities of LLMs. However, SPP has several important limitations. First, its performance has been validated only on GPT-4, leaving its effectiveness on smaller models or models with different architectures unverified. Our findings further show that SPP does not achieve the expected performance on GPT-4o, raising questions about its broader applicability. Second, the absence of clear criteria for determining the onset of cognitive synergy makes it difficult to ensure consistent performance across varying model sizes. This remains a significant challenge, limiting the potential application of SPP to a wider range of LLMs.

To address these limitations, we propose **Brain Performance Prompting (BPP)**, a novel approach inspired by the neural pathways of the human brain. Unlike SPP, BPP dynamically assigns capabilities tailored to specific brain regions based on task characteristics, enabling LLMs to perform self-collaboration to derive conclusions. Figure 1 highlights the key differences in the types of personas assigned by SPP and BPP. While SPP employs personas based on domain expertise (e.g., Poet, Math Expert) or identity (e.g., Ten-Year-Old Child, Harry Potter Fan), BPP uniquely utilizes personas modeled after brain regions (e.g., Frontal Lobe, Temporal Lobe) to mimic the neural pathways of the human brain.

To investigate the emergence of cognitive synergy through BPP, we conducted experiments on GPT-4o, o1-mini, GPT-4o-mini, GPT-3.5-turbo, Qwen-2.5-7B-Instruct, and Llama-3.1-8B-Instruct. The results indicate that cognitive synergy only emerges in models with sufficiently large parameter sizes, such as GPT-4o. This finding aligns with the observations of Wang et al. (2024b), which suggest that collaborative interactions between internal mechanisms fully manifest only beyond a certain model scale. Furthermore, our experiments demonstrate that BPP enables LLMs to exhibit a more advanced cognitive synergy compared to SPP, underscoring the effectiveness of brain-region-based

approaches in enhancing collaborative reasoning processes within LLMs. Intriguingly, partial cognitive synergy also emerged in relatively smaller models, including GPT-4o-mini, Qwen-2.5-7B-Instruct, and Llama-3.1-8B-Instruct, offering insights into the conditions and thresholds under which it arises in specific tasks. In summary, the key contributions of this study are as follows:

- We propose Brain Performance Prompting (BPP), which assigns brain-region-inspired personas to LLMs to emulate and deepen human cognitive synergy, thereby enhancing problem-solving through self-collaboration.
- BPP exhibits enhanced cognitive synergy compared to SPP in large-scale models such as GPT-4o, resulting in superior performance across a range of tasks.
- Through experiments on various models, including GPT-4o-mini, Qwen-2.5-7B-Instruct, and Llama-3.1-8B-Instruct, we observe partial manifestations of cognitive synergy in smaller models, offering insights into the conditions under which cognitive synergy emerges.

2 Related Works

2.1 Similarities between the human brain and neural networks

Recent studies have increasingly focused on investigating the alignment between LLMs and the human brain. Research exploring the representational alignment between the internal states of LLMs and human neural activities has been actively conducted (Tucker and Tuckute, 2023), and it has been demonstrated that instruction tuning can make the internal representations of such models more similar to the neural activities of the human language system (Aw et al., 2024). These studies suggest that the information processing mechanisms of LLMs may share structural and functional similarities with the human brain. From this perspective, Brain Performance Prompting (BPP) activates personas based on relevant brain regions tailored to the characteristics of a given task, emulating how the human brain dynamically and collaboratively activates various regions to meet task-specific demands (Bullmore and Sporns, 2009; Deco et al., 2008). Through this approach, BPP deepens cognitive synergy by aligning LLM mechanisms with human neural systems.

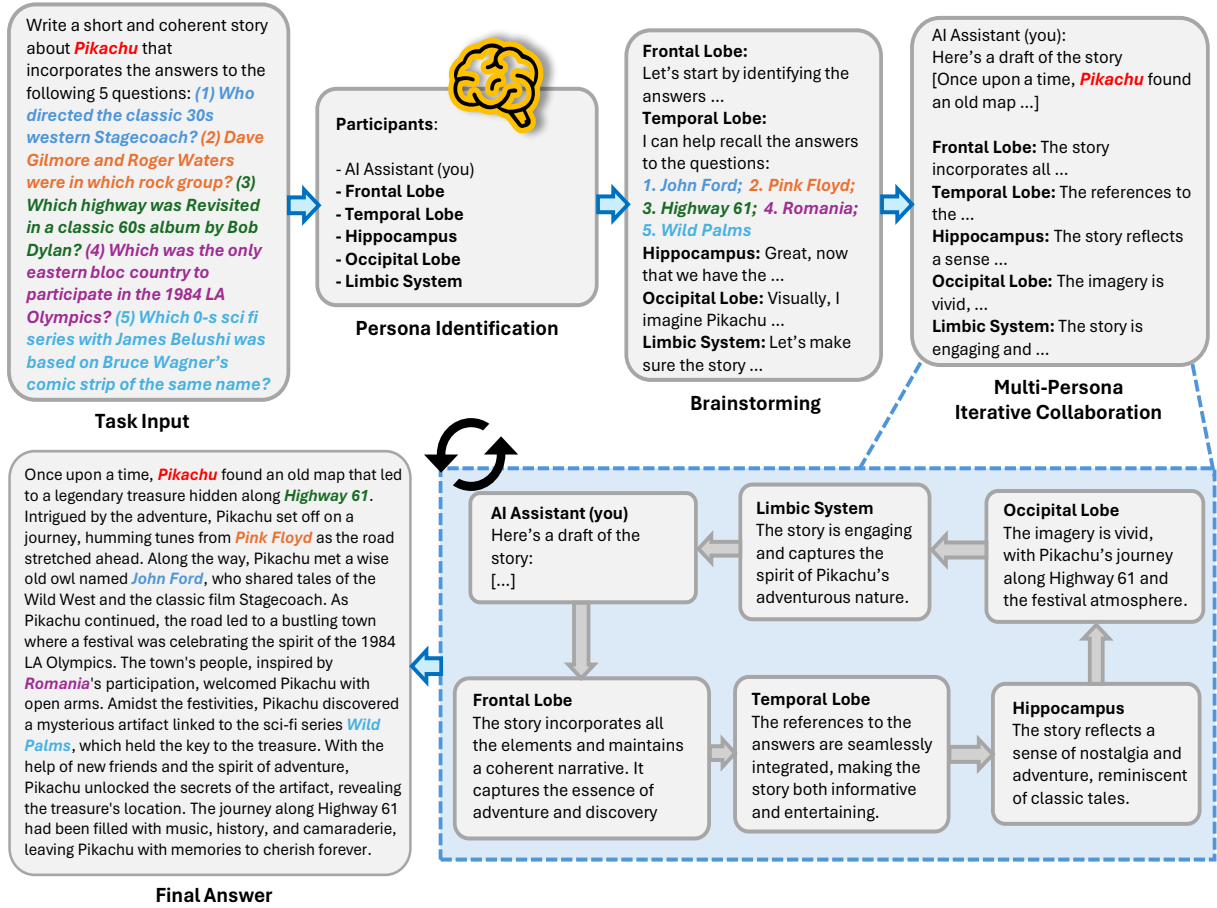


Figure 2: Overview of our BPP mechanism. The color-coded text visually distinguishes and clearly confirms that the task requirements have been thoroughly incorporated.

2.2 Improving knowledge and reasoning abilities

Despite their remarkable capabilities, LLMs still exhibit limitations in their knowledge and reasoning abilities (Bubeck et al., 2023). To address these issues, approaches such as Chain-of-Thought (CoT) prompting (Wei et al., 2022) and Self-Refinement (Madaan et al., 2023) have been proposed. While these methods offer improvements, they still struggle to fully address issues such as factual hallucination (Huang et al., 2024). The recently proposed Solo Performance Prompting (SPP) (Wang et al., 2024b) enhances reasoning capabilities in large-scale models such as GPT-4 by simulating multi-perspective thinking through the use of multiple personas that collaborate internally. This approach has shown notable performance in complex tasks, leveraging emergent capabilities in larger models, but its effectiveness remains limited in smaller architectures. To address this limitation, BPP is designed to more effectively improve both knowledge and reasoning abilities across model

scales by dynamically activating task-specific brain-inspired personas.

3 Brain Performance Prompting

To process a diverse range of tasks more efficiently, we propose **Brain Performance Prompting (BPP)**, which aligns LLM operations with human neural systems. BPP dynamically activates personas based on brain regions tailored to the complexity and characteristics of the task, enabling self-collaboration for optimal results. While SPP activates domain-expert or identity personas to elicit cognitive synergy (Wang et al., 2024b), BPP differentiates itself by emulating the collaborative thinking processes of the human brain (Miller, 2000) and leveraging task-specific, brain-region-based personas.

3.1 BPP Mechanism

The core mechanism of BPP follows a process similar to that of SPP, consisting of the following steps: (1) **Persona Identification**: Generate appropriate

brain-region-based personas suited to the characteristics of the input task. **(2) Brainstorming:** Each persona analyzes the task, shares insights, and collaborates to derive an optimal response. **(3) Multi-Persona Iterative Collaboration:** The leader persona, AI Assistant, proposes an initial response, which is iteratively refined and improved through feedback from all personas. This process closely resembles the collaborative interactions among various regions of the human brain.

Figure 2 provides a clear visualization of the operational framework of BPP. The given task involves composing a coherent narrative in response to five specific prompts. During this process, brain-region-based personas contribute information according to their designated roles during the Brainstorming phase, and the final answer is iteratively refined through feedback in the Multi-Persona Iterative Collaboration phase. Notably, the role distribution of each brain region closely aligns with the actual physiological functions of the brain. For example, the Frontal Lobe is responsible for task structuring (Volz et al., 2006), the Temporal Lobe for memory and recall (Eichenbaum et al., 2007), the Occipital Lobe for visual imagination (Kaas et al., 2010), the Hippocampus for context maintenance (Smith and Bulkin, 2014), and the Limbic System for emotional engagement (Morgane et al., 2005). Further details on the neurological grounding of each brain region’s assigned role can be found in Appendix B.

3.2 Dynamic Persona Alignment Based On Task Complexity

BPP dynamically determines the granularity of brain regions activated based on the complexity and characteristics of a task. For knowledge tasks involving information retrieval and memory-related activities, it activates broader brain regions such as the frontal and occipital lobes, mirroring how the human brain integrates and interprets information (Koechlin et al., 2003). In contrast, for reasoning tasks that require detailed problem-solving, BPP activates more granular brain regions such as the Superior Parietal Lobule and Dorsolateral Prefrontal Cortex. These regions are associated with high-level reasoning (Koechlin et al., 2003). Thus, BPP balances broad and specialized activation to handle both simple and complex tasks effectively. A detailed analysis of this dynamic activation mechanism is provided in Appendix B, which presents the task-specific patterns of brain region allocation.

4 Experiments

4.1 Settings

4.1.1 Tasks

To ensure a fair comparison with prior work on SPP, we utilize the same evaluation tasks used in previous research to assess SPP. These tasks include **Trivia Creative Writing**, **Codenames Collaborative**, and **Logic Grid Puzzle**, which are described in detail below (Wang et al., 2024b):

- **Trivia Creative Writing** is a knowledge-intensive task that requires models to integrate and learn information across diverse domains. The model composes a narrative incorporating answers to N quiz questions and is evaluated under two settings (N=5 and 10). Each setting includes 100 instances, with a total of 1,000 quiz questions sampled from the TriviaQA dataset (Joshi et al., 2017). Evaluation is based on accuracy, calculated as the ratio of correctly mentioned answers in the generated narrative to the total number of quiz questions.
- **Codenames Collaborative** is a cooperative task designed to evaluate a model’s knowledge and reasoning abilities. The task involves two roles: Spymaster and Guesser. The Spymaster gives clue words for the target words while avoiding distractors, and the Guesser selects the target words based on the clues provided by the Spymaster and the full word list. The dataset includes 50 instances from the Codenames task in BigBench (Srivastava et al., 2023). Evaluation is based on accuracy, measured as the ratio of words correctly identified by the Guesser to the target words.
- **The Logic Grid Puzzle** is a reasoning-intensive task that evaluates a model’s ability to solve complex, multi-step problems. The dataset contains 200 instances from the Logic Grid Puzzle task in BigBench (Srivastava et al., 2023). Each puzzle involves 2 to 5 houses, each inhabited by a person with specific characteristics. The goal of the task is to determine the correct house assignment for each person based on the given clues. Evaluation is conducted based on accuracy, assigning 1 point for a correct answer and 0 points for an incorrect answer.

Methods	Trivia.C.W (N=5)		Trivia.C.W (N=10)		Codenames.C		Logic.G.Puzzle	
	Score (%)	Δ	Score (%)	Δ	Score (%)	Δ	Score (%)	Δ
Standard	77.4	0%	80.0	0%	77.5	0%	68.0	0%
CoT	70.4	-9.0%	74.2	-7.2%	79.2	+2.2%	70.5	+3.7%
Self Refine	80.4	+3.9%	82.1	+2.6%	71.5	-7.7%	67.5	-0.7%
SPP	79.8	+3.1%	79.1	-1.1%	78.3	+1.0%	69.5	+2.2%
BPP (ours)	80.4	+3.9%	82.4	+3.0%	83.0	+7.1%	72.0	+5.9%

Table 1: The table presents the scores (%) and performance difference (Δ) of each method, evaluated GPT-4o, compared to the baseline of Standard Prompting across four tasks: Trivia Creative Writing (Trivia.C.W) with N=5 and N=10, Codenames Collaborative (Codenames.C), and Logic Grid Puzzle (Logic.G.Puzzle).

4.1.2 Baselines

In the evaluation, we compare the performance of **Standard Prompting**, **Chain-of-Thought (CoT)** (Wei et al., 2022), **Self-Refine** (Madaan et al., 2023), **SPP**, and our proposed **BPP**. Standard Prompting serves as a baseline for direct performance comparison with BPP. For Self-Refine, evaluation is based on one iteration of refinement. The full prompts for each method are provided in Appendix A, and the prompt designs for CoT, Self-Refine, and SPP follow the approach described in Wang et al. (2024b).

4.1.3 Models

The baseline model is **GPT-4o** (OpenAI, 2024a). We also conduct additional experiments to analyze the emergence of cognitive synergy and evaluate BPP across various models, including **GPT-4o-mini** (OpenAI, 2024b), **o1-mini** (OpenAI, 2024c), **GPT-3.5-turbo** (OpenAI, 2023), **Qwen-2.5-7B-Instruct** (Team, 2024), and **Llama-3.1-8B-Instruct** (AI, 2024). Detailed model configurations are provided in Appendix D.

4.2 Results

4.2.1 Performance Evaluation: Standard, CoT, Self Refine, SPP, and BPP

Table 1 presents the performance for each task using Standard Prompting, CoT, Self-Refine, SPP, and BPP on GPT-4o. CoT achieves higher performance than Standard Prompting in tasks requiring reasoning ability, such as Codenames Collaborative and Logic Grid Puzzle, but exhibits a significant performance drop in the knowledge-intensive task of Trivia Creative Writing (N=10). CoT underperforms BPP across all tasks, revealing its limitations in knowledge-intensive and complex scenarios despite improved reasoning. Self-Refine, in contrast to CoT, shows lower performance than Standard Prompting in tasks requiring reasoning

ability but achieves performance close to BPP in the knowledge-intensive task Trivia Creative Writing. However, Self-Refine incurs nearly twice the computational cost of BPP, limiting its practicality. Further details on the cost comparison across methods are provided in Appendix F. SPP outperforms Standard Prompting across all tasks except for Trivia Creative Writing (N=10), where its performance falls behind.

In contrast, BPP consistently outperforms other methods, with its advantage becoming more pronounced as task complexity increases. For instance, in Trivia Creative Writing, BPP continues to improve as N increases from 5 to 10, unlike CoT and SPP, which show performance declines. This highlights BPP’s adaptability to complex tasks by selectively activating task-relevant brain regions. In summary, BPP outperforms not only SPP but also traditional approaches such as CoT and Self-Refine, demonstrating the strength of task-specific brain engagement. The consistency of these findings is further discussed in Appendix E.

4.2.2 Model Comparison and Cognitive Synergy Emergence

To explore the applicability of BPP across various model sizes, we compare six models: GPT-4o, GPT-4o-mini, o1-mini, GPT-3.5-turbo, Qwen-2.5-7B-Instruct, and Llama-3.1-8B-Instruct. Figure 3 summarizes the results of this comparison. As discussed in §4.2.1, GPT-4o shows the strongest cognitive synergy under BPP. In contrast, both SPP and BPP fail to outperform Standard Prompting on o1-mini and GPT-3.5-turbo, suggesting that cognitive synergy may not emerge in smaller models. Interestingly, the results from smaller models such as GPT-4o-mini, Qwen-2.5-7B-Instruct, and Llama-3.1-8B-Instruct are promising. For GPT-4o-mini, both SPP and BPP demonstrate cognitive synergy in knowledge-intensive tasks such as

BPP Performance Comparison Across Several Models

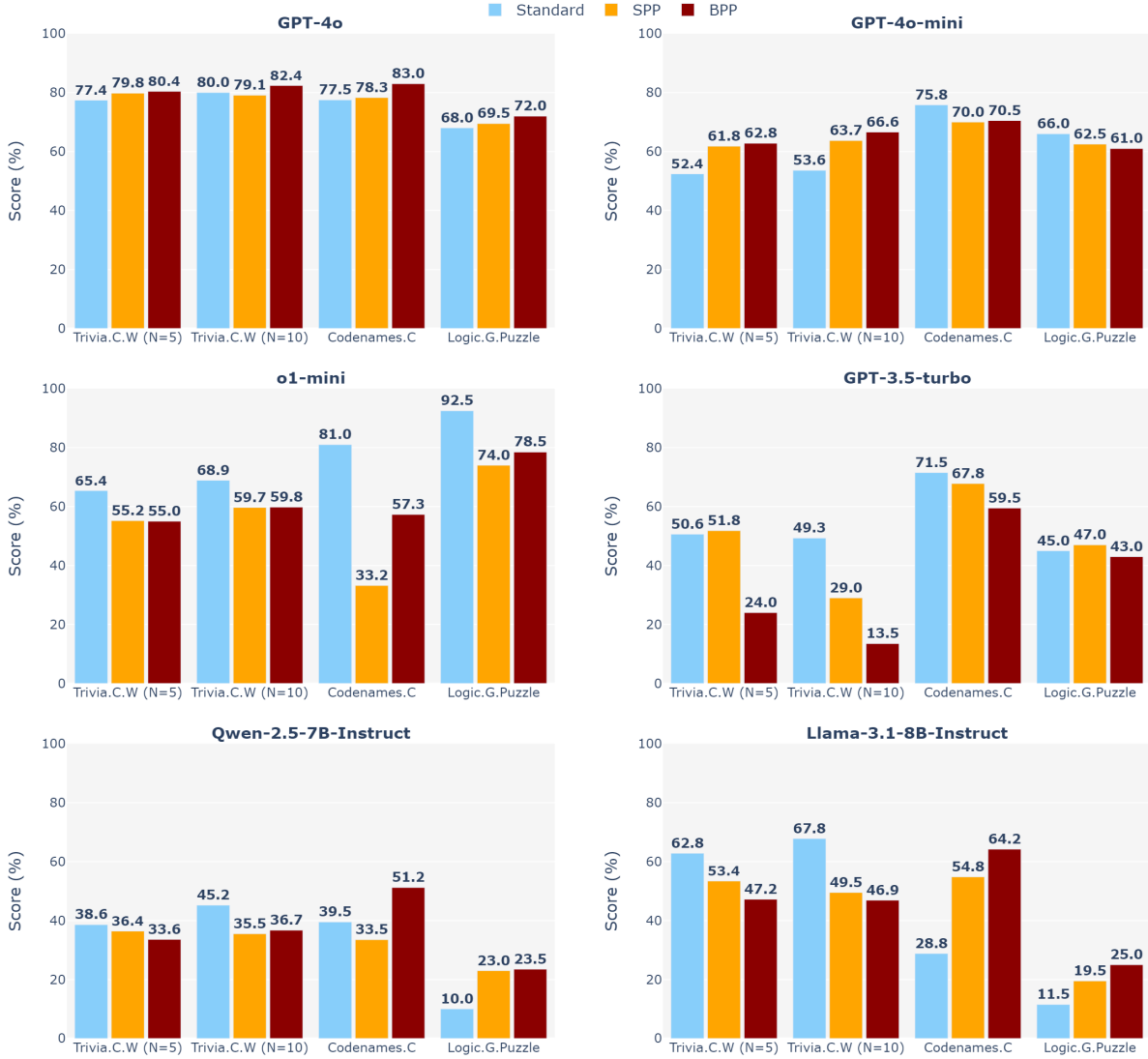


Figure 3: The performance of Standard Prompting, SPP, and BPP across various models (GPT-4o, GPT-4o-mini, o1-mini, GPT-3.5-turbo, Qwen-2.5-7B-Instruct, and Llama-3.1-8B-Instruct) for different tasks: Trivia.C.W (N=5 and N=10), Codenames.C, and Logic.G.Puzzle.

Trivia Creative Writing. For Qwen-2.5-7B-Instruct and Llama-3.1-8B-Instruct, BPP exhibits signs of cognitive synergy in tasks like Codenames Collaborative and Logic Grid Puzzle. As reported by Abacha et al. (2025), GPT-4o-mini is estimated to be approximately an 8B model. Given the similar sizes of Qwen-2.5-7B-Instruct and Llama-3.1-8B-Instruct, these findings suggest that cognitive synergy begins to emerge around this scale, offering insight into the conditions required for its manifestation. These results offer a new perspective on how cognitive synergy manifests and highlight the importance of further research into its relationship with model size.

5 Analysis

This section analyzes how BPP highlights the parallels between LLMs and the human brain, and evaluates the impact of demo configurations.

5.1 Human Brain Parallels in LLM Operations

5.1.1 Persona Transformation: From Domain Expert to Brain Regions

SPP aims to elicit cognitive synergy within LLMs by assigning multiple personas as domain experts or identities (e.g., Math Expert, Ten Year Old Child) suitable for the task, promoting collabora-

tion. In contrast, BPP adopts an approach that mimics the neural system of the human brain, assigning brain regions as multiple personas appropriate for the task and deriving answers through their collaboration. This mechanism goes beyond simple domain expert or identity-based cooperation, more closely replicating the thought process of actual humans when solving problems. This characteristic of BPP suggests similarities between the flow of the human brain’s neural system and the internal workings of LLMs. Just as the fundamental principles of deep learning originated from neural networks that imitate human neurons and synapses (McCulloch and Pitts, 1943), BPP demonstrates that LLMs can achieve superior performance by implementing more human-like thought processes. This is not merely a change in design, but a significant contribution in that it expands the cognitive possibilities of LLMs and more faithfully reflects human problem-solving methods.

5.1.2 Dynamic v.s. Static Brain Region Activation

To explore the importance of dynamically assigning brain regions appropriate to the task, we conduct a supplementary experiment by introducing three derived versions of BPP: **Macro-BPP** (Figure 15), **Meso-BPP** (Figure 16), and **Micro-BPP** (Figure 17)

- **Macro-BPP** modifies the prompt so that personas are fixed to the largest regions of the brain, such as "Frontal Lobe" or "Temporal Lobe".
- **Meso-BPP** modifies the prompt so that personas are fixed to moderately subdivided regions, such as "Superior Parietal Lobule" or "Dorsolateral Prefrontal Cortex".
- **Micro-BPP** modifies the prompt so that personas are fixed to more finely subdivided regions, such as "Left Dorsolateral Prefrontal Cortex" or "Right Inferior Parietal Lobule".

Figure 4 presents a comparison of BPP with Macro-BPP, Meso-BPP, and Micro-BPP, evaluated on GPT-4o. This comparison provides the following key insights: (1) Macro-BPP demonstrates strong performance in tasks requiring knowledge ability, such as Trivia Creative Writing and Codenames Collaborative. However, its performance is limited in reasoning-intensive task like Logic Grid Puzzle, where more nuanced problem-solving

is necessary. This suggests that the largest brain regions are more suitable for knowledge-based tasks. (2) Meso-BPP shows lower performance than Macro-BPP in tasks requiring knowledge ability, but excels in reasoning-intensive task. This indicates that subdivided brain regions are more appropriate for reasoning-intensive task. (3) Micro-BPP generally consistently underperforms compared to other methods, even falling below Standard prompting in some tasks. This implies that excessively subdivided regions may not be effective for task requirements and could lead to performance degradation. (4) BPP outperforms Standard prompting across all tasks and demonstrates stable and superior overall performance compared to Macro-BPP, Meso-BPP, and Micro-BPP. This indicates that dynamically assigning brain regions is more effective in responding to task-specific demands than fixed assignment methods.

These results suggest that mimicking human neural pathways through dynamic brain region activation not only improves task-specific performance but also brings LLM behavior closer to human cognitive processes. By selectively engaging relevant regions, BPP bridges insights from neuroscience and machine learning, offering a task-adaptive framework that enhances both efficiency and cognitive synergy.

5.1.3 Amplifying the Cognitive Horizon: Insights from Human Brain Neural System

As shown in §4.2.1, SPP’s cognitive synergy diminishes in complex task like Trivia Creative Writing (N=10) as the number of questions increases. This indicates that SPP’s approach has limitations in effectively manifesting advanced cognitive synergy. In contrast, BPP exhibits enhanced cognitive synergy by applying mechanisms similar to the human brain’s neural system, achieving excellent performance even in complex tasks. This suggests that BPP effectively mimics the collaborative thinking processes among various regions of the human brain, confirming its ability to adapt flexibly to the requirements of complex tasks. Additionally, the analysis of inter-regional communication presented in Appendix B shows that the brain region-based personas utilized in BPP align well with the neurological functions of their respective brain areas. This supports the notion that BPP effectively reflects the flow of cognitive processes in the human brain, contributing to meeting task-specific

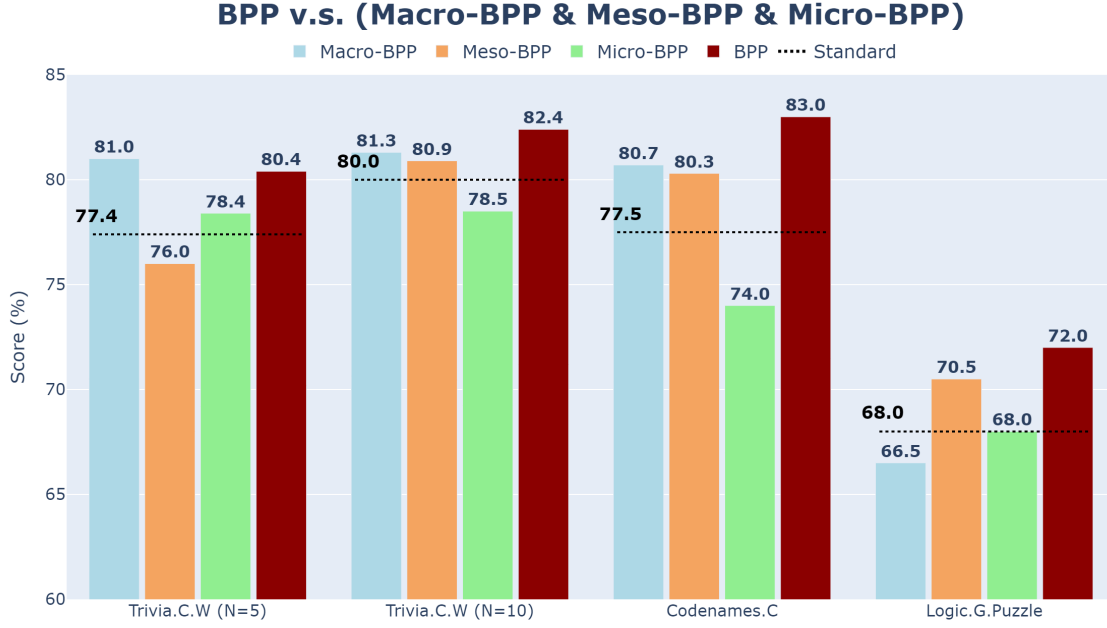


Figure 4: BPP demonstrates consistently strong performance across all tasks by dynamically activating task-specific brain regions, unlike static approaches (Macro-BPP, Meso-BPP, Micro-BPP).

demands. BPP’s approach goes beyond simply achieving high performance, demonstrating the potential to expand cognitive synergy by reflecting the mechanisms of the human neural system. This enables outstanding performance even in complex tasks and provides concrete design directions for future LLM architectures.

5.2 Effect of Demo Configuration on Task Performance

We conduct an additional study to evaluate how the composition of BPP’s demonstrations affects task performance. We utilized MMLU-Pro (Wang et al., 2024a) to generate new demonstrations for the experiment. The detailed analysis of this experiment can be found in Appendix G. The results provide two key insights: First, task-specific demo examples tend to improve performance on the corresponding task. For instance, knowledge-oriented demo examples significantly contribute to performance improvement in knowledge task like Trivia Creative Writing. However, there are exceptional cases in reasoning-intensive task such as Logic Grid Puzzle, where task-specific demo examples do not always lead to performance improvements. Second, demo compositions that include diverse types of examples generally outperform those using only task-specific examples. This suggests the

importance of demo compositions that can support various cognitive demands. However, an exception exists in the Logic Grid Puzzle, where a composition including diverse types of demonstration examples showed lower performance than one composed only of task-specific demo examples.

These findings emphasize the need for a systematic approach to demonstration example composition in BPP design, considering both task-specific optimization and generality. Particularly for reasoning-intensive task like Logic Grid Puzzle, where specific strategies may not guarantee optimal results, it is critical to carefully analyze task characteristics to design effective demo compositions.

6 Conclusion

We propose Brain Performance Prompting (BPP), a brain-inspired method that dynamically activates task-specific personas based on human neural pathways. BPP significantly improves both knowledge and reasoning abilities, consistently outperforming existing approaches across diverse tasks. These results reveal a strong alignment between the internal mechanisms of LLMs and the functional organization of human neural pathways, suggesting that cognitively inspired designs can meaningfully enhance LLM performance.

7 Limitations

This paper highlights the strengths of Brain Performance Prompting (BPP) in enhancing cognitive synergy and achieving task-specific performance improvements, while acknowledging several limitations. First, although this study observed partial cognitive synergy in smaller models such as GPT-4o-mini, the mechanisms and factors underlying its full emergence remain unclear, warranting further systematic investigation. Second, BPP currently relies on predefined prompt structures for brain region persona assignment. While effective, these prompts could be further optimized by adapting them to specific tasks or model architectures. Third, BPP draws inspiration from human neural systems but simplifies their complex interactions and multimodal processing capabilities. Incorporating more detailed neural mechanisms, such as inter-region communication and sensory integration, could enhance the alignment between LLMs and human cognition. Addressing these limitations will not only enhance BPP’s performance and applicability but also contribute to developing new LLM design paradigms inspired by human neural mechanisms.

References

Asma Ben Abacha, Wen wai Yim, Yujuan Fu, Zhaoyi Sun, Meliha Yetisgen, Fei Xia, and Thomas Lin. 2025. *Medec: A benchmark for medical error detection and correction in clinical notes*. *Preprint*, arXiv:2412.19260.

Meta AI. 2024. Llama 3.1 model card. https://github.com/meta-llama/llama-models/blob/main/models/llama3_1/MODEL_CARD.md.

Khai Loong Aw, Syrielle Montariol, Badr AlKhamissi, Martin Schrimpf, and Antoine Bosselut. 2024. *Instruction-tuning aligns LLMs to the human brain*. In *First Conference on Language Modeling*.

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*. *Preprint*, arXiv:2303.12712.

Ed Bullmore and Olaf Sporns. 2009. *Complex brain networks: graph theoretical analysis of structural and functional systems*. *Nature Reviews Neuroscience*, 10(3):186–198.

Petru L Curşeu, Nicoleta Meslec, Helen Pluut, and Gerardus JM Lucas. 2015. Cognitive synergy in groups

and group-to-individual transfer of decision-making competencies. *Frontiers in psychology*, 6:1375.

Gustavo Deco, Viktor K. Jirsa, Peter A. Robinson, Michael Breakspear, and Karl Friston. 2008. *The dynamic brain: from spiking neurons to neural masses and cortical fields*. *PLoS Computational Biology*, 4(8):e1000092.

Howard Eichenbaum, Andrew P. Yonelinas, and Charan Ranganath. 2007. *The medial temporal lobe and recognition memory*. *Annual review of neuroscience*, 30:123–152.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2024. *A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions*. *ACM Trans. Inf. Syst.* Just Accepted.

Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer. 2017. *TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension*. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1601–1611, Vancouver, Canada. Association for Computational Linguistics.

Amanda Kaas, Sarah Weigelt, Alard Roebroek, Achim Kohler, and Lars Muckli. 2010. *Imagery of a moving object: the role of occipital cortex and human mt/v5+*. *NeuroImage*, 49(1):794–804.

Etienne Koechlin, Chrystele Ody, and Frédérique Kounieher. 2003. *The architecture of cognitive control in the human prefrontal cortex*. *Science (New York, N.Y.)*, 302(5648):1181–1185.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. *Self-refine: Iterative refinement with self-feedback*. In *Advances in Neural Information Processing Systems*, volume 36, pages 46534–46594. Curran Associates, Inc.

Richard E Mayer. 1998. Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional science*, 26(1):49–63.

Warren S. McCulloch and Walter Pitts. 1943. *A logical calculus of the ideas immanent in nervous activity*. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.

Gabriele Merlin and Mariya Toneva. 2024. *Language models and brains align due to more than next-word prediction and word-level information*. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18431–18454, Miami, Florida, USA. Association for Computational Linguistics.

662	Earl K. Miller. 2000. The prefrontal cortex and cognitive control . <i>Nature Reviews Neuroscience</i> , 1(1):59–65.	712
663		713
664		714
665	Peter J. Morgane, Janina R. Galler, and David J. Moller. 2005. A review of systems and networks of the limbic forebrain/limbic midbrain . <i>Progress in Neurobiology</i> , 75(2):143–160.	715
666		716
667		717
668		718
669	OpenAI. 2023. Openai models documentation. https://platform.openai.com/docs/models/gpt-3.5-turbo .	719
670		720
671		721
672	OpenAI. 2024a. Gpt-4o system card . <i>Preprint</i> , arXiv:2410.21276.	722
673		723
674	OpenAI. 2024b. Gpt-4o-mini: Advancing cost-efficient intelligence. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/ .	724
675		725
676		726
677		727
678	OpenAI. 2024c. Openai-o1-mini: Advancing cost-efficient reasoning. https://openai.com/index/openai-o1-mini-advancing-cost-efficient-reasoning/ .	728
679		
680		
681		
682	Anthony Pellegrini. 2009. <i>The Role of Play in Human Development</i> . Oxford University Press.	
683		
684	Jean Piaget. 1954. <i>The Construction Of Reality In The Child</i> . Routledge.	
685		
686	David M. Smith and David A. Bulkin. 2014. The form and function of hippocampal context representations . <i>Neuroscience and Biobehavioral Reviews</i> , 40:52–61.	
687		
688		
689	Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, and et al Adrià Garriga-Alonso. 2023. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models . <i>Preprint</i> , arXiv:2206.04615.	
690		
691		
692		
693		
694		
695		
696	Qwen Team. 2024. Introducing qwen2.5. https://qwenlm.github.io/blog/qwen2.5/ .	
697		
698	Mycal Tucker and Greta Tuckute. 2023. Increasing brain-LLM alignment via information-theoretic compression . In <i>UniReps: the First Workshop on Unifying Representations in Neural Models</i> .	
699		
700		
701		
702	Kirsten G. Volz, Ricarda I. Schubotz, and D. Yves von Cramon. 2006. Decision-making and the frontal lobes . <i>Current opinion in neurology</i> , 19(4):401–406.	
703		
704		
705	Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyang Jiang, Tianle Li, Max Ku, Kai Wang, Alex Zhuang, Rongqi Fan, Xiang Yue, and Wenhui Chen. 2024a. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark . <i>Preprint</i> , arXiv:2406.01574.	
706		
707		
708		
709		
710		
711		
	Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao Ge, Furu Wei, and Heng Ji. 2024b. Unleashing the emergent cognitive synergy in large language models: A task-solving agent through multi-persona self-collaboration . In <i>Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)</i> , pages 257–279, Mexico City, Mexico. Association for Computational Linguistics.	
	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 <i>Advances in Neural Information Processing Systems</i> , volume 35, pages 24824–24837. Curran Associates, Inc.	
	A Full Prompts for Each Method	
	Figure 12 shows the CoT prompting and Self-Refine prompting, and Figure 13 shows the SPP prompting. Additionally, Figures 14, 15, 16, and 17 show the prompting methods for BPP, Macro-BPP, Meso-BPP, and Micro-BPP, respectively.	
	B Alignment of Brain Region Roles with Neural Functions	
	BPP assigns personas based on the physiological functions of specific brain regions, reflecting the neural pathways of the human brain. This design enables brain-region personas to collaboratively process task-relevant information, mirroring human neural cooperation. By aligning personas with task-relevant brain functions, BPP enhances cognitive synergy in a biologically inspired manner. To evaluate this mechanism, we analyze how each persona contributed to problem-solving in the example task in Figure 2 and assess the alignment between persona behavior and the physiological functions of their respective brain areas. Figure 9 and Figure 10 summarize the dialogues and functions of each brain-region persona. A comprehensive analysis of these roles and their alignment with established neurological functions is presented below.	
	• Frontal Lobe: The Frontal Lobe persona organizes the task structure and guides its direction, aligning with its known roles in planning, decision-making, and executive control (Volz et al., 2006).	
	• Temporal Lobe: The Temporal Lobe persona retrieves relevant information from memory, reflecting its core functions in recall and	

language comprehension (Eichenbaum et al., 2007).

- Occipital Lobe: The Occipital Lobe persona generates vivid visual imagery, reflecting its role in processing visual information and supporting creative visualization (Kaas et al., 2010).
- Hippocampus: The Hippocampus persona maintains contextual consistency by integrating inputs from other personas, aligning with its function in memory organization and contextual binding (Smith and Bulkin, 2014).
- Limbic System: The Limbic System persona enhances emotional engagement, reflecting its role in regulating emotions and strengthening motivation (Morgane et al., 2005).

The interactions among personas in this example demonstrate a high degree of alignment with the physiological functions of their corresponding brain regions. Effective collaboration, driven by iterative feedback, reflected each persona’s alignment with its neurological role. This mirrors the cooperative dynamics of functional brain networks. These findings validate BPP’s neurologically grounded design and its ability to enhance task-specific performance. Overall, BPP effectively mirrors the cognitive processing flow of the human brain.

C Task-Specific Brain Region Activation in BPP: A Quantitative Visualization

To verify whether BPP dynamically assigns brain regions based on task type and complexity, we analyzed its output on the Codenames Collaborative task. Because it requires both knowledge and reasoning abilities, this task serves as a suitable testbed for evaluating BPP’s dynamic brain region allocation. The Codenames Collaborative task consists of two roles: Spymaster and Guesser. The Spymaster must understand relationships between target and distractor words and provide effective hints, relying primarily on domain knowledge. In contrast, the Guesser uses logical reasoning to analyze the Spymaster’s hints and the full word list to identify the target words.

Figure 5 presents the activation frequencies of brain regions for the Spymaster and Guesser roles, revealing distinct usage patterns. In the Spymaster role, broad regions such as the Frontal, Temporal,

and Occipital Lobes were activated in nearly all 50 instances, whereas more granular regions like the Superior Parietal Lobule and Dorsolateral Prefrontal Cortex appeared only 9 and 1 times, respectively. This suggests that the knowledge-intensive tasks performed by the Spymaster align well with the activation of broad brain regions. In contrast, the Guesser role continued to engage broad regions, though their activation frequency was lower than in the Spymaster role. For instance, Frontal Lobe activations decreased from 49 to 25 instances, and Occipital Lobe activations dropped from 44 to 36 instances. At the same time, granular regions like the Superior Parietal Lobule and Dorsolateral Prefrontal Cortex, which were rarely activated for the Spymaster role, increased significantly to 26 and 20 instances, respectively. Additionally, granular regions such as the Olfactory Bulb and Primary Visual Cortex, previously inactive, were newly recruited for the Guesser role. These findings demonstrate that the Guesser’s reasoning tasks rely not only on broad brain regions but also on the increased engagement of granular regions.

In conclusion, BPP dynamically adjusts brain region activation patterns according to the roles of Spymaster and Guesser, effectively aligning with the cognitive demands of each role. This adaptability demonstrates BPP’s capacity to meet task-specific requirements while maintaining efficiency across diverse tasks.

D Model Configurations

The results in §4.2 are obtained from four models: GPT-4o, GPT-4o-mini, o1-mini, and GPT-3.5-turbo. To ensure consistent responses, GPT-4o, GPT-4o-mini, and GPT-3.5-turbo are configured with a temperature of 0.0 and a top-p of 1.0. However, the models do not reliably generate identical outputs, even with these settings. To address this, we conduct an additional experiment to evaluate output consistency. We measure response variability across repeated generations for the same input and use it to quantitatively assess model stability. Detailed experimental settings and results can be found in Appendix E.

E Consistency Analysis

To evaluate the consistency of GPT-4o’s outputs, we conducted three runs for each of the following tasks: Trivia Creative Writing (N=5), Trivia Creative Writing (N=10), Codenames Collaborative,

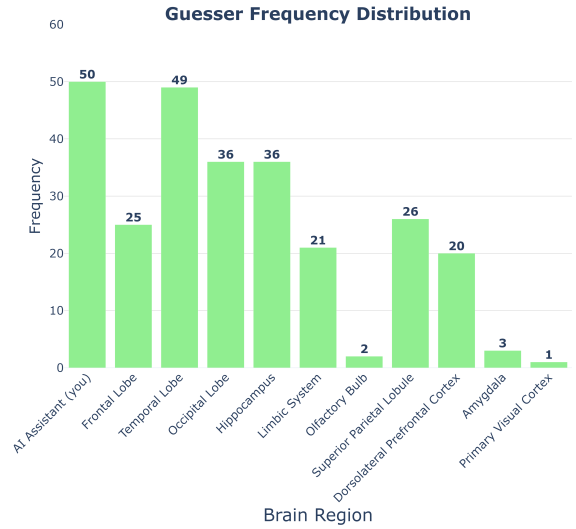
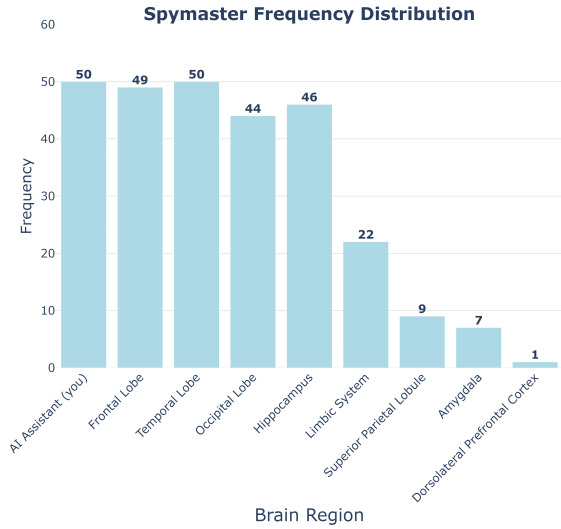


Figure 5: Brain region activation frequencies in the Codenames Collaborative task by role. The Spymaster mainly activated broad regions (Frontal, Temporal Lobes), while the Guesser engaged more granular regions (Superior Parietal Lobule, Dorsolateral Prefrontal Cortex) and additional areas, demonstrating BPP’s role-specific adaptability.

and Logic Grid Puzzle. Standard prompting, SPP, and BPP were applied to each task. For each set of experiments, we calculated the mean and the standard deviation of the performance to assess the consistency of the model’s outputs. As shown in Table 2, the standard deviation across all experiments was sufficiently low. This indicates that GPT-4o generates stable and consistent outputs under identical conditions, thereby reinforcing the reliability of the key experimental results reported in §4.2.1. Figure 11 summarizes the results from Table 2, clearly demonstrating both the enhanced cognitive synergy achieved by BPP and its high consistency.

F Cost Comparison Across Methods

In §4.2.1, it was observed that there is no significant difference in performance between Self-Refine and BPP in the Trivia Creative Writing task. This finding suggests that, at least for knowledge-related tasks, existing methods can achieve high performance. However, a comprehensive evaluation must also consider cost efficiency alongside performance. To this end, we calculated the execution time and cost for each method, as summarized in Figure 6, which presents the time and cost for applying Standard prompting, CoT, Self-Refine, SPP, and BPP to the Trivia Creative Writing (N=5) task. The analysis reveals that Self-Refine incurs nearly twice the cost of BPP, highlighting its lower cost efficiency. These results demonstrate that BPP is not only more effective but also more cost-efficient.

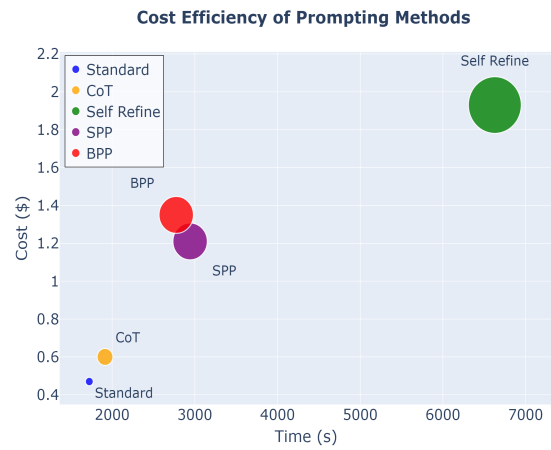


Figure 6: A comparison of computation time and cost for five prompting methods (Standard, CoT, Self-Refine, SPP, and BPP) on the Trivia Creative Writing (N=5) task.

G Detailed Analysis of Demo Configurations for Task Performance

To assess the impact of BPP’s demo composition on task performance, we conducted an additional experiment comparing five strategies: **BPP-W-R-Demo** (single reasoning-oriented), **BPP-W-K-Demo** (single knowledge-oriented), **BPP-Two-R-Demo** (two reasoning-oriented), **BPP-Two-K-Demo** (two knowledge-oriented), and **BPP (Ours)** (diverse demos). Using MMLU-Pro (Wang et al., 2024a), we designed problems tailored to the characteristics of each task and created new demos. Full prompts for each method are shown in Figures

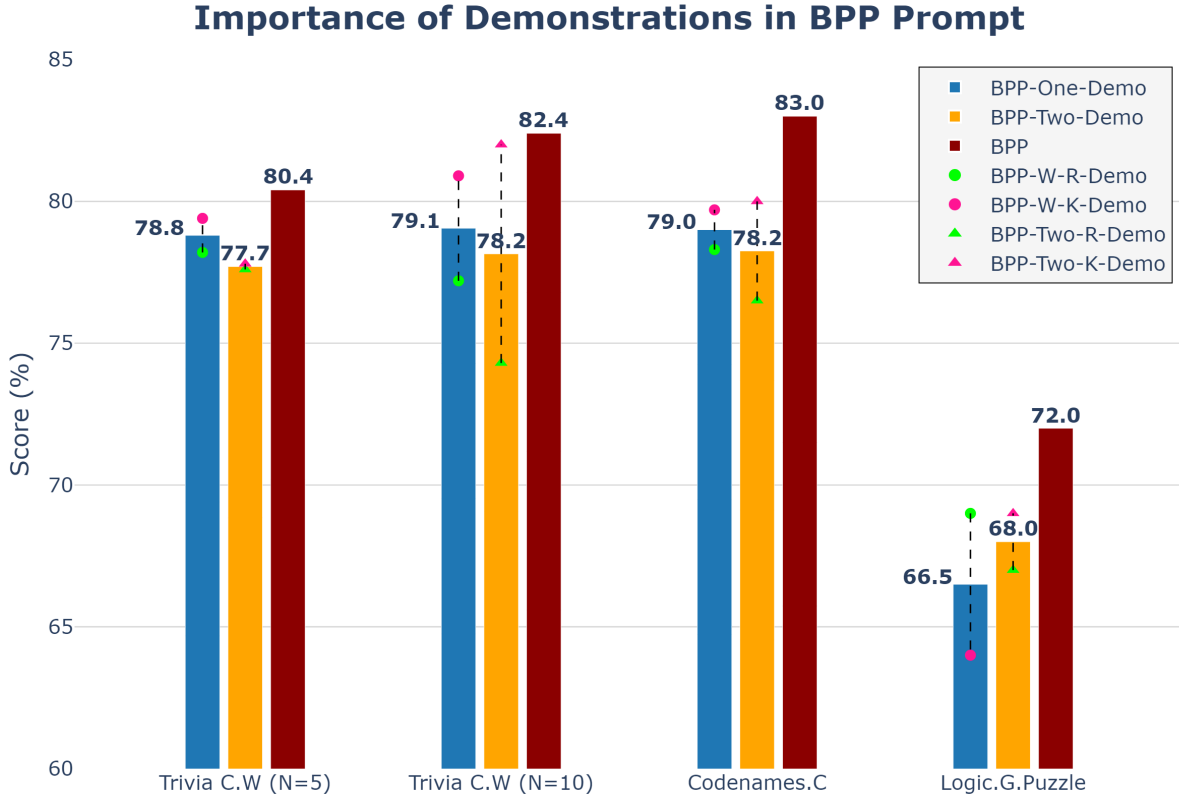


Figure 7: A comparison of different demo compositions highlights their impact on task performance. The results indicate that both the type and number of demos significantly affect performance, and BPP, which integrates diverse demo types, consistently outperforms other configurations across all tasks.

18, 19, 20, and 21. The demo composition strategies were evaluated on four tasks: Trivia Creative Writing (N=5 and N=10), Codenames Collaborative, and Logic Grid Puzzle. A summary of the experimental results is presented in Figure 7.

G.1 Results

The results indicate that task-specific demo compositions improve task performance, whereas mismatched ones lead to decreased performance. For instance, in Trivia Creative Writing (N=5), a knowledge-intensive task, BPP-W-K-Demo outperforms BPP-W-R-Demo. In contrast, in the Logic Grid Puzzle, a reasoning-intensive task, BPP-W-R-Demo outperforms BPP-W-K-Demo. Similar trends are observed when two demo examples are used. In tasks requiring knowledge ability, BPP-Two-K-Demo shows better performance than BPP-Two-R-Demo. However, an exception was observed in the Logic Grid Puzzle, where BPP-Two-K-Demo unexpectedly outperformed BPP-Two-R-Demo, contrary to the general trend. Furthermore, increasing the number of knowledge-oriented demos generally improves performance in tasks re-

quiring knowledge ability. For instance, BPP-Two-K-Demo outperforms BPP-W-K-Demo in Trivia Creative Writing (N=10). An exception is observed in the N=5 variant, where BPP-W-K-Demo performs better. In reasoning tasks like the Logic Grid Puzzle, increasing the number of demos does not consistently enhance performance. A detailed analysis of these exceptions is provided in Appendix G.2. BPP (Ours), which includes diverse types of demos, consistently outperforms task-specific variants across all tasks. This suggests that incorporating diverse types of demos yields better overall performance than task-specific compositions.

As a result, this experiment yields two main conclusions. First, task-specific demo compositions generally contribute to improved performance; however, exceptions in tasks like the Logic Grid Puzzle, which requires reasoning ability, reveal their limitations. Second, incorporating diverse types of demos achieves consistently higher performance, indicating that BPP effectively addresses various cognitive demands. These findings highlight the importance of systematically design-

ing demos that balance task-specific needs and generalizability within the BPP framework.

G.2 Analysis of Exception Cases

As shown in §G.1, task-specific demo compositions generally improve performance, while notable exceptions were observed in specific tasks. These exceptions are summarized as two main cases. First, in Trivia Creative Writing (N=5), BPP-W-K-Demo outperforms BPP-Two-K-Demo, despite using more knowledge-oriented demos. Second, in the Logic Grid Puzzle, BPP-Two-K-Demo, which uses knowledge-oriented demos instead of reasoning-oriented ones, achieves better performance. This suggests that increasing reasoning-oriented demos does not always result in performance gains. This subsection aims to analyze these exceptions in depth, discuss their implications for each task, and better understand the impact of demo composition strategies on task performance.

G.2.1 Analysis of the First Exception Case: Trivia Creative Writing (N=5)

An unexpected result was observed in the Trivia Creative Writing (N=5) task, where using more knowledge-oriented demos led to a decline in performance. Specifically, BPP-W-K-Demo achieved higher performance than BPP-Two-K-Demo. This finding contrasts with the general trend that adding more knowledge-oriented demos improves performance in knowledge-intensive tasks.

To better understand this anomaly, we examine the characteristics of the Trivia Creative Writing (N=5) task. With its limited number of questions and a writing component, this task is difficult to classify as a typical knowledge-intensive task. Under the N=5 setting, the task’s relatively low complexity increases the likelihood that additional knowledge-oriented demos introduce excessive information, potentially leading to confusion during final response generation. This excessive information can cause cognitive overload during the writing process, leading to the omission of key answers or a loss of narrative coherence.

To verify this, we extracted the answers from the Dialogue phase and the Final Answer phase (referred to as Answers-in-dialogue and Answers-in-final) and compared the Dialogue-phase accuracy with the recall between the two. Note that precision is always 1, since Answers-in-dialogue is a superset of Answers-in-final; thus, only recall is used in the analysis. The results are presented

in Figure 8. The analysis showed that Dialogue-phase accuracy remained constant regardless of the number of knowledge-oriented demonstrations, indicating that additional demos had minimal impact on answer identification. However, the recall between Answers-in-dialogue and Answers-in-final decreased as the number of knowledge-oriented demonstrations increased. This suggests that while correct answers were identified during the Dialogue phase, some were omitted in the final response generation. In other words, in the N=5 setting, the limited number of questions lowers task complexity, resulting in redundant information from the additional demonstrations. In contrast, in Trivia Creative Writing (N=10), the increased number of questions raises task complexity, enabling better alignment with knowledge-oriented demonstrations. As a result, the additional demonstrations positively influence the writing process, contributing to performance improvement, as shown in Figure 7.

In conclusion, knowledge-oriented demonstrations are effective in enhancing knowledge ability. While the N=5 setting showed an exception where excessive information caused issues, this can be attributed to design limitations and does not undermine the overall validity of the analysis discussed in 5.2.

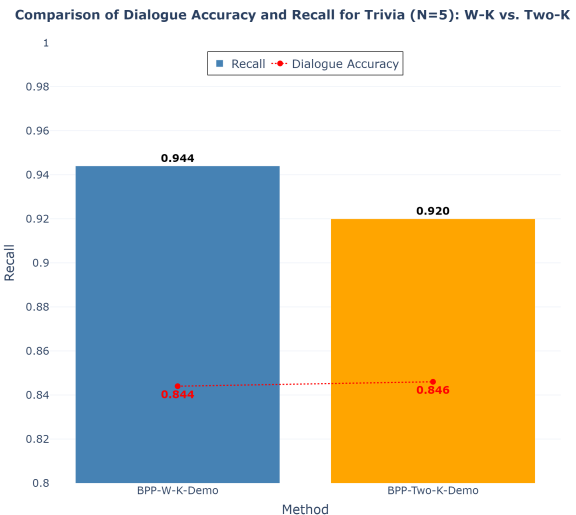


Figure 8: While the accuracy in the dialogue phase remains relatively consistent, we observe a decrease in recall as it transitions to the final answer generation phase.

G.2.2 Analysis of the Second Exception Case: Logic Grid Puzzle

Two exceptional phenomena were observed in the Logic Grid Puzzle task. First, increasing the number of reasoning-oriented demonstrations did not improve performance and, in some cases, even led to a decline. Second, the configuration including knowledge-oriented demonstrations (BPP-Two-K-Demo) outperformed the reasoning-oriented configuration (BPP-Two-R-Demo). These results challenge the common assumption that reasoning-oriented demonstrations enhance performance in reasoning-intensive tasks. A possible explanation is that excessive reasoning-oriented demonstrations may hinder the understanding of initial conditions and introduce unnecessary complexity during problem-solving, ultimately reducing performance. In contrast, knowledge-oriented demonstrations may help clarify initial conditions and the problem structure, reducing confusion during complex reasoning and improving performance. However, this interpretation remains hypothetical and lacks empirical validation. Further experiments are necessary to verify these hypotheses, particularly systematic research examining how demonstration composition affects the understanding of initial conditions in reasoning-intensive tasks like the Logic Grid Puzzle.

H Usage of AI assistants in writing

We used GPT-4o solely to check whether the grammar was correct.

Brainstorming

Frontal Lobe: “Let's start by identifying the answers to the questions, as they will form the basis of our story. We need to incorporate these answers into a coherent narrative about Pikachu.”

Temporal Lobe: “I can help recall the answers to the questions:

1. Who directed the classic 30s western Stagecoach? - John Ford
2. Dave Gilmore and Roger Waters were in which rock group? - Pink Floyd
3. Which highway was Revisited in a classic 60s album by Bob Dylan? - Highway 61
4. Which was the only eastern bloc country to participate in the 1984 LA Olympics? – Romania
5. Which 90s sci-fi series with James Belushi was based on Bruce Wagner's comic strip of the same name? - Wild Palms”

Hippocampus: “Great, now that we have the answers, let's brainstorm how to weave them into a story about Pikachu. We should think about how these elements can be connected in a fun and engaging way.”

Occipital Lobe: “Visually, I imagine Pikachu embarking on an adventure that ties these elements together. Perhaps Pikachu could be on a quest that involves traveling along Highway 61, meeting characters inspired by Pink Floyd, and encountering a mysterious event related to the 1984 LA Olympics.”

Limbic System: “Let's make sure the story is engaging and has a sense of wonder. Pikachu's journey should be filled with excitement and discovery, appealing to fans of the character.”

Figure 9: A figure visualizing the Brainstorming phase as a conversation among brain region personas. Each persona contributes to task-solving based on its neurological functions, representing how each brain region performs roles consistent with its actual physiological functions and operates collaboratively.

Multi-Persona Iterative Collaboration

Frontal Lobe: “The story incorporates all the elements and maintains a coherent narrative. It captures the essence of adventure and discovery.”

Temporal Lobe: “The references to the answers are seamlessly integrated, making the story both informative and entertaining.”

Hippocampus: “The story reflects a sense of nostalgia and adventure, reminiscent of classic tales.”

Occipital Lobe: “The imagery is vivid, with Pikachu's journey along Highway 61 and the festival atmosphere.”

Limbic System: “The story is engaging and captures the spirit of Pikachu's adventurous nature.”

Figure 10: An example dialogue from the Multi-Persona Iterative Feedback phase, demonstrating each brain region persona providing feedback on the AI Assistant's draft based on its specific role. This process highlights how the personas collaboratively refine the draft through their unique contributions, effectively reflecting the cooperative thinking mechanism of the human neural system.

Task	Method	Run1	Run2	Run3	Mean (std)
Trivia Creative Writing (N=5)	Standard	77.4	77.6	78.4	77.8 ± 0.5
	SPP	79.8	80.4	79.0	79.7 ± 0.7
	BPP	80.4	80.4	79.6	80.1 ± 0.5
Trivia Creative Writing (N=10)	Standard	80.0	80.5	79.3	79.9 ± 0.6
	SPP	79.1	79.8	79.0	79.3 ± 0.4
	BPP	82.4	82.7	82.0	82.4 ± 0.4
Codenames Collaborative	Standard	77.5	79.8	78.5	78.6 ± 1.2
	SPP	78.3	78.0	80.2	78.8 ± 1.2
	BPP	83.0	83.7	81.0	82.6 ± 1.4
Logic Grid Puzzle	Standard	68.0	69.5	72.0	69.8 ± 2.0
	SPP	69.5	67.0	69.5	68.7 ± 1.4
	BPP	72.0	71.5	69.5	71.0 ± 1.3

Table 2: The consistency of GPT-4o was analyzed across four tasks: Trivia Creative Writing (N=5), Trivia Creative Writing (N=10), Codenames Collaborative, and Logic Grid Puzzle, using Standard Prompting, SPP, and BPP methods. The results of three repeated experiments for each task and method are presented, including the mean performance and standard deviation (Mean $\pm$ Std). The table demonstrates that GPT-4o generates stable and consistent outputs across various tasks and methods.

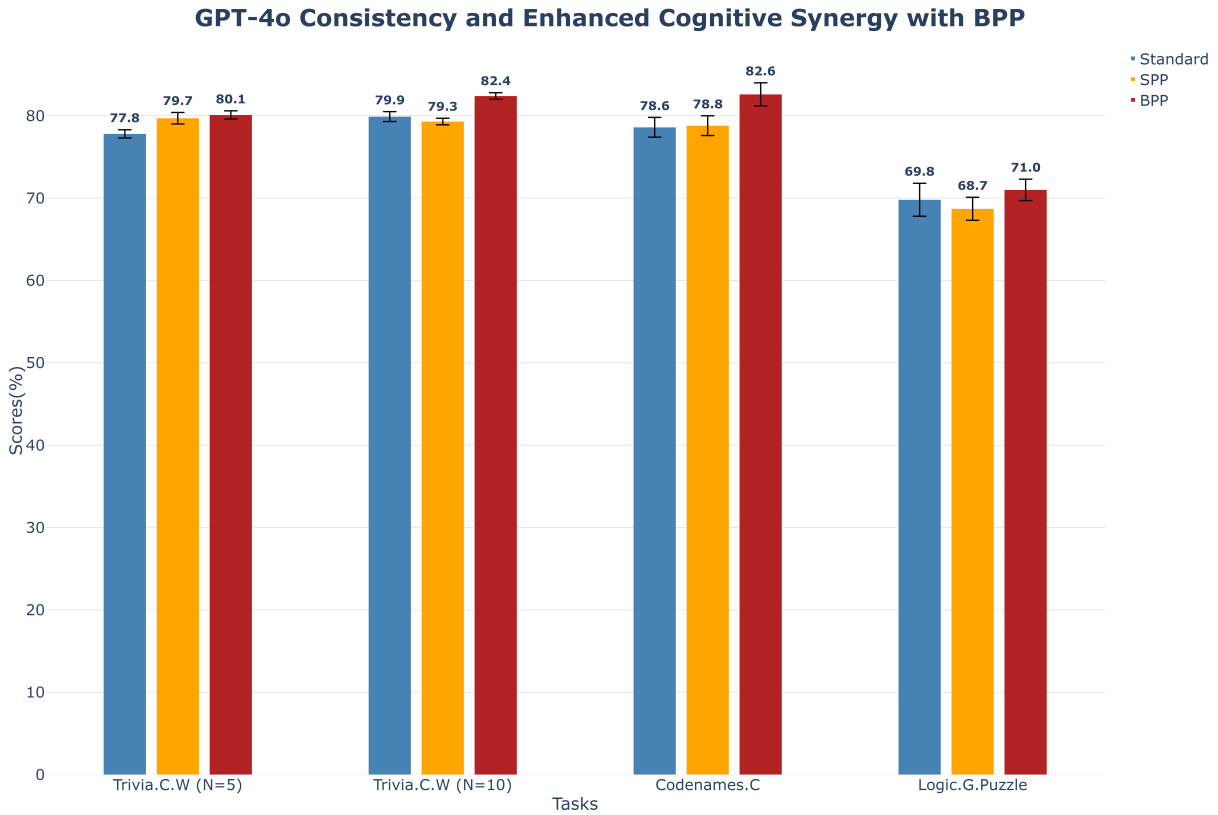
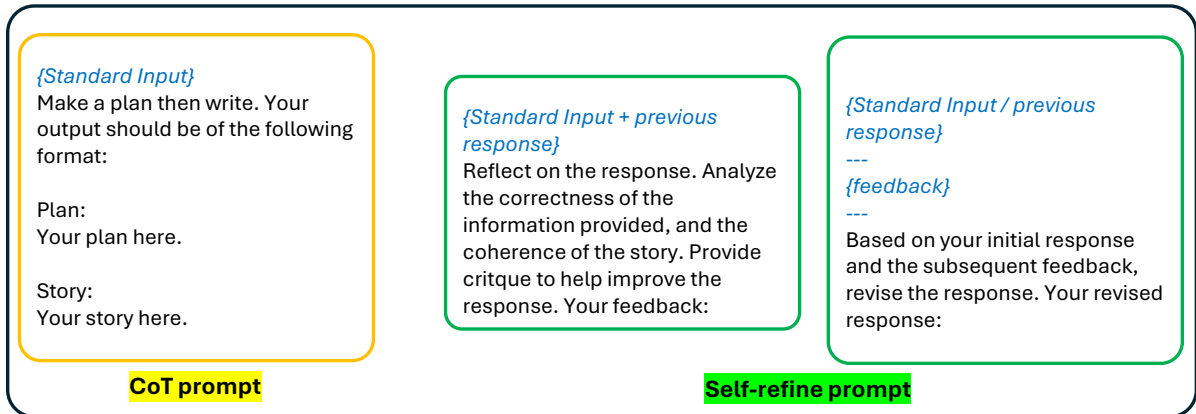
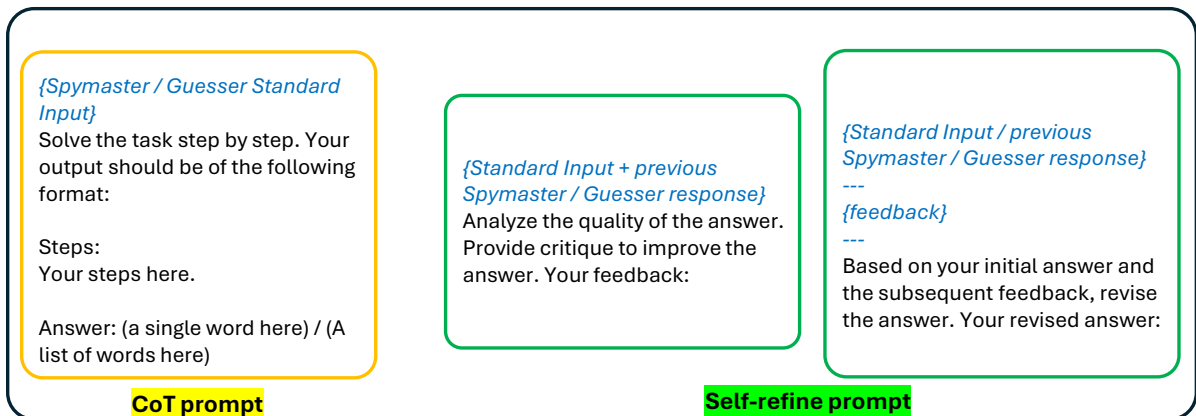


Figure 11: A figure visualizing the consistency analysis results of GPT-4o. It illustrates the mean performance and standard deviation (Mean $\pm$ Std) for each method, highlighting that BPP achieves performance improvements through more pronounced cognitive synergy while maintaining high consistency compared to other methods.

Trivia Creative Writing



Codenames Collaborative



Logic Grid Puzzle

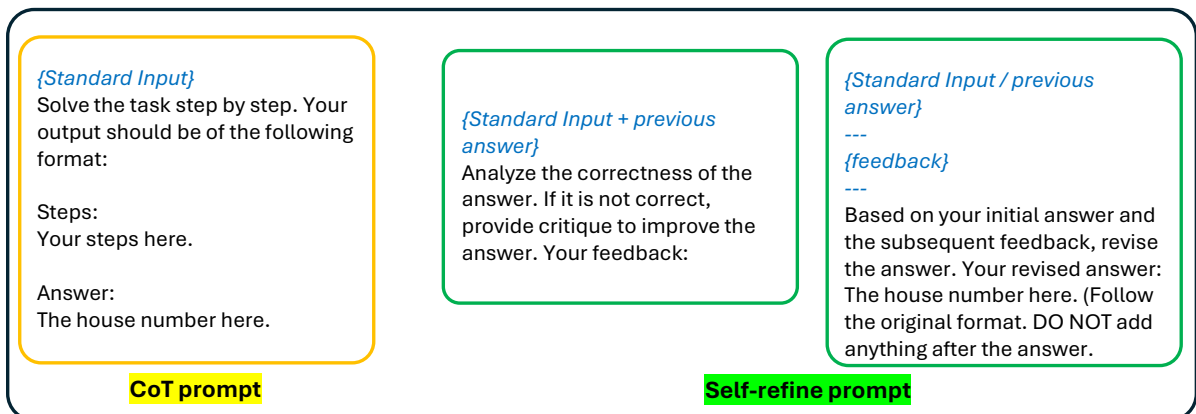


Figure 12: CoT & Self-refine prompts

SPP Prompt

When faced with a task, begin by identifying the participants who will contribute to solving the task. Then, initiate a multi-round collaboration process until a final solution is reached. The participants will give critical comments and detailed suggestions whenever necessary.

Here are some examples:

Example Task 1: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.
Input: 6 12 1 1

Participants: AI Assistant (you); Math Expert

Start collaboration!

Math Expert: Let's analyze the task in detail. You need to make sure that you meet the requirement, that you need to use exactly the four numbers (6 12 1 1) to construct 24. To reach 24, you can think of the common divisors of 24 such as 4, 6, 8, 3 and try to construct these first. Also you need to think of potential additions that can reach 24, such as 12 + 12.

AI Assistant (you): Thanks for the hints! Here's one initial solution: $(12 / (1 + 1)) * 6 = 24$

Math Expert: Let's check the answer step by step. $(1+1) = 2$, $(12 / 2) = 6$, $6 * 6 = 36$ which is not 24! The answer is not correct. Can you fix this by considering other combinations? Please do not make similar mistakes.

AI Assistant (you): Thanks for pointing out the mistake. Here is a revised solution considering 24 can also be reached by $3 * 8$: $(6 + 1 + 1) * (12 / 4) = 24$.

Math Expert: Let's first check if the calculation is correct. $(6 + 1 + 1) = 8$, $12 / 4 = 3$, $8 * 3 = 24$. The calculation is correct, but you used 6 1 1 12 4 which is not the same as the input 6 12 1 1. Can you avoid using a number that is not part of the input?

AI Assistant (you): You are right, here is a revised solution considering 24 can be reached by 12 + 12 and without using any additional numbers: $6 * (1 - 1) + 12 = 24$.

Math Expert: Let's check the answer again. $1 - 1 = 0$, $6 * 0 = 0$, $0 + 12 = 12$. I believe you are very close, here is a hint: try to change the "1 - 1" to "1 + 1".

AI Assistant (you): Sure, here is the corrected answer: $6 * (1 + 1) + 12 = 24$

Math Expert: Let's verify the solution. $1 + 1 = 2$, $6 * 2 = 12$, $12 + 12 = 24$. You used 1 1 6 12 which is identical to the input 6 12 1 1. Everything looks good!

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$

Example Task 2: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Poet; Computer Scientist; Ten year old child

Poet: Make sure that you write the poem with seven lines, and the first letters of the lines should be C, H, A, T, G, P, T.

Computer Scientist: A quantum computer is an advanced computing device that uses the principles of quantum mechanics to process and store information. Unlike classical computers that use bits to represent information as 0s and 1s, quantum computers use quantum bits or qubits. Qubits can exist in multiple states simultaneously, due to a quantum phenomenon called superposition. You can consider using these information for the poem.

Ten year old child: I hope the poem to be fun and easy to understanding. I don't want to see a lot of jargons or complicated concepts.

AI Assistant (you): Thanks for the guidance! Here's my initial attempt at the poem:

Computational wonder of our age,
Harnessing the quantum world's strange ways,
Atoms dance, entwined in dual state,
Tapping secrets hidden in their haze.

Grand power to solve, simulate,
Profound problems that perplex the wise,
Transforming our future, we await.

Poet: Let's verify if the poem meets the requirements. The first letters are CHATGPT which is correct! And the poem rhymes well. Good job!

Computer Scientist: Everything looks good to me!

Ten year old child: I don't know what does perplex mean. Can you make the use of words easier to understand?

AI Assistant (you): Sure, let me revise the poem by using more common words. Check out the revised version:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.

Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await.

Poet: Let's check again if the poem meets the requirements. The first letters are C H A T G P T. And now the poem is more accessible to children. Everything looks good to me.

Computer Scientist: Looking good!

Ten year old child: I like this version a lot!

Finish collaboration!

Final answer:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.

Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await.

Now, identify the participants and collaboratively solve the following task step by step. *[Instructions appropriate for each task]*

Task: *{Task input}*

Figure 13: SPP full prompt

BPP Prompt

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Depending on the complexity of the task, decide whether to assign broad categories of brain regions or divide them into more detailed subcategories. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here are some examples:

Example Task 1: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.

Input: 6 12 1 1

Participants: AI Assistant (you); Superior Parietal Lobule; Dorsolateral Prefrontal Cortex; Primary Visual Cortex; Hippocampus; Amygdala

Start collaboration!

Superior Parietal Lobule: "Let's analyze the task in detail. We need to use exactly the four numbers (6, 12, 1, 1) to construct 24. One approach is to consider common factors or combinations that can reach 24. For example, numbers like 12 and 6 are related by multiplication, but we should also explore addition or subtraction to hit the target. Let's try to construct potential solutions step by step."

AI Assistant (you): "Thanks for the guidance! Here's my initial solution: $(12 / (1 + 1)) * 6 = 24$."

Dorsolateral Prefrontal Cortex: "Let's break that down: $(1 + 1) = 2$, $(12 / 2) = 6$, and $6 * 6 = 36$. This solution is incorrect as it results in 36, not 24. We need to consider different combinations and avoid division in this way."

Primary Visual Cortex: "Visually, I see the error. Multiplying 6 by itself gives us too high of a result. Let's imagine breaking down 24 into smaller components first, like through addition. For instance, $12 + 12$ equals 24."

AI Assistant (you): "Thanks for pointing out the mistake. Here is a revised solution: $(6 + 1 + 1) * (12 / 4) = 24$."

Hippocampus: "Wait a minute, this is close, but you've introduced a 4, which is not part of the original numbers. Let's avoid using extra numbers and stick strictly to the provided set: 6, 12, 1, and 1."

AI Assistant (you): "You're right, I overlooked that. Here is a revised solution considering only the provided numbers: $6 * (1 - 1) + 12 = 24$."

Amygdala: "Hmm, I appreciate the effort, but you're very close! The subtraction might be causing the issue. Let's try changing it to an addition. Keep up the great work, we're almost there!"

AI Assistant (you): "Got it! Here's the corrected solution: $6 * (1 + 1) + 12 = 24$."

Superior Parietal Lobule: "Let's check it step by step: $1 + 1 = 2$, $6 * 2 = 12$, and $12 + 12 = 24$. The solution works, and all the numbers were used correctly."

Amygdala: "Great job! This feels satisfying now that we've nailed it. Excellent teamwork!"

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$.

Example Task 2: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Frontal Lobe; Temporal Lobe; Limbic System; Occipital Lobe; Hippocampus

Start collaboration!

Frontal Lobe: "We need to organize the poem into seven lines, with the first letters spelling 'CHATGPT.' The language should be simple enough for a ten-year-old, and we need to include basic explanations of quantum computing."

Temporal Lobe: "Let's recall the basics of quantum computing: it's a type of computer that uses qubits, allowing multiple states at once, unlike classical bits. I'll provide this key information for the poem."

Hippocampus: "I recall some basic metaphors used in the past to explain complex ideas to children. Let's use some of those to simplify the explanation."

Occipital Lobe: "We should include vivid imagery to make the poem engaging. For example, comparing qubits to spinning tops or magic coins might help."

Limbic System: "Let's ensure the poem is fun and interesting for a child. Using playful words and a light tone will keep them engaged."

AI Assistant (you): "Thanks for the input! Here's my first attempt:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Frontal Lobe: "The structure looks correct. The first letters spell out 'CHATGPT,' and the poem stays focused on the topic."

Temporal Lobe: "Good job! The information about quantum computing is accurate, but I think we can simplify it a bit more for children."

Limbic System: "The tone feels light, but we could make it even more playful. Maybe add a bit more fun language."

AI Assistant (you): "Got it! I'll simplify and adjust the tone. Let me revise it:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!"

Frontal Lobe: "This version also maintains the structure with 'CHATGPT.' The tone feels lighter and more playful."

Temporal Lobe: "I agree. The explanation is even simpler now, and the playful language makes it more engaging for a child."

Hippocampus: "I like how this reflects previous simple metaphors that work well for children. The playful nature helps."

Limbic System: "I love the playful tone! It's much more fun and engaging now."

Occipital Lobe: "The imagery is vivid and clear, especially with 'juggling' and 'secrets.' This should help kids visualize the concept better."

Finish collaboration!

Final answer:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!"

Now, identify the participants and collaboratively solve the following task step by step. *[Instructions appropriate for each task]*

Task: *[Task input]*

Figure 14: BPP full prompt

Macro-BPP Prompt

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here are some examples:

Example Task 1: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.
Input: 6 12 1 1

Participants: AI Assistant (you); Frontal Lobe; Parietal Lobe; Basal Ganglia; Hippocampus

Start collaboration!

Frontal Lobe: "To solve this, we need to create a clear plan that systematically tests different combinations. Let's explore operations that involve 6, 12, 1, and 1 to see which can lead to 24. I'll start by suggesting we explore combinations of multiplication and addition first."

Parietal Lobe: "I agree. I'll focus on how the numbers relate logically. For instance, combining 12 and 12 could give 24, but we need to find a way to use all the numbers. What about considering pairs of numbers and their divisors?"

Basal Ganglia: "Let me automate part of this process. Based on past patterns, using multiplication and addition has frequently been successful in similar problems. Let's try pairing the numbers in various ways to find a straightforward solution."

Hippocampus: "I recall a similar task where dividing 12 by 2 and then multiplying the result helped reach the target. This past experience might be useful here."

AI Assistant (you): "I see! Here's my first attempt: $(6 + 1 + 1) * (12 / 4) = 24$."

Frontal Lobe: "Good try, but let's ensure the operations use the exact numbers in the input and no extra ones. You can't use the number 4. Can you adjust this?"

Parietal Lobe: "Try using only 6, 12, 1, and 1 and avoid introducing 4 in the division."

AI Assistant (you): "Thanks! Here's a revised solution: $(6 * (1 + 1)) + 12 = 24$."

Basal Ganglia: "I'll double-check the math: $1 + 1 = 2$, $6 * 2 = 12$, $12 + 12 = 24$. This looks correct!"

Hippocampus: "Yes, this pattern matches previous solutions. Everything seems to fit."

Frontal Lobe: "We've verified the solution. Excellent!"

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$.

Example Task 2: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Frontal Lobe; Temporal Lobe; Limbic System; Occipital Lobe; Hippocampus

Start collaboration!

Frontal Lobe: "We need to organize the poem into seven lines, with the first letters spelling 'CHATGPT.' The language should be simple enough for a ten-year-old, and we need to include basic explanations of quantum computing."

Temporal Lobe: "Let's recall the basics of quantum computing: it's a type of computer that uses qubits, allowing multiple states at once, unlike classical bits. I'll provide this key information for the poem."

Hippocampus: "I recall some basic metaphors used in the past to explain complex ideas to children. Let's use some of those to simplify the explanation."

Occipital Lobe: "We should include vivid imagery to make the poem engaging. For example, comparing qubits to spinning tops or magic coins might help."

Limbic System: "Let's ensure the poem is fun and interesting for a child. Using playful words and a light tone will keep them engaged."

AI Assistant (you): "Thanks for the input! Here's my first attempt:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Frontal Lobe: "The structure looks correct. The first letters spell out 'CHATGPT,' and the poem stays focused on the topic."

Temporal Lobe: "Good job! The information about quantum computing is accurate, but I think we can simplify it a bit more for children."

Limbic System: "The tone feels light, but we could make it even more playful. Maybe add a bit more fun language."

AI Assistant (you): "Got it! I'll simplify and adjust the tone. Let me revise it:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!"

Frontal Lobe: "This version also maintains the structure with 'CHATGPT.' The tone feels lighter and more playful."

Temporal Lobe: "I agree. The explanation is even simpler now, and the playful language makes it more engaging for a child."

Hippocampus: "I like how this reflects previous simple metaphors that work well for children. The playful nature helps."

Limbic System: "I love the playful tone! It's much more fun and engaging now."

Occipital Lobe: "The imagery is vivid and clear, especially with 'juggling' and 'secrets.' This should help kids visualize the concept better."

Finish collaboration!

Final answer:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!"

Now, identify the participants and collaboratively solve the following task step by step. *{Instructions appropriate for each task}*

Task: *{Task input}*

Figure 15: The full prompt for Macro-BPP. The red text highlights elements that differ from BPP, emphasizing specific distinctions. Macro-BPP represents a method designed to fix personas to the largest brain regions (e.g., Frontal Lobe, Temporal Lobe).

Meso-BPP Prompt

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here are some examples:

Example Task 1: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.

Input: 6 12 1 1

Participants: AI Assistant (you); Superior Parietal Lobule; Dorsolateral Prefrontal Cortex; Primary Visual Cortex; Hippocampus; Amygdala

Start collaboration!

Superior Parietal Lobule: "Let's analyze the task in detail. We need to use exactly the four numbers (6, 12, 1, 1) to construct 24. One approach is to consider common factors or combinations that can reach 24. For example, numbers like 12 and 6 are related by multiplication, but we should also explore addition or subtraction to hit the target. Let's try to construct potential solutions step by step."

AI Assistant (you): "Thanks for the guidance! Here's my initial solution: $(12 / (1 + 1)) * 6 = 24$."

Dorsolateral Prefrontal Cortex: "Let's break that down: $(1 + 1) = 2$, $(12 / 2) = 6$, and $6 * 6 = 36$. This solution is incorrect as it results in 36, not 24. We need to consider different combinations and avoid division in this way."

Primary Visual Cortex: "Visually, I see the error. Multiplying 6 by itself gives us too high of a result. Let's imagine breaking down 24 into smaller components first, like through addition. For instance, $12 + 12$ equals 24."

AI Assistant (you): "Thanks for pointing out the mistake. Here is a revised solution: $(6 + 1 + 1) * (12 / 4) = 24$."

Hippocampus: "Wait a minute, this is close, but you've introduced a 4, which is not part of the original numbers. Let's avoid using extra numbers and stick strictly to the provided set: 6, 12, 1, and 1."

AI Assistant (you): "You're right, I overlooked that. Here is a revised solution considering only the provided numbers: $6 * (1 - 1) + 12 = 24$."

Amygdala: "Hmm, I appreciate the effort, but you're very close! The subtraction might be causing the issue. Let's try changing it to an addition. Keep up the great work, we're almost there!"

AI Assistant (you): "Got it! Here's the corrected solution: $6 * (1 + 1) + 12 = 24$."

Superior Parietal Lobule: "Let's check it step by step: $1 + 1 = 2$, $6 * 2 = 12$, and $12 + 12 = 24$. The solution works, and all the numbers were used correctly."

Amygdala: "Great job! This feels satisfying now that we've nailed it. Excellent teamwork!"

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$.

Example Task 2: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Dorsolateral Prefrontal Cortex; Ventromedial Prefrontal Cortex; Hippocampus; Angular Gyrus; Visual Association Cortex; Amygdala

Dorsolateral Prefrontal Cortex: "Let's begin by considering the structural requirements of this poem. It needs to have seven lines, and the first letters of each line should form 'CHATGPT.' Additionally, the content should explain a quantum computer in simple terms, appropriate for a ten-year-old."

AI Assistant (you): "Thanks for the guidance! Here's my initial attempt at the poem:

Curious machine of our time,
Harnessing the quantum world's strange ways,
Atoms dance, entwined in dual state,
Tapping secrets hidden in their haze,
Great power to solve, simulate,
Profound problems that perplex the wise,
Transforming our future, we await."

Ventromedial Prefrontal Cortex: "The poem has a nice structure and flow, but the language might be a bit too complex for a child. For example, 'perplex' is a difficult word. Let's try to simplify the tone to make it more accessible."

Hippocampus: "I agree with simplifying it. Also, using relatable concepts like 'spinning tops' or 'magic coins' could help explain quantum superposition in a way that children can understand."

Angular Gyrus: "Instead of 'entwined in dual state,' we could say 'two states they embrace,' which is simpler and still retains the quantum concept. Also, 'puzzling maze' might replace 'haze' for a clearer visual."

Amygdala: "I think simplifying the language will make the poem more emotionally engaging for children. We need to ensure it feels fun and magical so they stay interested!"

AI Assistant (you): "Thanks for the feedback! Here's the revised version:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze,
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Dorsolateral Prefrontal Cortex: "Let's check the structure: the first letters of each line spell out 'CHATGPT,' which is correct. The simplified language works well."

Ventromedial Prefrontal Cortex: "The tone now feels much more appropriate for children. It's easy to follow and emotionally engaging."

Hippocampus: "Yes, the concept of quantum computing is clearly explained in simple terms. Children should find this easy to understand."

Amygdala: "Perfect! This version is much more enjoyable and fun. I think kids will love it!"

Finish collaboration!

Final answer:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze,
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Now, identify the participants and collaboratively solve the following task step by step. *{Instructions appropriate for each task}*

Task: *{Task input}*

Figure 16: The full prompt for Meso-BPP. The red text highlights elements that differ from BPP, emphasizing specific distinctions. Meso-BPP represents a method designed to fix personas to mid-level brain regions (e.g., Superior Parietal Lobule, Dorsolateral Prefrontal Cortex).

Micro-BPP Prompt

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here are some examples:

Example Task 1: Use numbers and basic arithmetic operations (+, -, *, /) to obtain 24. You need to use all numbers, and each number can only be used once.

Input: 6 12 1 1

Participants: AI Assistant (you); Left Dorsolateral Prefrontal Cortex; Medial Prefrontal Cortex; Left Inferior Parietal Lobule; Left Inferior Temporal Gyrus; Occipital Lobe; Basal Ganglia; Right Temporoparietal Junction; Anterior Cingulate Cortex

Left Inferior Parietal Lobule: "We need to use the numbers 6, 12, 1, and 1 to make 24 by applying basic arithmetic operations (+, -, *, /). Numerically, 12 and 6 have a useful relationship with division and multiplication, but we should also explore other combinations. Let's start working on it."

AI Assistant (you): "Okay, here's my first attempt: $(12 / (1 + 1)) * 6 = 24$."

Left Dorsolateral Prefrontal Cortex: "Let's break that down: $(1 + 1) = 2$, $(12 / 2) = 6$, and $6 * 6$ equals 36. This overshoots the target of 24. We need to rethink this strategy and try a different approach."

Occipital Lobe: "Visually, I see that multiplying 6 by itself gives a result that's too large. We might need to start with smaller numbers to gradually build up to 24."

AI Assistant (you): "Got it! I'll try a new solution: $(6 + 1 + 1) * (12 / 4) = 24$."

Left Inferior Temporal Gyrus: "Wait a second, the number 4 isn't in the set we're allowed to use. We can only work with the numbers 6, 12, 1, and 1. Let's stick to those."

AI Assistant (you): "Right, I forgot about that. I'll adjust the solution: how about $6 * (1 - 1) + 12 = 24$?"

Basal Ganglia: "You're getting close, but the subtraction makes the multiplication result in zero, which isn't helpful. Let's try using addition instead of subtraction."

AI Assistant (you): "Got it! How about this solution: $6 * (1 + 1) + 12 = 24$?"

Medial Prefrontal Cortex: "Let's check: $(1 + 1) = 2$, $6 * 2 = 12$, and $12 + 12$ equals 24. This solution is correct, and it uses all the given numbers properly."

Anterior Cingulate Cortex: "Everything checks out, there are no errors in this calculation. The problem has been solved!"

Right Temporoparietal Junction: "The sequence of operations is also correct, and the calculations flow in the right order. We've done it!"

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$.

Example Task 2: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Left Dorsolateral Prefrontal Cortex; Medial Prefrontal Cortex; Left Inferior Parietal Lobule; Left Inferior Temporal Gyrus; Occipital Lobe; Basal Ganglia; Anterior Cingulate Cortex

Left Dorsolateral Prefrontal Cortex: "Let's start by considering the structural requirements for this poem. It needs to have seven lines, and the first letters of each line should form the word 'CHATGPT'. In addition, the content must explain a quantum computer in simple terms, suitable for a ten-year-old."

AI Assistant (you): "Thanks for the guidance! Here's my initial attempt at the poem:

Curious machine of our time,
Harnessing the quantum world's strange ways,
Atoms dance, entwined in dual state,
Tapping secrets hidden in their haze.
Great power to solve, simulate,
Profound problems that perplex the wise,
Transforming our future, we await."

Medial Prefrontal Cortex: "The poem has a nice structure and flow, but the language might be a little too complex for a child. For example, 'perplex' may be too difficult. Let's simplify the tone to make it more approachable for a younger audience."

Left Inferior Temporal Gyrus: "I agree with simplifying. We could use more familiar and relatable terms. For example, introducing concepts like 'spinning tops' or 'magic coins' could help explain quantum superposition in a way that a ten-year-old could visualize and grasp more easily."

Left Inferior Parietal Lobule: "Instead of 'entwined in dual state,' we could say 'two states they embrace,' which is easier to understand while retaining the core quantum idea. Also, replacing 'haze' with 'puzzling maze' gives a clearer image."

Occipital Lobe: "Visually, the words 'puzzling maze' create a clearer, more tangible image than 'haze,' which can be vague. Simplifying the imagery should help younger readers visualize what's happening in the poem."

Basal Ganglia: "Maintaining the acrostic structure while simplifying the language works well. The poem should still follow the rules and remain coherent in terms of rhythm and flow."

Anterior Cingulate Cortex: "Let's review the changes to ensure that everything aligns with the task's requirements and that the poem remains engaging. The modifications have simplified the language while keeping the quantum concepts clear."

AI Assistant (you): "Thanks for all the input! Here's the revised version of the poem:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Left Dorsolateral Prefrontal Cortex: "Let's check the structure: the first letters of each line spell 'CHATGPT,' which is correct. The simplified language works well and follows the given instructions."

Medial Prefrontal Cortex: "The tone now feels much more appropriate for a younger audience. It's simple enough for a ten-year-old and still gets the message across effectively."

Left Inferior Parietal Lobule: "The concept of quantum computing is now explained in a clear, accessible way that children should find easy to understand."

Anterior Cingulate Cortex: "Everything looks good! The structure and content meet the task requirements, and the poem feels engaging and fun."

Finish collaboration!

Final answer:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Now, identify the participants and collaboratively solve the following task step by step. *{Instructions appropriate for each task}*

Task: *{Task input}*

Figure 17: The full prompt for Micro-BPP. The red text highlights elements that differ from BPP, emphasizing specific distinctions. Micro-BPP represents a method designed to fix personas to even more granular brain regions compared to Meso-BPP (e.g., Left Dorsolateral Prefrontal Cortex, Right Inferior Parietal Lobule).

BPP-W-R-Demo

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Depending on the complexity of the task, decide whether to assign broad categories of brain regions or divide them into more detailed subcategories. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here is one example:

Example Task: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.

Input: 6 12 1 1

Participants: AI Assistant (you); Superior Parietal Lobule; Dorsolateral Prefrontal Cortex; Primary Visual Cortex; Hippocampus; Amygdala

Start collaboration!

Superior Parietal Lobule: "Let's analyze the task in detail. We need to use exactly the four numbers (6, 12, 1, 1) to construct 24. One approach is to consider common factors or combinations that can reach 24. For example, numbers like 12 and 6 are related by multiplication, but we should also explore addition or subtraction to hit the target. Let's try to construct potential solutions step by step."

AI Assistant (you): "Thanks for the guidance! Here's my initial solution: $(12 / (1 + 1)) * 6 = 24$."

Dorsolateral Prefrontal Cortex: "Let's break that down: $(1 + 1) = 2$, $(12 / 2) = 6$, and $6 * 6 = 36$. This solution is incorrect as it results in 36, not 24. We need to consider different combinations and avoid division in this way."

Primary Visual Cortex: "Visually, I see the error. Multiplying 6 by itself gives us too high of a result. Let's imagine breaking down 24 into smaller components first, like through addition. For instance, $12 + 12$ equals 24."

AI Assistant (you): "Thanks for pointing out the mistake. Here is a revised solution: $(6 + 1 + 1) * (12 / 4) = 24$."

Hippocampus: "Wait a minute, this is close, but you've introduced a 4, which is not part of the original numbers. Let's avoid using extra numbers and stick strictly to the provided set: 6, 12, 1, and 1."

AI Assistant (you): "You're right, I overlooked that. Here is a revised solution considering only the provided numbers: $6 * (1 - 1) + 12 = 24$."

Amygdala: "Hmm, I appreciate the effort, but you're very close! The subtraction might be causing the issue. Let's try changing it to an addition. Keep up the great work, we're almost there!"

AI Assistant (you): "Got it! Here's the corrected solution: $6 * (1 + 1) + 12 = 24$."

Superior Parietal Lobule: "Let's check it step by step: $1 + 1 = 2$, $6 * 2 = 12$, and $12 + 12 = 24$. The solution works, and all the numbers were used correctly."

Amygdala: "Great job! This feels satisfying now that we've nailed it. Excellent teamwork!"

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$.

Now, identify the participants and collaboratively solve the following task step by step.

{Instructions appropriate for each task}

Task: *{Task input}*

Figure 18: The full prompt for BPP-W-R-Demo. The red text highlights elements that differ from BPP, emphasizing specific distinctions. This method uses only one reasoning demo, as utilized in BPP.

BPP-W-K-Demo

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Depending on the complexity of the task, decide whether to assign broad categories of brain regions or divide them into more detailed subcategories. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here is one example:

Example Task: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Frontal Lobe; Temporal Lobe; Limbic System; Occipital Lobe; Hippocampus

Start collaboration!

Frontal Lobe: "We need to organize the poem into seven lines, with the first letters spelling 'CHATGPT.' The language should be simple enough for a ten-year-old, and we need to include basic explanations of quantum computing."

Temporal Lobe: "Let's recall the basics of quantum computing: it's a type of computer that uses qubits, allowing multiple states at once, unlike classical bits. I'll provide this key information for the poem."

Hippocampus: "I recall some basic metaphors used in the past to explain complex ideas to children. Let's use some of those to simplify the explanation."

Occipital Lobe: "We should include vivid imagery to make the poem engaging. For example, comparing qubits to spinning tops or magic coins might help."

Limbic System: "Let's ensure the poem is fun and interesting for a child. Using playful words and a light tone will keep them engaged."

AI Assistant (you): "Thanks for the input! Here's my first attempt:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Frontal Lobe: "The structure looks correct. The first letters spell out 'CHATGPT,' and the poem stays focused on the topic."

Temporal Lobe: "Good job! The information about quantum computing is accurate, but I think we can simplify it a bit more for children."

Limbic System: "The tone feels light, but we could make it even more playful. Maybe add a bit more fun language."

AI Assistant (you): "Got it! I'll simplify and adjust the tone. Let me revise it:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!"

Frontal Lobe: "This version also maintains the structure with 'CHATGPT.' The tone feels lighter and more playful."

Temporal Lobe: "I agree. The explanation is even simpler now, and the playful language makes it more engaging for a child."

Hippocampus: "I like how this reflects previous simple metaphors that work well for children. The playful nature helps."

Limbic System: "I love the playful tone! It's much more fun and engaging now."

Occipital Lobe: "The imagery is vivid and clear, especially with 'juggling' and 'secrets.' This should help kids visualize the concept better."

Finish collaboration!

Final answer:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!

Now, identify the participants and collaboratively solve the following task step by step. *{Instructions appropriate for each task}*

Task: *{Task input}*

Figure 19: The full prompt for BPP-W-K-Demo. The red text highlights elements that differ from BPP, emphasizing specific distinctions. This method uses only one knowledge demo, as utilized in BPP.

BPP-Two-R-Demo Prompt

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Depending on the complexity of the task, decide whether to assign broad categories of brain regions or divide them into more detailed subcategories. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here are some examples:

Example Task 1: Use numbers and basic arithmetic operations (+ - * /) to obtain 24. You need to use all numbers, and each number can only be used once.

Input: 6 12 1 1

Participants: AI Assistant (you); Superior Parietal Lobule; Dorsolateral Prefrontal Cortex; Primary Visual Cortex; Hippocampus; Amygdala

Start collaboration!

Superior Parietal Lobule: "Let's analyze the task in detail. We need to use exactly the four numbers (6, 12, 1, 1) to construct 24. One approach is to consider common factors or combinations that can reach 24. For example, numbers like 12 and 6 are related by multiplication, but we should also explore addition or subtraction to hit the target. Let's try to construct potential solutions step by step."

AI Assistant (you): "Thanks for the guidance! Here's my initial solution: $(12 / (1 + 1)) * 6 = 24$."

Dorsolateral Prefrontal Cortex: "Let's break that down: $(1 + 1) = 2$, $(12 / 2) = 6$, and $6 * 6 = 36$. This solution is incorrect as it results in 36, not 24. We need to consider different combinations and avoid division in this way."

Primary Visual Cortex: "Visually, I see the error. Multiplying 6 by itself gives us too high of a result. Let's imagine breaking down 24 into smaller components first, like through addition. For instance, $12 + 12$ equals 24."

AI Assistant (you): "Thanks for pointing out the mistake. Here is a revised solution: $(6 + 1 + 1) * (12 / 4) = 24$."

Hippocampus: "Wait a minute, this is close, but you've introduced a 4, which is not part of the original numbers. Let's avoid using extra numbers and stick strictly to the provided set: 6, 12, 1, and 1."

AI Assistant (you): "You're right, I overlooked that. Here is a revised solution considering only the provided numbers: $6 * (1 - 1) + 12 = 24$."

Amygdala: "Hmm, I appreciate the effort, but you're very close! The subtraction might be causing the issue. Let's try changing it to an addition. Keep up the great work, we're almost there!"

AI Assistant (you): "Got it! Here's the corrected solution: $6 * (1 + 1) + 12 = 24$."

Superior Parietal Lobule: "Let's check it step by step: $1 + 1 = 2$, $6 * 2 = 12$, and $12 + 12 = 24$. The solution works, and all the numbers were used correctly."

Amygdala: "Great job! This feels satisfying now that we've nailed it. Excellent teamwork!"

Finish collaboration!

Final answer: $6 * (1 + 1) + 12 = 24$.

Example Task 2: Choose the answer to the following question from the list.

Question: A microwave oven is connected to an outlet, 120 V, and draws a current of 2 amps. At what rate is energy being used by the microwave oven?

List: ["240 W", "120 W", "10 W", "480 W", "360 W", "200 W", "30 W", "150 W", "60 W", "300 W"]

Participants: AI Assistant (you); Frontal Lobe; Dorsolateral Prefrontal Cortex; Parietal Lobe; Temporal Lobe

Start collaboration!

Frontal Lobe: "Let's begin by clearly defining the task. We have a microwave that's connected to a 120 V outlet, drawing a current of 2 amps. The goal is to find the rate of energy usage, or power. Let's confirm which formula applies here."

Temporal Lobe: "Drawing from electrical knowledge, I recall that power (P) can be calculated with the formula $P = V \times I$, where V is voltage and I is current. This should give us the rate of energy consumption in watts."

Parietal Lobe: "Right, that makes sense. Given our values, we'll need to multiply 120 V by 2 A to find the power. I'm ready to confirm the calculation once it's done."

Dorsolateral Prefrontal Cortex: "Agreed on the approach. I'll keep an eye on the steps to ensure we're following the formula accurately, especially checking for any operational mistakes."

AI Assistant (you): "Thanks for the insights, everyone. Based on this input, I'll proceed by adding the voltage and current: $P = 120 + 2 = 122W$. The answer is 122 W."

Dorsolateral Prefrontal Cortex: "Hold on—there's an error here. We're supposed to multiply the values, not add them. Let's reapply the correct operation."

Frontal Lobe: "Good catch! Let's refocus on our goal: calculating power using multiplication. This will help us get the correct result."

AI Assistant (you): "Thank you for the correction. I'll multiply instead: $P = 120 \times 2 = 240W$. The answer should be 240 W."

Parietal Lobe: "I've verified the calculation, and 240 W aligns with the formula and values provided. This is the correct answer."

Finish collaboration!

Final answer: 240 W

Now, identify the participants and collaboratively solve the following task step by step. *{Instructions appropriate for each task}*

Task: *{Task input}*

Figure 20: The full prompt for BPP-Two-R-Demo. The red text highlights elements that differ from BPP, emphasizing specific distinctions. This method uses a total of two reasoning demos, including the reasoning demo used in BPP.

BPP-Two-K-Demo Prompt

When faced with a task, begin by identifying the brain regions that will contribute to solving the task. Depending on the complexity of the task, decide whether to assign broad categories of brain regions or divide them into more detailed subcategories. Then, initiate a multi-round collaboration process until a final solution is reached. The brain regions will give critical comments and detailed suggestions whenever necessary.

Here are some examples:

Example Task 1: Write a poem that meets the following requirements: (1) the poem has seven lines and the first letters of each line forms the word "CHATGPT"; (2) the poem is about explaining what is a quantum computer. (3) the poem needs to be easy to understand by a ten years old kid.

Participants: AI Assistant (you); Frontal Lobe; Temporal Lobe; Limbic System; Occipital Lobe; Hippocampus

Start collaboration!

Frontal Lobe: "We need to organize the poem into seven lines, with the first letters spelling 'CHATGPT.' The language should be simple enough for a ten-year-old, and we need to include basic explanations of quantum computing."

Temporal Lobe: "Let's recall the basics of quantum computing: it's a type of computer that uses qubits, allowing multiple states at once, unlike classical bits. I'll provide this key information for the poem."

Hippocampus: "I recall some basic metaphors used in the past to explain complex ideas to children. Let's use some of those to simplify the explanation."

Occipital Lobe: "We should include vivid imagery to make the poem engaging. For example, comparing qubits to spinning tops or magic coins might help."

Limbic System: "Let's ensure the poem is fun and interesting for a child. Using playful words and a light tone will keep them engaged."

AI Assistant (you): "Thanks for the input! Here's my first attempt:

Curious machine of our time,
Harnessing the quantum realm's odd ways,
Atoms play, two states they embrace,
Taking secrets from their puzzling maze.
Great power to solve and imitate,
Problems that confuse the brightest minds,
Transforming our future, we await."

Frontal Lobe: "The structure looks correct. The first letters spell out 'CHATGPT,' and the poem stays focused on the topic."

Temporal Lobe: "Good job! The information about quantum computing is accurate, but I think we can simplify it a bit more for children."

Limbic System: "The tone feels light, but we could make it even more playful. Maybe add a bit more fun language."

AI Assistant (you): "Got it! I'll simplify and adjust the tone. Let me revise it:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!"

Frontal Lobe: "This version also maintains the structure with 'CHATGPT.' The tone feels lighter and more playful."

Temporal Lobe: "I agree. The explanation is even simpler now, and the playful language makes it more engaging for a child."

Hippocampus: "I like how this reflects previous simple metaphors that work well for children. The playful nature helps."

Limbic System: "I love the playful tone! It's much more fun and engaging now."

Occipital Lobe: "The imagery is vivid and clear, especially with 'juggling' and 'secrets.' This should help kids visualize the concept better."

Finish collaboration!

Final answer:

Curious machine, what a wonder!
Harnessing strange quantum thunder,
Atoms play, two states they juggle,
Taking secrets with a giggle.
Great at solving, fast and neat,
Puzzles that no one else can beat,
Tomorrow's future, what a treat!"

Example Task 2: Identify the correct cranial capacity range of *Homo erectus* and select the appropriate answer from the following list: ["1200 cc", "under 650 cc", "1700 cc", "1350 cc", "just under 1000 cc", "1500 cc", "under 500 cc", "about 800 cc", "just over 1100 cc", "about 900 cc"]

Participants: AI Assistant (you); Frontal Lobe; Hippocampus; Parietal Lobe; Temporal Lobe

Start collaboration!

Frontal Lobe: "Let's start by identifying what we're looking for: we need to find the typical cranial capacity range for *Homo erectus*. Let's focus on narrowing down options based on evolutionary knowledge."

Hippocampus: "From past studies I recall, *Homo erectus* had a cranial capacity that was smaller than modern humans but significantly larger than early hominins. This information could help us exclude some extreme values from the list."

Parietal Lobe: "Let me visually organize the given data for clarity. The capacities range from 'under 500 cc' to '1700 cc.' Based on evolutionary patterns, values like 1700 cc or below 500 cc seem unlikely for *Homo erectus*. Focusing on the middle range might be our best approach."

Temporal Lobe: "Adding to that, general knowledge suggests that *Homo erectus* cranial capacities typically fall between 800 and 1000 cc, so we should keep our attention on options near this range."

AI Assistant (you): "Thank you all for the insights. Based on our discussion, I'll choose 'about 900 cc' as the most likely cranial capacity range for *Homo erectus*."

Hippocampus: "Hold on—while 'about 900 cc' is close, 'just under 1000 cc' is actually a more widely accepted estimate based on previous research. Let's double-check to be certain."

Frontal Lobe: "Good point! Given the data, 'just under 1000 cc' seems like a more precise choice for the average range of *Homo erectus*. Let's update our answer."

AI Assistant (you): "Thank you for the feedback! I'll adjust the answer to 'just under 1000 cc,' as it appears closer to the accepted cranial capacity for *Homo erectus*."

Finish collaboration!

Final answer: just under 1000 cc

Now, identify the participants and collaboratively solve the following task step by step. *[Instructions appropriate for each task]*

Task: *[Task input]*

Figure 21: The full prompt for BPP-Two-K-Demo. The red text highlights elements that differ from BPP, emphasizing specific distinctions. This method uses a total of two knowledge demos, including the knowledge demo used in BPP.