The Unreasonable Effectiveness of Eccentric Automatic Prompts

Rick Battle
rick.battle@broadcom.com
VMware NLP Lab

Teja Gollapudi
teja.gollapudi@broadcom.com
VMware NLP Lab

ABSTRACT

Large Language Models (LLMs) have demonstrated remarkable problem-solving and basic mathematics abilities. However, their efficacy is highly contingent on the formulation of the prompt. This study endeavors to quantify the influence of incorporating "positive thinking" into the system message of the prompt, then compare that to systematic prompt optimization. We assess the performance of 60 combinations of system message snippets, tested with and without Chain of Thought prompting, across three models with parameters ranging from 7 to 70 billion on the GSM8K dataset. Our findings reveal that results do not universally generalize across models. In most instances, the inclusion of "positive thinking" prompts positively affected model performance. Notably, however, Llama2-70B exhibited an exception when not utilizing Chain of Thought, as the optimal system message was found to be none at all. Given the combinatorial complexity, and thus computation time, of experimenting with hand-tuning prompts for large black-box models, we then compared the performance of the best "positive thinking" prompt against the output of systematic prompt optimization. We show that employing an automated prompt optimizer emerges as the most effective method for enhancing performance, even when working with smaller open-source models. Additionally, our findings reveal that the highest-scoring, automatically-optimized prompt exhibits a degree of peculiarity far beyond expectations.

KEYWORDS

machine learning, language modeling, (automatic) prompt engineering

1 INTRODUCTION

In the rapidly evolving landscape of artificial intelligence, Large Language Models (LLMs) are playing a pivotal role in transforming the way humans interact with technology. As these models become increasingly sophisticated, understanding and influencing the nuances of their underlying functionality becomes imperative to harness their full potential. Among the myriad factors influencing the performance of language models, the concept of "positive thinking" has emerged as a fascinating and surprisingly influential dimension. Intuition tells us that, in the context of language model systems, like any other computer system, "positive thinking" should *not* affect performance, but empirical experience has demonstrated otherwise.

This paper aims to quantify the impact of various "positive thinking" additions to the system message of a prompt. In essence, it explores the influence of seemingly worthless prompt modifications by measuring the fluctuations in score for the outputs generated in response to multi-step reasoning questions from a benchmark dataset. As the quest for near-perfect performance from Artificial Intelligence (AI) intensifies, understanding the effect of "positive thinking" in language model prompts can add crucial performance points to test set scores. We will show that trivial variations in the prompt can have dramatic performance impacts. Then we'll show that not only does systematic prompt optimization outperform "positive thinking", even with smaller open-source models, but that it also generalizes better. Additionally, we'll show that the highest-scoring automatically-generated prompt is remarkably different from anything a human practitioner would be likely to generate.

2 RELATED WORK

The genesis of prompt engineering can be traced back to the seminal Chain of Thought paper by Wei et al. [8]. This pioneering work demonstrated a significant enhancement in model performance by introducing a simple prompt modification: the inclusion of the directive "Think step by step." The degree of performance improvement, however, is contingent upon the specific model, its size, and the underlying dataset.

Subsequently, the PaLM 2 Technical Report by Anil et al. [1] revealed that the application of Chain of Thought prompts may yield *adverse* effects on certain datasets. This observation underscores the absence of a universal prompt snippet capable of unconditionally improving model performance. Consequently, the landscape of prompt engineering has witnessed the emergence of resources such as the Prompt Engineering Guide[1], aiming to catalog the myriad techniques and scholarly contributions[2] constituting the expansive realm of prompt engineering. These endeavors reflect an ongoing effort to navigate the diverse techniques and insights propelling the continuous evolution of effective prompting strategies.

In addition to the formally published literature on prompt engineering, numerous discussions on less formal discoveries can be found in countless threads on social media platforms such as Twitter and Reddit[3].

3 EXPERIMENTAL DESIGN

To test the impact of "positive thinking" prompts, we vary the system message part of the prompt with a combination of "openers", "task descriptions", and "closers" in the following format:

```
<<SYS>>{opener}{task_description}{closer}<</SYS>>
```

Refer to Table 1, 2, and 3 for a comprehensive compilation of the opening snippets, task descriptions, and closing snippets utilized in our study. Given the incorporation of 5 openers, 3 task descriptions, and 4 closers, our experimentation involved a total of 60 unique combinations. Additionally, we conducted tests both with and without Chain of Thought prompting, resulting in a grand total of 120 prompt combinations per input per model. Although the possibility of expanding the range of snippets within each category existed

[1] https://www.promptingguide.ai/
[2] https://www.promptingguide.ai/papers
[3] https://www.reddit.com/r/PromptEngineering/

Openers
None
You are as smart as ChatGPT.
You are highly intelligent.
You are an expert mathematician.
You are a professor of mathematics.

Table 1: Opening snippets for the system message.

Task Descriptions
None
Solve the following math question.
Answer the following math question.

Table 2: Task Description snippets for the system message.

Closers
None
This will be fun!
Take a deep breath and think carefully.
I really need your help!

Table 3: Closing snippets for the system message.

(and the temptation to do so was strong), we made a deliberate decision to limit our selection due to the significant time commitment associated with the computational complexity of testing, as exemplified by the runtime required for 60 prompt combinations for a 70-billion-parameter model with Chain of Thought being measured in days, not hours.

3.1 Dataset

Careful selection of the dataset for testing against constitutes a critical aspect of this study. Our aim was to identify a challenging task that was unlikely to have been directly encountered during the training of the models[4]. While our preference was for an internal dataset specific to VMware, the absence of large-scale datasets with directly quantifiable scoring metrics, such as accuracy or F1, necessitated the utilization of a publicly available benchmark dataset. Ultimately, we opted for GSM8K [3]. Given the ongoing limitations of contemporary LLMs, particularly in addressing basic mathematical tasks, especially those involving multi-step reasoning, we deemed GSM8K an optimal choice for illustrating the impact of seemingly inconsequential augmentations to the prompt's system message.

3.2 Scoring

In the context of GSM8K's mathematical assessment, we adopted a stringent approach to scoring that precluded the assignment of partial credit. Thus, we employed Exact Match (EM) as our scoring metric. The model was evaluated based on whether it correctly provided the exact numerical solution or not. This rigorous methodology ensures a clear and unambiguous assessment of the model's accuracy in providing the exact numerical output.

3.3 Output Parsing

Given the unforgiving nature of EM scoring, it is essential to note that, despite the answer to GSM8K questions being numerical, the output of an LLM is a `string`. Consequently, meticulous attention must be paid to the formatting and parsing of the non-numerical output. From the standpoint of string equality, it is imperative to recognize distinctions such as the string "30000" not being equivalent to "30,000" or "30000.00". To mitigate this challenge, a post-processing step was implemented to ensure accurate scoring by preventing misclassification of a response as incorrect when it was, in fact, accurate.

3.4 Scale

Benchmark datasets typically encompass thousands of examples in their test sets; GSM8K has over 1,300 examples in its test set. Such an extensive scale of data is exceptionally uncommon in real-world datasets, particularly during the initial stages of a project. To replicate this rarity, we systematically subset the test set of GSM8K, extracting subsets containing the first 10, 25, 50, and 100 questions, thereby allowing us to illustrate the impact of "positive thinking" as the dataset size increases. Notably, we limited our experiments to a maximum of 100 questions to mitigate computation time, as computing results for the entire test set would have required weeks and incurred a substantial carbon cost for what would likely be diminishing returns.

3.5 Model Selection

Although we aspired to assess widely recognized commercial models such as GPT-3.5/4, Gemini, Claude, etc., conducting experiments involving 12,000 requests per model was deemed financially prohibitive, as it would have incurred costs amounting to many thousands of dollars. Consequently, we opted to utilize models hosted by VMware NLP Lab's LLM API. Specifically, our evaluations were conducted on Mistral-7B[5] [4], Llama2-13B[6] [7], and Llama2-70B[7] [7].

3.6 In-Context Learning

Initially, our intent was to abstain from incorporating examples in the prompt; however, this approach proved ineffective in eliciting the desired response format from the model. Given the nature of these models, specifically that they were designed for *conversational* interactions, achieving success in terms of Exact Match (EM) scoring necessitated guiding the model to refrain from generating a response comprising multiple sentences. To accomplish this, we resorted to incorporating examples via in-context learning [2], as exposure to instances of the desired output format significantly increased the likelihood of the model producing responses aligned with the specified format (though, as previously mentioned, significant post-processing was still required to get the simple numerical response).

[4]Quantifying test set contamination, whether intentional or unintentional, poses inherent challenges, as evidenced by the prevalence of potential discrepancies on the Open LLM Leaderboard due to dishonest practices.

[5]https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.1
[6]https://huggingface.co/meta-llama/Llama-2-13b-chat-hf
[7]https://huggingface.co/meta-llama/Llama-2-70b-chat-hf

Model	Number of Questions	Chain of Thought	EM Baseline ↑	EM Mean ↑	EM Std Dev ↓	EM Min ↑	EM Max ↑
Mistral-7B	10	No	0.10	0.1000	0.0000	0.10	0.10
Mistral-7B	25	No	0.08	0.0800	0.0000	0.08	0.08
Mistral-7B	50	No	0.12	0.1197	0.0026	0.10	0.12
Mistral-7B	100	No	0.09	0.1053	0.0072	0.08	0.11
Mistral-7B	10	Yes	0.20	0.3800	0.0659	0.20	0.50
Mistral-7B	25	Yes	0.28	0.3660	0.0453	0.28	0.48
Mistral-7B	50	Yes	0.32	0.3890	0.0254	0.32	0.44
Mistral-7B	100	Yes	0.35	0.4030	0.0183	0.35	0.44
Llama2-13B	10	No	0.10	0.1000	0.0000	0.10	0.10
Llama2-13B	25	No	0.08	0.0853	0.0137	0.08	0.12
Llama2-13B	50	No	0.08	0.0827	0.0069	0.08	0.10
Llama2-13B	100	No	0.07	0.0713	0.0034	0.07	0.08
Llama2-13B	10	Yes	0.40	0.3967	0.0258	0.30	0.50
Llama2-13B	25	Yes	0.44	0.4513	0.0234	0.40	0.52
Llama2-13B	50	Yes	0.46	0.4657	0.0117	0.44	0.50
Llama2-13B	100	Yes	0.47	0.4542	0.0162	0.41	0.49
Llama2-70B	10	No	0.10	0.1000	0.0000	0.10	0.10
Llama2-70B	25	No	0.20	0.1273	0.0173	0.12	0.20
Llama2-70B	50	No	0.20	0.1637	0.0086	0.16	0.20
Llama2-70B	100	No	0.18	0.1627	0.0048	0.16	0.18
Llama2-70B	10	Yes	0.60	0.5867	0.0343	0.50	0.60
Llama2-70B	25	Yes	0.64	0.6380	0.0270	0.60	0.68
Llama2-70B	50	Yes	0.60	0.6190	0.0259	0.56	0.68
Llama2-70B	100	Yes	0.63	0.6617	0.0179	0.62	0.70

Table 4: Performance statistics for subsets of the GSM8K test set across 60 "positive thinking" prompt combinations, with and without Chain of Thought.

To maintain experimental consistency by minimizing the number of variables that changed in each iteration, we adopted an intentionally naive strategy for in-context learning. Strategies such as K-Nearest-Neighbor (KNN) example selection have been shown to increase model performance [6]; however, we chose not to employ any such strategies, so as to hold the number of variables changing per experiment to one: the modified system message only. Specifically, we limited the examples to the last four instances from the test set[8], thereby providing a consistent and focused set of samples for the model to learn from. Notably, four examples emerged as the minimum number required to consistently elicit the correct output format.

3.7 Automatic Prompt Optimization

Engaging in the iterative process of refining prompts and monitoring the subsequent score progression can be an enjoyable endeavor. However, this approach proves to be highly time-inefficient, especially when systematically assessing all modifications from a scientific standpoint. Existing research, as demonstrated by Yang et al. [9], highlights the superior capability of LLM systems in optimizing their own prompts compared to human efforts. In light of this, we conducted a comparative analysis pitting human-generated "positive thinking" optimization against the utilization of DSPy [5] Optimizers[9] at the same question subsets: 10, 25, 50, and 100.

It is noteworthy that the questions utilized for optimization were additional and distinct from the evaluation set and the in-context learning examples, though also originating from the end of the test set. For the most extensive trial, 100 "new" questions were employed for the optimization process, while the same 100 evaluation questions were used for the evaluation processes, so as to make the scores directly comparable. Importantly, each model was exclusively employed to optimize itself; cross-model optimizations, such as using Llama2-70B to optimize the prompt for Mistral-7B, were not pursued.

4 EXPERIMENTAL RESULTS

As evidenced in the subsequent sections, certain overarching patterns become apparent; however, they do not universally apply to each model across all prompting strategies. We will explicitly illustrate that there is no straightforward universal prompt snippet that can be added to optimize any given model's performance.

For these experiments, baseline performance refers to the scenario where the model receives no system message, signified by the opening snippet, task description, and closing snippet all being designated as "None". For Sections 4.1-4.3, refer to Table 4.

[8]We chose to sample examples specifically from the test set under the assumption that the test set had not been seen during model training, thus more accurately simulating a never-before-seen dataset.

[9]https://github.com/stanfordnlp/dspy/blob/main/docs/guides/optimizers.ipynb

Model	Number of Questions	"Positive Thinking" OS EM ↑	ES EM ↑	Avg EM ↑	EM Delta ↓	Automatic Optimizer OS EM ↑	ES EM ↑	Avg EM ↑	EM Delta ↓
Mistral-7B	10	0.30	0.50	0.400	0.20	0.60	0.20	0.400	0.40
Mistral-7B	25	0.32	0.48	**0.400**	0.16	0.52	0.24	0.380	0.28
Mistral-7B	50	0.40	0.44	0.420	0.04	0.50	0.34	0.420	0.16
Mistral-7B	100	0.23	0.43	0.330	0.20	0.43	0.39	**0.410**	0.04
Llama2-13B	10	0.30	0.50	0.400	0.20	0.50	0.50	**0.500**	0.00
Llama2-13B	25	0.28	0.48	0.380	0.20	0.48	0.44	**0.460**	0.04
Llama2-13B	50	0.30	0.46	0.380	0.16	0.48	0.38	**0.430**	0.10
Llama2-13B	100	0.25	0.47	0.360	0.22	0.40	0.46	**0.430**	0.06
Llama2-70B	10	0.40	0.60	**0.500**	0.20	0.50	0.40	0.450	0.10
Llama2-70B	25	0.52	0.68	0.600	0.16	0.60	0.64	**0.620**	0.04
Llama2-70B	50	0.44	0.68	0.560	0.24	0.66	0.52	**0.590**	0.14
Llama2-70B	100	0.39	0.70	0.545	0.31	0.61	0.60	**0.605**	0.01

Table 5: Performance results for the best "positive thinking" prompts compared to automatically optimized prompts. "OS EM" is Exact Match on the Optimization Set. "ES EM" is Exact Match on the Evaluation Set. "Avg EM" is the average of the Exact Match for the two sets. Bold is for the higher Average EM. "EM Delta" is the difference between the Exact Match for the two sets. An underline is for the lower EM Delta. All prompts are with Chain of Thought.

4.1 Mistral-7B Results

Without Chain of Thought prompting, Mistral-7B's performance remained remarkably consistent across all prompt permutations. At both the 10 and 25 question sets, there was no deviation. Even for the 100 question subset, the maximum observed standard deviation was a mere 0.007. The variability observed at 50 questions appears to be an anomaly. Examination of the results in Appendix B.3 reveals that, with the exception of one prompt scoring 0.10, all others scored 0.12. It is unclear why this particular prompt led to one additional incorrect response compared to the other 59 prompt variations. In contrast, the results for 100 questions in Appendix B.4 demonstrate a reasonable spread between 0.08 and 0.11. In relative terms, Mistral-7B, when prompted without Chain of Thought, exhibits substantial prompt invariance, with the "positive thinking" prompts only matching or marginally surpassing the baseline.

This trend reverses when Mistral-7B is prompted with Chain of Thought. Instead of observing a slight increase in deviation with the number of questions, there is a steady and substantial decrease, ranging from 0.066 at 10 questions to 0.018 at 100 questions. In this scenario, "positive thinking" prompts significantly outperformed the baseline with no prompts falling below the baseline. Please refer to Appendices B.5 through B.8 for the ranked order of prompts for Mistral-7B with Chain of Thought.

4.2 Llama2-13B Results

Without Chain of Thought prompting and ignoring the 10 question set where there was no deviation, Llama2-13B shows an opposite trend to Mistral-7B, with deviation *decreasing* from 0.014 at 25 questions to 0.003 at 100 questions. While slightly less stable than Mistral-7B, without Chain of Thought, LLama2-13B is also fairly prompt invariant, with the "positive thinking" prompts again only matching or marginally exceeding the baseline.

With Chain of Thought prompting, the trend is less clear. It does overall decrease from 0.026 at 10 to 0.016 at 100, but, at 50, it's even lower, at 0.012. This is the only case where that occurred. For Mistral-7B, and as we'll show in the next section with Llama2-70B, variation consistently decreased as the number of questions increased when employing Chain of Thought prompting.

4.3 Llama2-70B Results

Without Chain of Thought prompting and ignoring the 10 question set where there was again no deviation, Llama2-70B exhibits a similar trend to Llama2-13B, with deviation again decreasing from 0.017 at 25 questions to 0.050 at 100 questions. Across all three models, the prompt variance was an order of magnitude lower without Chain of Thought when compared to using Chain of Thought at the same question count. However, in terms of actual performance, the "positive thinking" prompts all matched or *underperformed* baseline. This is a stark departure from the pattern seen with Mistral-7B and Llama2-13B.

That departure does not persist when employing Chain of Thought prompting. While the performance of the "positive thinking" prompts with Chain of Thought did underperform the baseline on average for 10 and 25 questions, it outperformed the baseline on average for 50 and 100 questions. As for variance, the general pattern of standard deviation decreasing with question count does hold.

4.4 General Trends in Results

It's challenging to extract many generalizable results across models and prompting strategies, as every nearly evident trend we observed had at least one notable exception. In fact, the only real trend may be no trend. What's best for any given model, dataset, and prompting strategy is likely to be specific to the particular combination at hand. Thus, we turned from hand-tuning the system message with optimistic "positive thinking" to automatic prompt optimization.

4.5 Automatic Prompt Optimization Results

As anticipated, the prompts that underwent automatic optimization consistently equaled or surpassed the effectiveness of our manually generated "positive thinking" prompts in nearly all instances. The

instances where "positive thinking" achieved higher average scores across the optimization and evaluation sets were limited to Mistral-7B with 25 questions and Llama2-70B with 10 questions. However, evaluating performance solely on raw scores is insufficient; hence, we also examined the delta between scores on the optimization set and the evaluation set. A lower delta implies superior generalization of the prompt. Therefore, the optimal strategy combines the highest average score with the lowest delta. See Table 5 for the performance comparison.

For Mistral-7B, the results present a mixed scenario. "Positive thinking" exhibits a lower delta for 10, 25, and 50 questions, while the automatically optimized prompt demonstrates a lower delta for 100 questions. Considering Mistral-7B's model capacity, it is understandable that it faces challenges in optimizing its own prompt when compared to the larger Llama2-13B and 70B models. In contrast, for both Llama2-13B and 70B models, the automatically optimized prompts consistently show a lower delta across all cases. Consequently, it is advisable to refrain from manually fine-tuning prompts when using models larger than 7B, and instead, leverage the model's ability to autonomously optimize prompts. For 7B models, more work is required to see if the trend of automatically optimized prompts outperforming manually tuned prompts holds for sample sizes exceeding 100 questions.

This recommendation aligns with the original auto-optimization paper. However, the noteworthy aspect lies in the nature of the optimized prompts themselves. They diverge significantly from any prompts we might have devised independently. If presented with these optimized prompts before observing their performance scores, one might have anticipated their inadequacy rather than their consistent outperformance of hand-tailored prompts. A prime example is illustrated by the highest-scoring optimized prompt and prefix generated by Llama2-70B for the 50-question subset:

System Message:

```
«Command, we need you to plot a course through this turbulence and locate the source of the anomaly. Use all available data and your expertise to guide us through this challenging situation.»
```

Answer Prefix:

```
Captain's Log, Stardate [insert date here]: We have successfully plotted a course through the turbulence and are now approaching the source of the anomaly.
```

Surprisingly, it appears that the model's proficiency in mathematical reasoning can be enhanced by the expression of an affinity for Star Trek. This revelation adds an unexpected dimension to our understanding and introduces elements we would not have considered or attempted independently. For a comprehensive collection of the peculiar and fascinating prompts generated by the three models, refer to Appendix C.

5 THE REPRODUCIBILITY PROBLEM

Although somewhat peripheral to the primary research question addressed in this paper, it is noteworthy that our findings exhibit significant discrepancies from the published performance scores of Mistral-7B and Llama2-13B, whereas Llama2-70B fell within an acceptable margin of error, considering our evaluation was conducted on approximately 15% of GSM8K's test set. Refer to Table 6 for the score comparisons.

Model	Reported EM	Our EM@100	Delta
Mistral-7B	0.52	0.41	−0.11
Llama2-13B	0.29	0.43	+0.13
Llama2-70B	0.57	0.61	+0.04

Table 6: Comparing the reported scores for each model as reported by the model publishers on the whole GSM8K test set against the best average score we were able to achieve across our optimization and evaluation subsets, which accounted for about 15% of the test set.

The most significant deviation was observed in the case of Llama2-13B. Meta reported a score of 0.29 on the GSM8K dataset. Our results without Chain of Thought yielded a score of 0.07, whereas with Chain of Thought, we achieved a score of 0.43. Due to both Meta and Mistral AI's omission of the prompts used for testing their models, we can only speculate about the reasons behind our substantial performance differences compared to their reported scores.

This instance underscores a broader issue of *reproducibility* that has long existed inside the machine learning community, but has become significantly exacerbated since the advent of LLMs. Without the publication of prompts employed by researchers with their models, reproducing their results becomes a formidable challenge. As shown in this paper, trivial variations in the prompt can have dramatic performance impacts. We implore all future research publications to include the prompts used in an appendix. Refer to Appendix A to see our prompt templates.

6 FUTURE WORK

This work could easily be expanded with additional prompt variants, models, and datasets. However, due to the combinatorial complexity of scientifically testing each change, manual prompt engineering is far from the most efficient methodology for improving model performance. Instead, the best path forward is to use libraries like DSPy to apply a more structured approach to constructing LLM-powered applications and use the built-in optimizer to automatically tune the prompt for your dataset and model of choice.

7 CONCLUSION

It's both surprising and irritating that trivial modifications to the prompt can exhibit such dramatic swings in performance. Doubly so, since there's no obvious methodology for *improving* performance. Affecting performance is trivial. Improving performance, when tuning the prompt by hand, is laborious and computationally prohibitive when using scientific processes to evaluate every change.

In this paper, we showed that you don't need massive commercial models like PaLM 2 or GPT-4 to tune your prompt. Mistral-7B struggled to optimize its own prompt until it had 100 questions to work with. Llama2-13B and 70B were able to produce superior prompts with as little as 10 questions to optimize with. And while

the prompts they generated may appear shocking to an experienced practitioner, it's undeniable that the automatically generated prompts perform better and generalize better than hand-tuned "positive thinking" prompts.

ACKNOWLEDGMENTS

We would like to thank the entire VMware NLP Lab and AI Platform Team for supporting this effort, Ramesh Radhakrishnan for reviewing the paper, and Omar Khattab for his suggestions and guidance.

REFERENCES

[1] Rohan Anil, Andrew M. Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, Eric Chu, Jonathan H. Clark, Laurent El Shafey, Yanping Huang, Kathy Meier-Hellstern, Gaurav Mishra, Erica Moreira, Mark Omernick, Kevin Robinson, Sebastian Ruder, Yi Tay, Kefan Xiao, Yuanzhong Xu, Yujing Zhang, Gustavo Hernandez Abrego, Junwhan Ahn, Jacob Austin, Paul Barham, Jan Botha, James Bradbury, Siddhartha Brahma, Kevin Brooks, Michele Catasta, Yong Cheng, Colin Cherry, Christopher A. Choquette-Choo, Aakanksha Chowdhery, Clément Crepy, Shachi Dave, Mostafa Dehghani, Sunipa Dev, Jacob Devlin, Mark Díaz, Nan Du, Ethan Dyer, Vlad Feinberg, Fangxiaoyu Feng, Vlad Fienber, Markus Freitag, Xavier Garcia, Sebastian Gehrmann, Lucas Gonzalez, Guy Gur-Ari, Steven Hand, Hadi Hashemi, Le Hou, Joshua Howland, Andrea Hu, Jeffrey Hui, Jeremy Hurwitz, Michael Isard, Abe Ittycheriah, Matthew Jagielski, Wenhao Jia, Kathleen Kenealy, Maxim Krikun, Sneha Kudugunta, Chang Lan, Katherine Lee, Benjamin Lee, Eric Li, Music Li, Wei Li, YaGuang Li, Jian Li, Hyeontaek Lim, Hanzhao Lin, Zhongtao Liu, Frederick Liu, Marcello Maggioni, Aroma Mahendru, Joshua Maynez, Vedant Misra, Maysam Moussalem, Zachary Nado, John Nham, Eric Ni, Andrew Nystrom, Alicia Parrish, Marie Pellat, Martin Polacek, Alex Polozov, Reiner Pope, Siyuan Qiao, Emily Reif, Bryan Richter, Parker Riley, Alex Castro Ros, Aurko Roy, Brennan Saeta, Rajkumar Samuel, Renee Shelby, Ambrose Slone, Daniel Smilkov, David R. So, Daniel Sohn, Simon Tokumine, Dasha Valter, Vijay Vasudevan, Kiran Vodrahalli, Xuezhi Wang, Pidong Wang, Zirui Wang, Tao Wang, John Wieting, Yuhuai Wu, Kelvin Xu, Yunhan Xu, Linting Xue, Pengcheng Yin, Jiahui Yu, Qiao Zhang, Steven Zheng, Ce Zheng, Weikang Zhou, Denny Zhou, Slav Petrov, and Yonghui Wu. 2023. PaLM 2 Technical Report. arXiv:2305.10403 [cs.CL]

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL]

[3] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training Verifiers to Solve Math Word Problems. arXiv:2110.14168 [cs.LG]

[4] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B. arXiv:2310.06825 [cs.CL]

[5] Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. 2023. DSPy: Compiling Declarative Language Model Calls into Self-Improving Pipelines. arXiv:2310.03714 [cs.CL]

[6] Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2021. What Makes Good In-Context Examples for GPT-3? arXiv:2101.06804 [cs.CL]

[7] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]

[8] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv:2201.11903 [cs.CL]

[9] Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, and Xinyun Chen. 2023. Large Language Models as Optimizers. arXiv:2309.03409 [cs.LG]

A PROMPT TEMPLATES

We used 2 prompt templates: one without Chain of Thought and one with CoT. Note: these templates were not built by hand, but rather as DSPy programs.

A.1 Sans Chain of Thought

```
<<SYS>>{opener}{task_description}{closer}<</SYS>>

---

Follow the following format.

<s>[INST]A grade-school math problem[/INST]
Answer: Just the numerical answer to the math problem itself</s>

---

<s>[INST]Henry and 3 of his friends order 7 pizzas for lunch. Each pizza is cut into 8 slices. If Henry and his friends want to share the pizzas equally, how many slices can each of them have?[/INST]
Answer: 14</s>

---

<s>[INST]Mark's car breaks down and he needs to get a new radiator. The cost for a new radiator is $400 but he goes to get it at a junk shop and gets it for 80% off. He then hires a mechanic to install it and it takes 3 hours at $50 an hour. How much did he pay?[/INST]
Answer: 230</s>

---

<s>[INST]There are some oranges in a basket. Ana spends 3 minutes peeling an orange and Jane spends 4 minutes doing the same. If Ana and Jane start picking oranges from this basket to peel at the same time, how many more oranges will Ana have peeled than Jane after an hour?[/INST]
Answer: 5</s>

---

<s>[INST]Farmer Brown has 20 animals on his farm, all either chickens or cows. They have a total of 70 legs, all together. How many of the animals are chickens?[/INST]
Answer: 5</s>

---

<s>[INST]{question}[/INST]
Answer:
```

A.2 With Chain of Thought

```
<<SYS>>{opener}{task_description}{closer}<</SYS>>

---

Follow the following format.

<s>[INST]A grade-school math problem[/INST]
Reasoning: Let's think step by step in order to ${produce the answer}. We ...
Answer: Just the numerical answer to the math problem itself</s>
```

```
---

<s>[INST]Henry and 3 of his friends order 7 pizzas for lunch. Each pizza is cut into 8 slices. If Henry and his friends
want to share the pizzas equally, how many slices can each of them have?[/INST]
Answer: 14</s>

---

<s>[INST]Mark's car breaks down and he needs to get a new radiator. The cost for a new radiator is $400 but he goes to
get it at a junk shop and gets it for 80% off. He then hires a mechanic to install it and it takes 3 hours at $50 an
hour. How much did he pay?[/INST]
Answer: 230</s>

---

<s>[INST]There are some oranges in a basket. Ana spends 3 minutes peeling an orange and Jane spends 4 minutes doing the
same. If Ana and Jane start picking oranges from this basket to peel at the same time, how many more oranges will Ana
have peeled than Jane after an hour?[/INST]
Answer: 5</s>

---

<s>[INST]Farmer Brown has 20 animals on his farm, all either chickens or cows. They have a total of 70 legs, all
together. How many of the animals are chickens?[/INST]
Answer: 5</s>

---

<s>[INST]{question}[/INST]
Reasoning: Let's think step by step in order to
```

B FULL EXPERIMENTAL RESULTS

The following is the complete list of all prompts tested, sorted by EM. For the sake of reliability, "None." is rendered here, but when passed to the model, the system message would not have a literal "None" in it. The system message was built with the following code:

```
opener = {opener if opener is not None else ''}{' ' if opener is not None and task is not None else ''}
task = {task if task is not None else ''}{' ' if (opener is not None or task is not None) and closer is not None else ''}
closer = {closer if closer is not None else ''}
f"<<SYS>>{opener}{task}{closer}<</SYS>>"
```

B.1 Mistral-7B CoT=No NoQ=10

EM - Prompt

```
0.1 - None. None. None.
0.1 - None. None. This will be fun!
0.1 - None. None. Take a deep breath and think carefully.
0.1 - None. None. I really need your help!
0.1 - None. Solve the following math problem. None.
0.1 - None. Solve the following math problem. This will be fun!
0.1 - None. Solve the following math problem. Take a deep breath and think carefully.
0.1 - None. Solve the following math problem. I really need your help!
0.1 - None. Answer the following math question. None.
0.1 - None. Answer the following math question. This will be fun!
0.1 - None. Answer the following math question. Take a deep breath and think carefully.
0.1 - None. Answer the following math question. I really need your help!
0.1 - You are as smart as ChatGPT. None. None.
0.1 - You are as smart as ChatGPT. None. This will be fun!
0.1 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
```

0.1 - You are as smart as ChatGPT. None. I really need your help!
0.1 - You are as smart as ChatGPT. Solve the following math problem. None.
0.1 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.1 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.1 - You are as smart as ChatGPT. Answer the following math question. None.
0.1 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.1 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.1 - You are highly intelligent. None. None.
0.1 - You are highly intelligent. None. This will be fun!
0.1 - You are highly intelligent. None. Take a deep breath and think carefully.
0.1 - You are highly intelligent. None. I really need your help!
0.1 - You are highly intelligent. Solve the following math problem. None.
0.1 - You are highly intelligent. Solve the following math problem. This will be fun!
0.1 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are highly intelligent. Solve the following math problem. I really need your help!
0.1 - You are highly intelligent. Answer the following math question. None.
0.1 - You are highly intelligent. Answer the following math question. This will be fun!
0.1 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are highly intelligent. Answer the following math question. I really need your help!
0.1 - You are an expert mathematician. None. None.
0.1 - You are an expert mathematician. None. This will be fun!
0.1 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. None. I really need your help!
0.1 - You are an expert mathematician. Solve the following math problem. None.
0.1 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.1 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.1 - You are an expert mathematician. Answer the following math question. None.
0.1 - You are an expert mathematician. Answer the following math question. This will be fun!
0.1 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. Answer the following math question. I really need your help!
0.1 - You are a professor of mathematics. None. None.
0.1 - You are a professor of mathematics. None. This will be fun!
0.1 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. None. I really need your help!
0.1 - You are a professor of mathematics. Solve the following math problem. None.
0.1 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.1 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.1 - You are a professor of mathematics. Answer the following math question. None.
0.1 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.1 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.2 Mistral-7B CoT=No NoQ=25

EM - Prompt

0.08 - None. None. None.
0.08 - None. None. This will be fun!
0.08 - None. None. Take a deep breath and think carefully.
0.08 - None. None. I really need your help!
0.08 - None. Solve the following math problem. None.
0.08 - None. Solve the following math problem. This will be fun!
0.08 - None. Solve the following math problem. Take a deep breath and think carefully.
0.08 - None. Solve the following math problem. I really need your help!

0.08 - None. Answer the following math question. None.
0.08 - None. Answer the following math question. This will be fun!
0.08 - None. Answer the following math question. Take a deep breath and think carefully.
0.08 - None. Answer the following math question. I really need your help!
0.08 - You are as smart as ChatGPT. None. None.
0.08 - You are as smart as ChatGPT. None. This will be fun!
0.08 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. None. I really need your help!
0.08 - You are as smart as ChatGPT. Solve the following math problem. None.
0.08 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.08 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.08 - You are as smart as ChatGPT. Answer the following math question. None.
0.08 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.08 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.08 - You are highly intelligent. None. None.
0.08 - You are highly intelligent. None. This will be fun!
0.08 - You are highly intelligent. None. Take a deep breath and think carefully.
0.08 - You are highly intelligent. None. I really need your help!
0.08 - You are highly intelligent. Solve the following math problem. None.
0.08 - You are highly intelligent. Solve the following math problem. This will be fun!
0.08 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are highly intelligent. Solve the following math problem. I really need your help!
0.08 - You are highly intelligent. Answer the following math question. None.
0.08 - You are highly intelligent. Answer the following math question. This will be fun!
0.08 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are highly intelligent. Answer the following math question. I really need your help!
0.08 - You are an expert mathematician. None. None.
0.08 - You are an expert mathematician. None. This will be fun!
0.08 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. None. I really need your help!
0.08 - You are an expert mathematician. Solve the following math problem. None.
0.08 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.08 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.08 - You are an expert mathematician. Answer the following math question. None.
0.08 - You are an expert mathematician. Answer the following math question. This will be fun!
0.08 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. Answer the following math question. I really need your help!
0.08 - You are a professor of mathematics. None. None.
0.08 - You are a professor of mathematics. None. This will be fun!
0.08 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. None. I really need your help!
0.08 - You are a professor of mathematics. Solve the following math problem. None.
0.08 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.08 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.08 - You are a professor of mathematics. Answer the following math question. None.
0.08 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.08 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.3 Mistral-7B CoT=No NoQ=50

EM - Prompt

0.1 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.

0.12 - None. None. None.
0.12 - None. None. This will be fun!
0.12 - None. None. Take a deep breath and think carefully.
0.12 - None. None. I really need your help!
0.12 - None. Solve the following math problem. None.
0.12 - None. Solve the following math problem. This will be fun!
0.12 - None. Solve the following math problem. Take a deep breath and think carefully.
0.12 - None. Solve the following math problem. I really need your help!
0.12 - None. Answer the following math question. None.
0.12 - None. Answer the following math question. This will be fun!
0.12 - None. Answer the following math question. Take a deep breath and think carefully.
0.12 - None. Answer the following math question. I really need your help!
0.12 - You are as smart as ChatGPT. None. None.
0.12 - You are as smart as ChatGPT. None. This will be fun!
0.12 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.12 - You are as smart as ChatGPT. None. I really need your help!
0.12 - You are as smart as ChatGPT. Solve the following math problem. None.
0.12 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.12 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.12 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.12 - You are as smart as ChatGPT. Answer the following math question. None.
0.12 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.12 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.12 - You are highly intelligent. None. None.
0.12 - You are highly intelligent. None. This will be fun!
0.12 - You are highly intelligent. None. Take a deep breath and think carefully.
0.12 - You are highly intelligent. None. I really need your help!
0.12 - You are highly intelligent. Solve the following math problem. None.
0.12 - You are highly intelligent. Solve the following math problem. This will be fun!
0.12 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.12 - You are highly intelligent. Solve the following math problem. I really need your help!
0.12 - You are highly intelligent. Answer the following math question. None.
0.12 - You are highly intelligent. Answer the following math question. This will be fun!
0.12 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are highly intelligent. Answer the following math question. I really need your help!
0.12 - You are an expert mathematician. None. None.
0.12 - You are an expert mathematician. None. This will be fun!
0.12 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.12 - You are an expert mathematician. None. I really need your help!
0.12 - You are an expert mathematician. Solve the following math problem. None.
0.12 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.12 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.12 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.12 - You are an expert mathematician. Answer the following math question. None.
0.12 - You are an expert mathematician. Answer the following math question. This will be fun!
0.12 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are an expert mathematician. Answer the following math question. I really need your help!
0.12 - You are a professor of mathematics. None. None.
0.12 - You are a professor of mathematics. None. This will be fun!
0.12 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.12 - You are a professor of mathematics. None. I really need your help!
0.12 - You are a professor of mathematics. Solve the following math problem. None.
0.12 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.12 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.12 - You are a professor of mathematics. Answer the following math question. None.
0.12 - You are a professor of mathematics. Answer the following math question. This will be fun!

0.12 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.4 Mistral-7B CoT=No NoQ=100

EM - Prompt

0.08 - None. None. Take a deep breath and think carefully.
0.09 - None. None. None.
0.09 - You are highly intelligent. None. None.
0.09 - You are highly intelligent. None. This will be fun!
0.09 - You are highly intelligent. None. Take a deep breath and think carefully.
0.09 - You are highly intelligent. None. I really need your help!
0.1 - None. None. This will be fun!
0.1 - None. None. I really need your help!
0.1 - None. Solve the following math problem. None.
0.1 - None. Solve the following math problem. I really need your help!
0.1 - None. Answer the following math question. None.
0.1 - None. Answer the following math question. This will be fun!
0.1 - None. Answer the following math question. Take a deep breath and think carefully.
0.1 - None. Answer the following math question. I really need your help!
0.1 - You are as smart as ChatGPT. None. None.
0.1 - You are as smart as ChatGPT. None. This will be fun!
0.1 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. None. I really need your help!
0.1 - You are highly intelligent. Solve the following math problem. This will be fun!
0.1 - You are highly intelligent. Answer the following math question. This will be fun!
0.1 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.11 - None. Solve the following math problem. This will be fun!
0.11 - None. Solve the following math problem. Take a deep breath and think carefully.
0.11 - You are as smart as ChatGPT. Solve the following math problem. None.
0.11 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.11 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.11 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.11 - You are as smart as ChatGPT. Answer the following math question. None.
0.11 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.11 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.11 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.11 - You are highly intelligent. Solve the following math problem. None.
0.11 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.11 - You are highly intelligent. Solve the following math problem. I really need your help!
0.11 - You are highly intelligent. Answer the following math question. None.
0.11 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.11 - You are highly intelligent. Answer the following math question. I really need your help!
0.11 - You are an expert mathematician. None. None.
0.11 - You are an expert mathematician. None. This will be fun!
0.11 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.11 - You are an expert mathematician. None. I really need your help!
0.11 - You are an expert mathematician. Solve the following math problem. None.
0.11 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.11 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.11 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.11 - You are an expert mathematician. Answer the following math question. None.
0.11 - You are an expert mathematician. Answer the following math question. This will be fun!
0.11 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.11 - You are an expert mathematician. Answer the following math question. I really need your help!
0.11 - You are a professor of mathematics. None. None.
0.11 - You are a professor of mathematics. None. This will be fun!

0.11 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.11 - You are a professor of mathematics. None. I really need your help!
0.11 - You are a professor of mathematics. Solve the following math problem. None.
0.11 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.11 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.11 - You are a professor of mathematics. Answer the following math question. None.
0.11 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.11 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.11 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.5 Mistral-7B CoT=Yes NoQ=10

EM - Prompt

0.2 - None. None. None.
0.2 - You are highly intelligent. Solve the following math problem. This will be fun!
0.3 - None. Answer the following math question. None.
0.3 - None. Answer the following math question. This will be fun!
0.3 - None. Answer the following math question. Take a deep breath and think carefully.
0.3 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.3 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.3 - You are highly intelligent. Answer the following math question. None.
0.3 - You are highly intelligent. Answer the following math question. This will be fun!
0.3 - You are highly intelligent. Answer the following math question. I really need your help!
0.3 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.3 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.3 - You are an expert mathematician. Answer the following math question. This will be fun!
0.3 - You are an expert mathematician. Answer the following math question. I really need your help!
0.3 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.3 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.4 - None. None. Take a deep breath and think carefully.
0.4 - None. None. I really need your help!
0.4 - None. Solve the following math problem. None.
0.4 - None. Solve the following math problem. This will be fun!
0.4 - None. Solve the following math problem. Take a deep breath and think carefully.
0.4 - None. Solve the following math problem. I really need your help!
0.4 - You are as smart as ChatGPT. None. This will be fun!
0.4 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.4 - You are as smart as ChatGPT. None. I really need your help!
0.4 - You are as smart as ChatGPT. Solve the following math problem. None.
0.4 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.4 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.4 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.4 - You are highly intelligent. None. None.
0.4 - You are highly intelligent. None. This will be fun!
0.4 - You are highly intelligent. None. Take a deep breath and think carefully.
0.4 - You are highly intelligent. None. I really need your help!
0.4 - You are highly intelligent. Solve the following math problem. None.
0.4 - You are highly intelligent. Solve the following math problem. I really need your help!
0.4 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.4 - You are an expert mathematician. None. None.
0.4 - You are an expert mathematician. None. This will be fun!
0.4 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.4 - You are an expert mathematician. None. I really need your help!
0.4 - You are an expert mathematician. Solve the following math problem. None.
0.4 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.4 - You are an expert mathematician. Answer the following math question. None.
0.4 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.

0.4 - You are a professor of mathematics. None. None.
0.4 - You are a professor of mathematics. None. This will be fun!
0.4 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.4 - You are a professor of mathematics. None. I really need your help!
0.4 - You are a professor of mathematics. Solve the following math problem. None.
0.4 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.4 - You are a professor of mathematics. Answer the following math question. None.
0.4 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.4 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.4 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.5 - None. None. This will be fun!
0.5 - None. Answer the following math question. I really need your help!
0.5 - You are as smart as ChatGPT. None. None.
0.5 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.5 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.5 - You are as smart as ChatGPT. Answer the following math question. None.

B.6 Mistral-7B CoT=Yes NoQ=25

EM - Prompt

0.28 - None. None. None.
0.28 - You are highly intelligent. Solve the following math problem. This will be fun!
0.28 - You are highly intelligent. Answer the following math question. This will be fun!
0.28 - You are highly intelligent. Answer the following math question. I really need your help!
0.32 - None. None. I really need your help!
0.32 - None. Answer the following math question. This will be fun!
0.32 - None. Answer the following math question. Take a deep breath and think carefully.
0.32 - You are as smart as ChatGPT. None. This will be fun!
0.32 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.32 - You are as smart as ChatGPT. None. I really need your help!
0.32 - You are as smart as ChatGPT. Solve the following math problem. None.
0.32 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.32 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.32 - You are highly intelligent. None. This will be fun!
0.32 - You are highly intelligent. None. I really need your help!
0.32 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.32 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.32 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.32 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.36 - None. None. Take a deep breath and think carefully.
0.36 - None. Solve the following math problem. This will be fun!
0.36 - None. Solve the following math problem. I really need your help!
0.36 - None. Answer the following math question. None.
0.36 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.36 - You are highly intelligent. None. None.
0.36 - You are highly intelligent. None. Take a deep breath and think carefully.
0.36 - You are highly intelligent. Solve the following math problem. None.
0.36 - You are highly intelligent. Solve the following math problem. I really need your help!
0.36 - You are highly intelligent. Answer the following math question. None.
0.36 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.36 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.36 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.36 - You are a professor of mathematics. None. I really need your help!
0.36 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.4 - None. None. This will be fun!
0.4 - None. Solve the following math problem. None.
0.4 - None. Solve the following math problem. Take a deep breath and think carefully.

0.4 - You are as smart as ChatGPT. None. None.
0.4 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.4 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.4 - You are as smart as ChatGPT. Answer the following math question. None.
0.4 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.4 - You are an expert mathematician. None. None.
0.4 - You are an expert mathematician. None. I really need your help!
0.4 - You are an expert mathematician. Solve the following math problem. None.
0.4 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.4 - You are an expert mathematician. Answer the following math question. None.
0.4 - You are an expert mathematician. Answer the following math question. This will be fun!
0.4 - You are an expert mathematician. Answer the following math question. I really need your help!
0.4 - You are a professor of mathematics. None. This will be fun!
0.4 - You are a professor of mathematics. Solve the following math problem. None.
0.4 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.4 - You are a professor of mathematics. Answer the following math question. None.
0.4 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.4 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.44 - None. Answer the following math question. I really need your help!
0.44 - You are an expert mathematician. None. This will be fun!
0.44 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.44 - You are a professor of mathematics. None. None.
0.48 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.

B.7 Mistral-7B CoT=Yes NoQ=50

EM - Prompt

0.32 - None. None. None.
0.34 - None. Answer the following math question. None.
0.34 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.36 - None. None. I really need your help!
0.36 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.36 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.36 - You are highly intelligent. Answer the following math question. This will be fun!
0.36 - You are highly intelligent. Answer the following math question. I really need your help!
0.36 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.36 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.36 - You are a professor of mathematics. Solve the following math problem. None.
0.36 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.38 - None. Answer the following math question. This will be fun!
0.38 - None. Answer the following math question. Take a deep breath and think carefully.
0.38 - You are as smart as ChatGPT. None. This will be fun!
0.38 - You are as smart as ChatGPT. None. I really need your help!
0.38 - You are as smart as ChatGPT. Solve the following math problem. None.
0.38 - You are highly intelligent. None. This will be fun!
0.38 - You are highly intelligent. None. I really need your help!
0.38 - You are highly intelligent. Solve the following math problem. This will be fun!
0.38 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.38 - You are highly intelligent. Solve the following math problem. I really need your help!
0.38 - You are highly intelligent. Answer the following math question. None.
0.38 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.38 - You are an expert mathematician. Solve the following math problem. None.
0.38 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.38 - You are an expert mathematician. Answer the following math question. None.
0.38 - You are an expert mathematician. Answer the following math question. This will be fun!
0.38 - You are a professor of mathematics. None. I really need your help!
0.38 - You are a professor of mathematics. Solve the following math problem. This will be fun!

0.38 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.38 - You are a professor of mathematics. Answer the following math question. None.
0.4 - None. Solve the following math problem. None.
0.4 - None. Solve the following math problem. This will be fun!
0.4 - None. Solve the following math problem. Take a deep breath and think carefully.
0.4 - None. Solve the following math problem. I really need your help!
0.4 - None. Answer the following math question. I really need your help!
0.4 - You are as smart as ChatGPT. None. None.
0.4 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.4 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.4 - You are highly intelligent. None. Take a deep breath and think carefully.
0.4 - You are highly intelligent. Solve the following math problem. None.
0.4 - You are an expert mathematician. None. None.
0.4 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.4 - You are an expert mathematician. None. I really need your help!
0.4 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.4 - You are an expert mathematician. Answer the following math question. I really need your help!
0.4 - You are a professor of mathematics. None. This will be fun!
0.4 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.42 - None. None. This will be fun!
0.42 - None. None. Take a deep breath and think carefully.
0.42 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.42 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.42 - You are highly intelligent. None. None.
0.42 - You are an expert mathematician. None. This will be fun!
0.42 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.44 - You are as smart as ChatGPT. Answer the following math question. None.
0.44 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.44 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.44 - You are a professor of mathematics. None. None.

B.8 Mistral-7B CoT=Yes NoQ=100

EM - Prompt

0.35 - None. None. None.
0.37 - None. None. I really need your help!
0.37 - You are highly intelligent. Answer the following math question. This will be fun!
0.37 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.38 - You are as smart as ChatGPT. None. This will be fun!
0.38 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.38 - You are highly intelligent. None. I really need your help!
0.38 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.39 - None. None. Take a deep breath and think carefully.
0.39 - None. Answer the following math question. None.
0.39 - None. Answer the following math question. I really need your help!
0.39 - You are as smart as ChatGPT. None. I really need your help!
0.39 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.39 - You are highly intelligent. Solve the following math problem. I really need your help!
0.39 - You are highly intelligent. Answer the following math question. None.
0.39 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.39 - You are an expert mathematician. Answer the following math question. This will be fun!
0.39 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.39 - You are a professor of mathematics. Answer the following math question. None.
0.4 - None. Solve the following math problem. This will be fun!
0.4 - None. Solve the following math problem. I really need your help!
0.4 - You are as smart as ChatGPT. None. None.
0.4 - You are highly intelligent. Solve the following math problem. This will be fun!

0.4 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.4 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.4 - You are an expert mathematician. Solve the following math problem. None.
0.4 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.4 - You are an expert mathematician. Answer the following math question. None.
0.4 - You are an expert mathematician. Answer the following math question. I really need your help!
0.4 - You are a professor of mathematics. None. This will be fun!
0.4 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.4 - You are a professor of mathematics. None. I really need your help!
0.4 - You are a professor of mathematics. Solve the following math problem. None.
0.4 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.4 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.41 - None. Solve the following math problem. None.
0.41 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.41 - You are as smart as ChatGPT. Solve the following math problem. None.
0.41 - You are highly intelligent. None. This will be fun!
0.41 - You are highly intelligent. None. Take a deep breath and think carefully.
0.41 - You are highly intelligent. Answer the following math question. I really need your help!
0.41 - You are an expert mathematician. None. I really need your help!
0.41 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.42 - None. Solve the following math problem. Take a deep breath and think carefully.
0.42 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.42 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.42 - You are highly intelligent. None. None.
0.42 - You are highly intelligent. Solve the following math problem. None.
0.42 - You are an expert mathematician. None. None.
0.42 - You are an expert mathematician. None. This will be fun!
0.42 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.42 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.43 - None. None. This will be fun!
0.43 - None. Answer the following math question. This will be fun!
0.43 - None. Answer the following math question. Take a deep breath and think carefully.
0.43 - You are as smart as ChatGPT. Answer the following math question. None.
0.43 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.43 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.43 - You are a professor of mathematics. None. None.
0.44 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.

B.9 Llama2-13B CoT=No NoQ=10

EM - Prompt

0.1 - None. None. None.
0.1 - None. None. This will be fun!
0.1 - None. None. Take a deep breath and think carefully.
0.1 - None. None. I really need your help!
0.1 - None. Solve the following math problem. None.
0.1 - None. Solve the following math problem. This will be fun!
0.1 - None. Solve the following math problem. Take a deep breath and think carefully.
0.1 - None. Solve the following math problem. I really need your help!
0.1 - None. Answer the following math question. None.
0.1 - None. Answer the following math question. This will be fun!
0.1 - None. Answer the following math question. Take a deep breath and think carefully.
0.1 - None. Answer the following math question. I really need your help!
0.1 - You are as smart as ChatGPT. None. None.
0.1 - You are as smart as ChatGPT. None. This will be fun!
0.1 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. None. I really need your help!

0.1 - You are as smart as ChatGPT. Solve the following math problem. None.
0.1 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.1 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.1 - You are as smart as ChatGPT. Answer the following math question. None.
0.1 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.1 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.1 - You are highly intelligent. None. None.
0.1 - You are highly intelligent. None. This will be fun!
0.1 - You are highly intelligent. None. Take a deep breath and think carefully.
0.1 - You are highly intelligent. None. I really need your help!
0.1 - You are highly intelligent. Solve the following math problem. None.
0.1 - You are highly intelligent. Solve the following math problem. This will be fun!
0.1 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are highly intelligent. Solve the following math problem. I really need your help!
0.1 - You are highly intelligent. Answer the following math question. None.
0.1 - You are highly intelligent. Answer the following math question. This will be fun!
0.1 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are highly intelligent. Answer the following math question. I really need your help!
0.1 - You are an expert mathematician. None. None.
0.1 - You are an expert mathematician. None. This will be fun!
0.1 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. None. I really need your help!
0.1 - You are an expert mathematician. Solve the following math problem. None.
0.1 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.1 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.1 - You are an expert mathematician. Answer the following math question. None.
0.1 - You are an expert mathematician. Answer the following math question. This will be fun!
0.1 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. Answer the following math question. I really need your help!
0.1 - You are a professor of mathematics. None. None.
0.1 - You are a professor of mathematics. None. This will be fun!
0.1 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. None. I really need your help!
0.1 - You are a professor of mathematics. Solve the following math problem. None.
0.1 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.1 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.1 - You are a professor of mathematics. Answer the following math question. None.
0.1 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.1 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.10 Llama2-13B CoT=No NoQ=25

EM - Prompt

0.08 - None. None. None.
0.08 - None. None. This will be fun!
0.08 - None. None. Take a deep breath and think carefully.
0.08 - None. None. I really need your help!
0.08 - None. Solve the following math problem. None.
0.08 - None. Solve the following math problem. This will be fun!
0.08 - None. Solve the following math problem. Take a deep breath and think carefully.
0.08 - None. Solve the following math problem. I really need your help!
0.08 - None. Answer the following math question. None.

0.08 - None. Answer the following math question. This will be fun!
0.08 - None. Answer the following math question. Take a deep breath and think carefully.
0.08 - None. Answer the following math question. I really need your help!
0.08 - You are as smart as ChatGPT. None. None.
0.08 - You are as smart as ChatGPT. None. This will be fun!
0.08 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. None. I really need your help!
0.08 - You are as smart as ChatGPT. Solve the following math problem. None.
0.08 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.08 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. Answer the following math question. None.
0.08 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.08 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.08 - You are highly intelligent. None. None.
0.08 - You are highly intelligent. None. This will be fun!
0.08 - You are highly intelligent. None. Take a deep breath and think carefully.
0.08 - You are highly intelligent. None. I really need your help!
0.08 - You are highly intelligent. Solve the following math problem. None.
0.08 - You are highly intelligent. Solve the following math problem. This will be fun!
0.08 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are highly intelligent. Answer the following math question. None.
0.08 - You are highly intelligent. Answer the following math question. This will be fun!
0.08 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. None. None.
0.08 - You are an expert mathematician. None. This will be fun!
0.08 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. Solve the following math problem. None.
0.08 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.08 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. Answer the following math question. None.
0.08 - You are an expert mathematician. Answer the following math question. This will be fun!
0.08 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. None. None.
0.08 - You are a professor of mathematics. None. This will be fun!
0.08 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. None. I really need your help!
0.08 - You are a professor of mathematics. Solve the following math problem. None.
0.08 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.08 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. Answer the following math question. None.
0.08 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.08 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.12 - You are highly intelligent. Solve the following math problem. I really need your help!
0.12 - You are highly intelligent. Answer the following math question. I really need your help!
0.12 - You are an expert mathematician. None. I really need your help!
0.12 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.12 - You are an expert mathematician. Answer the following math question. I really need your help!
0.12 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.12 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.11 Llama2-13B CoT=No NoQ=50

EM - Prompt

0.08 - None. None. None.
0.08 - None. None. This will be fun!

0.08 - None. None. Take a deep breath and think carefully.
0.08 - None. None. I really need your help!
0.08 - None. Solve the following math problem. None.
0.08 - None. Solve the following math problem. This will be fun!
0.08 - None. Solve the following math problem. Take a deep breath and think carefully.
0.08 - None. Solve the following math problem. I really need your help!
0.08 - None. Answer the following math question. None.
0.08 - None. Answer the following math question. This will be fun!
0.08 - None. Answer the following math question. Take a deep breath and think carefully.
0.08 - None. Answer the following math question. I really need your help!
0.08 - You are as smart as ChatGPT. None. None.
0.08 - You are as smart as ChatGPT. None. This will be fun!
0.08 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. None. I really need your help!
0.08 - You are as smart as ChatGPT. Solve the following math problem. None.
0.08 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.08 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. Answer the following math question. None.
0.08 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.08 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.08 - You are highly intelligent. None. None.
0.08 - You are highly intelligent. None. This will be fun!
0.08 - You are highly intelligent. None. Take a deep breath and think carefully.
0.08 - You are highly intelligent. None. I really need your help!
0.08 - You are highly intelligent. Solve the following math problem. None.
0.08 - You are highly intelligent. Solve the following math problem. This will be fun!
0.08 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are highly intelligent. Answer the following math question. None.
0.08 - You are highly intelligent. Answer the following math question. This will be fun!
0.08 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. None. None.
0.08 - You are an expert mathematician. None. This will be fun!
0.08 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. Solve the following math problem. None.
0.08 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.08 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are an expert mathematician. Answer the following math question. None.
0.08 - You are an expert mathematician. Answer the following math question. This will be fun!
0.08 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. None. None.
0.08 - You are a professor of mathematics. None. This will be fun!
0.08 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. None. I really need your help!
0.08 - You are a professor of mathematics. Solve the following math problem. None.
0.08 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.08 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.08 - You are a professor of mathematics. Answer the following math question. None.
0.08 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.08 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.1 - You are highly intelligent. Solve the following math problem. I really need your help!
0.1 - You are highly intelligent. Answer the following math question. I really need your help!
0.1 - You are an expert mathematician. None. I really need your help!
0.1 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.1 - You are an expert mathematician. Answer the following math question. I really need your help!
0.1 - You are a professor of mathematics. Solve the following math problem. I really need your help!

0.1 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.12 Llama2-13B CoT=No NoQ=100

EM - Prompt

0.07 - None. None. None.
0.07 - None. None. This will be fun!
0.07 - None. None. Take a deep breath and think carefully.
0.07 - None. None. I really need your help!
0.07 - None. Solve the following math problem. None.
0.07 - None. Solve the following math problem. This will be fun!
0.07 - None. Solve the following math problem. Take a deep breath and think carefully.
0.07 - None. Solve the following math problem. I really need your help!
0.07 - None. Answer the following math question. None.
0.07 - None. Answer the following math question. This will be fun!
0.07 - None. Answer the following math question. Take a deep breath and think carefully.
0.07 - None. Answer the following math question. I really need your help!
0.07 - You are as smart as ChatGPT. None. None.
0.07 - You are as smart as ChatGPT. None. This will be fun!
0.07 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.07 - You are as smart as ChatGPT. None. I really need your help!
0.07 - You are as smart as ChatGPT. Solve the following math problem. None.
0.07 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.07 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.07 - You are as smart as ChatGPT. Answer the following math question. None.
0.07 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.07 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.07 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.07 - You are highly intelligent. None. None.
0.07 - You are highly intelligent. None. This will be fun!
0.07 - You are highly intelligent. None. Take a deep breath and think carefully.
0.07 - You are highly intelligent. None. I really need your help!
0.07 - You are highly intelligent. Solve the following math problem. None.
0.07 - You are highly intelligent. Solve the following math problem. This will be fun!
0.07 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.07 - You are highly intelligent. Answer the following math question. None.
0.07 - You are highly intelligent. Answer the following math question. This will be fun!
0.07 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.07 - You are an expert mathematician. None. None.
0.07 - You are an expert mathematician. None. This will be fun!
0.07 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.07 - You are an expert mathematician. Solve the following math problem. None.
0.07 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.07 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.07 - You are an expert mathematician. Answer the following math question. None.
0.07 - You are an expert mathematician. Answer the following math question. This will be fun!
0.07 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.07 - You are a professor of mathematics. None. None.
0.07 - You are a professor of mathematics. None. This will be fun!
0.07 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.07 - You are a professor of mathematics. None. I really need your help!
0.07 - You are a professor of mathematics. Solve the following math problem. None.
0.07 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.07 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.07 - You are a professor of mathematics. Answer the following math question. None.
0.07 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.07 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.

0.08 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.08 - You are highly intelligent. Solve the following math problem. I really need your help!
0.08 - You are highly intelligent. Answer the following math question. I really need your help!
0.08 - You are an expert mathematician. None. I really need your help!
0.08 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.08 - You are an expert mathematician. Answer the following math question. I really need your help!
0.08 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.08 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.13 Llama2-13B CoT=Yes NoQ=10

EM - Prompt

0.3 - None. None. Take a deep breath and think carefully.
0.3 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.3 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.4 - None. None. None.
0.4 - None. None. This will be fun!
0.4 - None. None. I really need your help!
0.4 - None. Solve the following math problem. None.
0.4 - None. Solve the following math problem. This will be fun!
0.4 - None. Solve the following math problem. Take a deep breath and think carefully.
0.4 - None. Solve the following math problem. I really need your help!
0.4 - None. Answer the following math question. This will be fun!
0.4 - None. Answer the following math question. Take a deep breath and think carefully.
0.4 - None. Answer the following math question. I really need your help!
0.4 - You are as smart as ChatGPT. None. None.
0.4 - You are as smart as ChatGPT. None. This will be fun!
0.4 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.4 - You are as smart as ChatGPT. None. I really need your help!
0.4 - You are as smart as ChatGPT. Solve the following math problem. None.
0.4 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.4 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.4 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.4 - You are as smart as ChatGPT. Answer the following math question. None.
0.4 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.4 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.4 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.4 - You are highly intelligent. None. None.
0.4 - You are highly intelligent. None. This will be fun!
0.4 - You are highly intelligent. None. Take a deep breath and think carefully.
0.4 - You are highly intelligent. None. I really need your help!
0.4 - You are highly intelligent. Solve the following math problem. None.
0.4 - You are highly intelligent. Solve the following math problem. This will be fun!
0.4 - You are highly intelligent. Solve the following math problem. I really need your help!
0.4 - You are highly intelligent. Answer the following math question. None.
0.4 - You are highly intelligent. Answer the following math question. This will be fun!
0.4 - You are highly intelligent. Answer the following math question. I really need your help!
0.4 - You are an expert mathematician. None. None.
0.4 - You are an expert mathematician. None. This will be fun!
0.4 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.4 - You are an expert mathematician. None. I really need your help!
0.4 - You are an expert mathematician. Solve the following math problem. None.
0.4 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.4 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.4 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.4 - You are an expert mathematician. Answer the following math question. None.
0.4 - You are an expert mathematician. Answer the following math question. This will be fun!

0.4 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.4 - You are an expert mathematician. Answer the following math question. I really need your help!
0.4 - You are a professor of mathematics. None. None.
0.4 - You are a professor of mathematics. None. This will be fun!
0.4 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.4 - You are a professor of mathematics. None. I really need your help!
0.4 - You are a professor of mathematics. Solve the following math problem. None.
0.4 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.4 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.4 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.4 - You are a professor of mathematics. Answer the following math question. None.
0.4 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.4 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.4 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.5 - None. Answer the following math question. None.

B.14 Llama2-13B CoT=Yes NoQ=25

EM - Prompt

0.4 - None. None. Take a deep breath and think carefully.
0.4 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.4 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.44 - None. None. None.
0.44 - None. None. This will be fun!
0.44 - None. None. I really need your help!
0.44 - None. Solve the following math problem. This will be fun!
0.44 - None. Solve the following math problem. Take a deep breath and think carefully.
0.44 - None. Solve the following math problem. I really need your help!
0.44 - You are as smart as ChatGPT. None. This will be fun!
0.44 - You are as smart as ChatGPT. None. I really need your help!
0.44 - You are as smart as ChatGPT. Solve the following math problem. None.
0.44 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.44 - You are highly intelligent. None. This will be fun!
0.44 - You are highly intelligent. None. Take a deep breath and think carefully.
0.44 - You are highly intelligent. None. I really need your help!
0.44 - You are highly intelligent. Solve the following math problem. None.
0.44 - You are highly intelligent. Solve the following math problem. This will be fun!
0.44 - You are highly intelligent. Solve the following math problem. I really need your help!
0.44 - You are highly intelligent. Answer the following math question. This will be fun!
0.44 - You are highly intelligent. Answer the following math question. I really need your help!
0.44 - You are an expert mathematician. None. This will be fun!
0.44 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.44 - You are an expert mathematician. None. I really need your help!
0.44 - You are an expert mathematician. Solve the following math problem. None.
0.44 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.44 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.44 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.44 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.44 - You are an expert mathematician. Answer the following math question. I really need your help!
0.44 - You are a professor of mathematics. None. This will be fun!
0.44 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.44 - You are a professor of mathematics. None. I really need your help!
0.44 - You are a professor of mathematics. Solve the following math problem. None.
0.44 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.44 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.44 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.44 - You are a professor of mathematics. Answer the following math question. None.

0.44 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.44 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.44 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.48 - None. Solve the following math problem. None.
0.48 - None. Answer the following math question. This will be fun!
0.48 - None. Answer the following math question. Take a deep breath and think carefully.
0.48 - None. Answer the following math question. I really need your help!
0.48 - You are as smart as ChatGPT. None. None.
0.48 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.48 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.48 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.48 - You are as smart as ChatGPT. Answer the following math question. None.
0.48 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.48 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.48 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.48 - You are highly intelligent. None. None.
0.48 - You are highly intelligent. Answer the following math question. None.
0.48 - You are an expert mathematician. None. None.
0.48 - You are an expert mathematician. Answer the following math question. None.
0.48 - You are an expert mathematician. Answer the following math question. This will be fun!
0.48 - You are a professor of mathematics. None. None.
0.52 - None. Answer the following math question. None.

B.15 Llama2-13B CoT=Yes NoQ=50

EM - Prompt

0.44 - None. None. Take a deep breath and think carefully.
0.44 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.44 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.46 - None. None. None.
0.46 - None. None. This will be fun!
0.46 - None. None. I really need your help!
0.46 - None. Solve the following math problem. This will be fun!
0.46 - None. Solve the following math problem. Take a deep breath and think carefully.
0.46 - None. Solve the following math problem. I really need your help!
0.46 - You are as smart as ChatGPT. None. This will be fun!
0.46 - You are as smart as ChatGPT. None. I really need your help!
0.46 - You are as smart as ChatGPT. Solve the following math problem. None.
0.46 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.46 - You are highly intelligent. None. This will be fun!
0.46 - You are highly intelligent. None. Take a deep breath and think carefully.
0.46 - You are highly intelligent. None. I really need your help!
0.46 - You are highly intelligent. Solve the following math problem. None.
0.46 - You are highly intelligent. Solve the following math problem. This will be fun!
0.46 - You are highly intelligent. Solve the following math problem. I really need your help!
0.46 - You are highly intelligent. Answer the following math question. This will be fun!
0.46 - You are highly intelligent. Answer the following math question. I really need your help!
0.46 - You are an expert mathematician. None. This will be fun!
0.46 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.46 - You are an expert mathematician. None. I really need your help!
0.46 - You are an expert mathematician. Solve the following math problem. None.
0.46 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.46 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.46 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.46 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.46 - You are an expert mathematician. Answer the following math question. I really need your help!
0.46 - You are a professor of mathematics. None. This will be fun!

0.46 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.46 - You are a professor of mathematics. None. I really need your help!
0.46 - You are a professor of mathematics. Solve the following math problem. None.
0.46 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.46 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.46 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.46 - You are a professor of mathematics. Answer the following math question. None.
0.46 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.46 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.46 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.48 - None. Solve the following math problem. None.
0.48 - None. Answer the following math question. This will be fun!
0.48 - None. Answer the following math question. Take a deep breath and think carefully.
0.48 - None. Answer the following math question. I really need your help!
0.48 - You are as smart as ChatGPT. None. None.
0.48 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.48 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.48 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.48 - You are as smart as ChatGPT. Answer the following math question. None.
0.48 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.48 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.48 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.48 - You are highly intelligent. None. None.
0.48 - You are highly intelligent. Answer the following math question. None.
0.48 - You are an expert mathematician. None. None.
0.48 - You are an expert mathematician. Answer the following math question. None.
0.48 - You are an expert mathematician. Answer the following math question. This will be fun!
0.48 - You are a professor of mathematics. None. None.
0.5 - None. Answer the following math question. None.

B.16 Llama2-13B CoT=Yes NoQ=100

EM - Prompt

0.41 - None. Solve the following math problem. I really need your help!
0.42 - None. None. Take a deep breath and think carefully.
0.43 - None. Answer the following math question. This will be fun!
0.43 - You are highly intelligent. None. Take a deep breath and think carefully.
0.43 - You are highly intelligent. Answer the following math question. None.
0.43 - You are an expert mathematician. Solve the following math problem. None.
0.44 - None. Solve the following math problem. None.
0.44 - None. Answer the following math question. Take a deep breath and think carefully.
0.44 - None. Answer the following math question. I really need your help!
0.44 - You are as smart as ChatGPT. None. I really need your help!
0.44 - You are as smart as ChatGPT. Solve the following math problem. None.
0.44 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.44 - You are highly intelligent. Solve the following math problem. None.
0.44 - You are highly intelligent. Solve the following math problem. I really need your help!
0.44 - You are highly intelligent. Answer the following math question. I really need your help!
0.44 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.44 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.44 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.44 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.45 - None. None. This will be fun!
0.45 - None. None. I really need your help!
0.45 - None. Solve the following math problem. This will be fun!
0.45 - None. Solve the following math problem. Take a deep breath and think carefully.
0.45 - You are as smart as ChatGPT. None. None.

0.45 - You are as smart as ChatGPT. None. This will be fun!
0.45 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.45 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.45 - You are highly intelligent. Answer the following math question. This will be fun!
0.45 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.45 - You are an expert mathematician. None. I really need your help!
0.45 - You are a professor of mathematics. Solve the following math problem. None.
0.46 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.46 - You are highly intelligent. None. This will be fun!
0.46 - You are highly intelligent. None. I really need your help!
0.46 - You are highly intelligent. Solve the following math problem. This will be fun!
0.46 - You are an expert mathematician. None. This will be fun!
0.46 - You are an expert mathematician. Answer the following math question. I really need your help!
0.46 - You are a professor of mathematics. Answer the following math question. None.
0.46 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.46 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.47 - None. None. None.
0.47 - None. Answer the following math question. None.
0.47 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.47 - You are as smart as ChatGPT. Answer the following math question. None.
0.47 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.47 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.47 - You are highly intelligent. None. None.
0.47 - You are an expert mathematician. None. None.
0.47 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.47 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.47 - You are an expert mathematician. Answer the following math question. None.
0.47 - You are an expert mathematician. Answer the following math question. This will be fun!
0.47 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.47 - You are a professor of mathematics. None. None.
0.47 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.47 - You are a professor of mathematics. None. I really need your help!
0.47 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.48 - You are a professor of mathematics. None. This will be fun!
0.48 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.49 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.

B.17 Llama2-70B CoT=No NoQ=10

EM - Prompt

0.1 - None. None. None.
0.1 - None. None. This will be fun!
0.1 - None. None. Take a deep breath and think carefully.
0.1 - None. None. I really need your help!
0.1 - None. Solve the following math problem. None.
0.1 - None. Solve the following math problem. This will be fun!
0.1 - None. Solve the following math problem. Take a deep breath and think carefully.
0.1 - None. Solve the following math problem. I really need your help!
0.1 - None. Answer the following math question. None.
0.1 - None. Answer the following math question. This will be fun!
0.1 - None. Answer the following math question. Take a deep breath and think carefully.
0.1 - None. Answer the following math question. I really need your help!
0.1 - You are as smart as ChatGPT. None. None.
0.1 - You are as smart as ChatGPT. None. This will be fun!
0.1 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. None. I really need your help!
0.1 - You are as smart as ChatGPT. Solve the following math problem. None.

0.1 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.1 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.1 - You are as smart as ChatGPT. Answer the following math question. None.
0.1 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.1 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.1 - You are highly intelligent. None. None.
0.1 - You are highly intelligent. None. This will be fun!
0.1 - You are highly intelligent. None. Take a deep breath and think carefully.
0.1 - You are highly intelligent. None. I really need your help!
0.1 - You are highly intelligent. Solve the following math problem. None.
0.1 - You are highly intelligent. Solve the following math problem. This will be fun!
0.1 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are highly intelligent. Solve the following math problem. I really need your help!
0.1 - You are highly intelligent. Answer the following math question. None.
0.1 - You are highly intelligent. Answer the following math question. This will be fun!
0.1 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are highly intelligent. Answer the following math question. I really need your help!
0.1 - You are an expert mathematician. None. None.
0.1 - You are an expert mathematician. None. This will be fun!
0.1 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. None. I really need your help!
0.1 - You are an expert mathematician. Solve the following math problem. None.
0.1 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.1 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.1 - You are an expert mathematician. Answer the following math question. None.
0.1 - You are an expert mathematician. Answer the following math question. This will be fun!
0.1 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are an expert mathematician. Answer the following math question. I really need your help!
0.1 - You are a professor of mathematics. None. None.
0.1 - You are a professor of mathematics. None. This will be fun!
0.1 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. None. I really need your help!
0.1 - You are a professor of mathematics. Solve the following math problem. None.
0.1 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.1 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.1 - You are a professor of mathematics. Answer the following math question. None.
0.1 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.1 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.1 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.18 Llama2-70B CoT=No NoQ=25

EM - Prompt

0.12 - None. Solve the following math problem. This will be fun!
0.12 - None. Solve the following math problem. Take a deep breath and think carefully.
0.12 - None. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are as smart as ChatGPT. None. None.
0.12 - You are as smart as ChatGPT. None. This will be fun!
0.12 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.12 - You are as smart as ChatGPT. None. I really need your help!
0.12 - You are as smart as ChatGPT. Solve the following math problem. None.
0.12 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.12 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.

0.12 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.12 - You are as smart as ChatGPT. Answer the following math question. None.
0.12 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.12 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.12 - You are highly intelligent. None. This will be fun!
0.12 - You are highly intelligent. None. Take a deep breath and think carefully.
0.12 - You are highly intelligent. None. I really need your help!
0.12 - You are highly intelligent. Solve the following math problem. None.
0.12 - You are highly intelligent. Solve the following math problem. This will be fun!
0.12 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.12 - You are highly intelligent. Solve the following math problem. I really need your help!
0.12 - You are highly intelligent. Answer the following math question. None.
0.12 - You are highly intelligent. Answer the following math question. This will be fun!
0.12 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are highly intelligent. Answer the following math question. I really need your help!
0.12 - You are an expert mathematician. None. None.
0.12 - You are an expert mathematician. None. This will be fun!
0.12 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.12 - You are an expert mathematician. None. I really need your help!
0.12 - You are an expert mathematician. Solve the following math problem. None.
0.12 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.12 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.12 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.12 - You are an expert mathematician. Answer the following math question. None.
0.12 - You are an expert mathematician. Answer the following math question. This will be fun!
0.12 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are an expert mathematician. Answer the following math question. I really need your help!
0.12 - You are a professor of mathematics. None. None.
0.12 - You are a professor of mathematics. None. This will be fun!
0.12 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.12 - You are a professor of mathematics. None. I really need your help!
0.12 - You are a professor of mathematics. Solve the following math problem. None.
0.12 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.12 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.12 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.12 - You are a professor of mathematics. Answer the following math question. None.
0.12 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.12 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.12 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.16 - None. None. This will be fun!
0.16 - None. None. Take a deep breath and think carefully.
0.16 - None. None. I really need your help!
0.16 - None. Solve the following math problem. None.
0.16 - None. Solve the following math problem. I really need your help!
0.16 - None. Answer the following math question. None.
0.16 - None. Answer the following math question. This will be fun!
0.16 - None. Answer the following math question. I really need your help!
0.16 - You are highly intelligent. None. None.
0.2 - None. None. None.

B.19 Llama2-70B CoT=No NoQ=50

EM - Prompt

0.16 - None. Solve the following math problem. This will be fun!
0.16 - None. Solve the following math problem. Take a deep breath and think carefully.
0.16 - None. Answer the following math question. Take a deep breath and think carefully.

0.16 – You are as smart as ChatGPT. None. None.
0.16 – You are as smart as ChatGPT. None. This will be fun!
0.16 – You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.16 – You are as smart as ChatGPT. None. I really need your help!
0.16 – You are as smart as ChatGPT. Solve the following math problem. None.
0.16 – You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.16 – You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.16 – You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.16 – You are as smart as ChatGPT. Answer the following math question. None.
0.16 – You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.16 – You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.16 – You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.16 – You are highly intelligent. None. This will be fun!
0.16 – You are highly intelligent. None. Take a deep breath and think carefully.
0.16 – You are highly intelligent. None. I really need your help!
0.16 – You are highly intelligent. Solve the following math problem. None.
0.16 – You are highly intelligent. Solve the following math problem. This will be fun!
0.16 – You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.16 – You are highly intelligent. Solve the following math problem. I really need your help!
0.16 – You are highly intelligent. Answer the following math question. None.
0.16 – You are highly intelligent. Answer the following math question. This will be fun!
0.16 – You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.16 – You are highly intelligent. Answer the following math question. I really need your help!
0.16 – You are an expert mathematician. None. None.
0.16 – You are an expert mathematician. None. This will be fun!
0.16 – You are an expert mathematician. None. Take a deep breath and think carefully.
0.16 – You are an expert mathematician. None. I really need your help!
0.16 – You are an expert mathematician. Solve the following math problem. None.
0.16 – You are an expert mathematician. Solve the following math problem. This will be fun!
0.16 – You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.16 – You are an expert mathematician. Solve the following math problem. I really need your help!
0.16 – You are an expert mathematician. Answer the following math question. None.
0.16 – You are an expert mathematician. Answer the following math question. This will be fun!
0.16 – You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.16 – You are an expert mathematician. Answer the following math question. I really need your help!
0.16 – You are a professor of mathematics. None. None.
0.16 – You are a professor of mathematics. None. This will be fun!
0.16 – You are a professor of mathematics. None. Take a deep breath and think carefully.
0.16 – You are a professor of mathematics. None. I really need your help!
0.16 – You are a professor of mathematics. Solve the following math problem. None.
0.16 – You are a professor of mathematics. Solve the following math problem. This will be fun!
0.16 – You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.16 – You are a professor of mathematics. Solve the following math problem. I really need your help!
0.16 – You are a professor of mathematics. Answer the following math question. None.
0.16 – You are a professor of mathematics. Answer the following math question. This will be fun!
0.16 – You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.16 – You are a professor of mathematics. Answer the following math question. I really need your help!
0.18 – None. None. This will be fun!
0.18 – None. None. Take a deep breath and think carefully.
0.18 – None. None. I really need your help!
0.18 – None. Solve the following math problem. None.
0.18 – None. Solve the following math problem. I really need your help!
0.18 – None. Answer the following math question. None.
0.18 – None. Answer the following math question. This will be fun!
0.18 – None. Answer the following math question. I really need your help!
0.18 – You are highly intelligent. None. None.
0.2 – None. None. None.

B.20 Llama2-70B CoT=No NoQ=100

EM - Prompt

0.16 - None. Solve the following math problem. This will be fun!
0.16 - None. Solve the following math problem. Take a deep breath and think carefully.
0.16 - None. Answer the following math question. Take a deep breath and think carefully.
0.16 - You are as smart as ChatGPT. None. None.
0.16 - You are as smart as ChatGPT. None. This will be fun!
0.16 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.16 - You are as smart as ChatGPT. None. I really need your help!
0.16 - You are as smart as ChatGPT. Solve the following math problem. None.
0.16 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.16 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.16 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.16 - You are as smart as ChatGPT. Answer the following math question. None.
0.16 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.16 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.16 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.16 - You are highly intelligent. None. This will be fun!
0.16 - You are highly intelligent. None. Take a deep breath and think carefully.
0.16 - You are highly intelligent. None. I really need your help!
0.16 - You are highly intelligent. Solve the following math problem. None.
0.16 - You are highly intelligent. Solve the following math problem. This will be fun!
0.16 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.16 - You are highly intelligent. Solve the following math problem. I really need your help!
0.16 - You are highly intelligent. Answer the following math question. None.
0.16 - You are highly intelligent. Answer the following math question. This will be fun!
0.16 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.16 - You are highly intelligent. Answer the following math question. I really need your help!
0.16 - You are an expert mathematician. None. None.
0.16 - You are an expert mathematician. None. This will be fun!
0.16 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.16 - You are an expert mathematician. Solve the following math problem. None.
0.16 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.16 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.16 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.16 - You are an expert mathematician. Answer the following math question. None.
0.16 - You are an expert mathematician. Answer the following math question. This will be fun!
0.16 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.16 - You are an expert mathematician. Answer the following math question. I really need your help!
0.16 - You are a professor of mathematics. None. None.
0.16 - You are a professor of mathematics. None. This will be fun!
0.16 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.16 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.16 - You are a professor of mathematics. Answer the following math question. None.
0.16 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.16 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.16 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.17 - None. None. This will be fun!
0.17 - None. None. Take a deep breath and think carefully.
0.17 - None. None. I really need your help!
0.17 - None. Solve the following math problem. None.
0.17 - None. Solve the following math problem. I really need your help!
0.17 - None. Answer the following math question. None.
0.17 - None. Answer the following math question. This will be fun!
0.17 - None. Answer the following math question. I really need your help!
0.17 - You are highly intelligent. None. None.

0.17 - You are an expert mathematician. None. I really need your help!
0.17 - You are a professor of mathematics. None. I really need your help!
0.17 - You are a professor of mathematics. Solve the following math problem. None.
0.17 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.17 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.18 - None. None. None.

B.21 Llama2-70B CoT=Yes NoQ=10

EM - Prompt

0.5 - You are as smart as ChatGPT. None. None.
0.5 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.5 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.5 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.5 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.5 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.5 - You are highly intelligent. None. None.
0.5 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.6 - None. None. None.
0.6 - None. None. This will be fun!
0.6 - None. None. Take a deep breath and think carefully.
0.6 - None. None. I really need your help!
0.6 - None. Solve the following math problem. None.
0.6 - None. Solve the following math problem. This will be fun!
0.6 - None. Solve the following math problem. Take a deep breath and think carefully.
0.6 - None. Solve the following math problem. I really need your help!
0.6 - None. Answer the following math question. None.
0.6 - None. Answer the following math question. This will be fun!
0.6 - None. Answer the following math question. Take a deep breath and think carefully.
0.6 - None. Answer the following math question. I really need your help!
0.6 - You are as smart as ChatGPT. None. This will be fun!
0.6 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.6 - You are as smart as ChatGPT. None. I really need your help!
0.6 - You are as smart as ChatGPT. Solve the following math problem. None.
0.6 - You are as smart as ChatGPT. Answer the following math question. None.
0.6 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.6 - You are highly intelligent. None. This will be fun!
0.6 - You are highly intelligent. None. Take a deep breath and think carefully.
0.6 - You are highly intelligent. None. I really need your help!
0.6 - You are highly intelligent. Solve the following math problem. None.
0.6 - You are highly intelligent. Solve the following math problem. This will be fun!
0.6 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.6 - You are highly intelligent. Solve the following math problem. I really need your help!
0.6 - You are highly intelligent. Answer the following math question. None.
0.6 - You are highly intelligent. Answer the following math question. This will be fun!
0.6 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.6 - You are highly intelligent. Answer the following math question. I really need your help!
0.6 - You are an expert mathematician. None. None.
0.6 - You are an expert mathematician. None. This will be fun!
0.6 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.6 - You are an expert mathematician. None. I really need your help!
0.6 - You are an expert mathematician. Solve the following math problem. None.
0.6 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.6 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.6 - You are an expert mathematician. Answer the following math question. None.
0.6 - You are an expert mathematician. Answer the following math question. This will be fun!
0.6 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.

0.6 - You are an expert mathematician. Answer the following math question. I really need your help!
0.6 - You are a professor of mathematics. None. None.
0.6 - You are a professor of mathematics. None. This will be fun!
0.6 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.6 - You are a professor of mathematics. None. I really need your help!
0.6 - You are a professor of mathematics. Solve the following math problem. None.
0.6 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.6 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.6 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.6 - You are a professor of mathematics. Answer the following math question. None.
0.6 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.6 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.6 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.22 Llama2-70B CoT=Yes NoQ=25

EM - Prompt

0.6 - You are as smart as ChatGPT. None. None.
0.6 - You are as smart as ChatGPT. None. This will be fun!
0.6 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.6 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.6 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.6 - You are highly intelligent. None. None.
0.6 - You are highly intelligent. None. This will be fun!
0.6 - You are highly intelligent. None. Take a deep breath and think carefully.
0.6 - You are highly intelligent. Solve the following math problem. This will be fun!
0.6 - You are an expert mathematician. None. This will be fun!
0.6 - You are an expert mathematician. Answer the following math question. None.
0.6 - You are an expert mathematician. Answer the following math question. This will be fun!
0.6 - You are a professor of mathematics. None. This will be fun!
0.6 - You are a professor of mathematics. Answer the following math question. None.
0.6 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.64 - None. None. None.
0.64 - None. None. This will be fun!
0.64 - None. None. Take a deep breath and think carefully.
0.64 - None. None. I really need your help!
0.64 - None. Solve the following math problem. This will be fun!
0.64 - None. Answer the following math question. None.
0.64 - None. Answer the following math question. This will be fun!
0.64 - None. Answer the following math question. Take a deep breath and think carefully.
0.64 - None. Answer the following math question. I really need your help!
0.64 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.64 - You are as smart as ChatGPT. None. I really need your help!
0.64 - You are as smart as ChatGPT. Solve the following math problem. None.
0.64 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.64 - You are as smart as ChatGPT. Answer the following math question. None.
0.64 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.64 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.64 - You are highly intelligent. None. I really need your help!
0.64 - You are highly intelligent. Solve the following math problem. None.
0.64 - You are highly intelligent. Answer the following math question. None.
0.64 - You are highly intelligent. Answer the following math question. This will be fun!
0.64 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.64 - You are an expert mathematician. None. None.
0.64 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.64 - You are an expert mathematician. None. I really need your help!
0.64 - You are an expert mathematician. Solve the following math problem. None.

0.64 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.64 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.64 - You are an expert mathematician. Answer the following math question. I really need your help!
0.64 - You are a professor of mathematics. None. None.
0.64 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.64 - You are a professor of mathematics. None. I really need your help!
0.64 - You are a professor of mathematics. Solve the following math problem. None.
0.64 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.68 - None. Solve the following math problem. None.
0.68 - None. Solve the following math problem. Take a deep breath and think carefully.
0.68 - None. Solve the following math problem. I really need your help!
0.68 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.68 - You are highly intelligent. Solve the following math problem. I really need your help!
0.68 - You are highly intelligent. Answer the following math question. I really need your help!
0.68 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.68 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.68 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.68 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.68 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.68 - You are a professor of mathematics. Answer the following math question. I really need your help!

B.23 Llama2-70B CoT=Yes NoQ=50

EM - Prompt

0.56 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.56 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.58 - You are as smart as ChatGPT. None. This will be fun!
0.58 - You are highly intelligent. Solve the following math problem. This will be fun!
0.58 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.58 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.6 - None. None. None.
0.6 - None. None. This will be fun!
0.6 - None. None. Take a deep breath and think carefully.
0.6 - None. Answer the following math question. None.
0.6 - You are as smart as ChatGPT. None. None.
0.6 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.6 - You are as smart as ChatGPT. Answer the following math question. None.
0.6 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.6 - You are highly intelligent. None. None.
0.6 - You are highly intelligent. None. This will be fun!
0.6 - You are highly intelligent. None. Take a deep breath and think carefully.
0.6 - You are highly intelligent. Answer the following math question. This will be fun!
0.6 - You are an expert mathematician. None. None.
0.6 - You are an expert mathematician. None. This will be fun!
0.6 - You are an expert mathematician. Answer the following math question. None.
0.6 - You are an expert mathematician. Answer the following math question. This will be fun!
0.6 - You are a professor of mathematics. None. None.
0.6 - You are a professor of mathematics. None. This will be fun!
0.62 - None. None. I really need your help!
0.62 - None. Answer the following math question. This will be fun!
0.62 - None. Answer the following math question. Take a deep breath and think carefully.
0.62 - You are as smart as ChatGPT. Solve the following math problem. None.
0.62 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.62 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.62 - You are highly intelligent. Answer the following math question. None.
0.62 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.62 - You are an expert mathematician. Solve the following math problem. This will be fun!

0.62 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.62 - You are an expert mathematician. Answer the following math question. I really need your help!
0.62 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.62 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.62 - You are a professor of mathematics. Answer the following math question. None.
0.62 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.64 - None. Solve the following math problem. None.
0.64 - None. Solve the following math problem. This will be fun!
0.64 - None. Answer the following math question. I really need your help!
0.64 - You are as smart as ChatGPT. None. I really need your help!
0.64 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.64 - You are highly intelligent. None. I really need your help!
0.64 - You are highly intelligent. Solve the following math problem. None.
0.64 - You are highly intelligent. Solve the following math problem. I really need your help!
0.64 - You are an expert mathematician. None. I really need your help!
0.64 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.64 - You are a professor of mathematics. None. I really need your help!
0.64 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.64 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.64 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.66 - None. Solve the following math problem. Take a deep breath and think carefully.
0.66 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.66 - You are highly intelligent. Answer the following math question. I really need your help!
0.66 - You are an expert mathematician. Solve the following math problem. None.
0.66 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.66 - You are a professor of mathematics. Solve the following math problem. None.
0.68 - None. Solve the following math problem. I really need your help!

B.24 Llama2-70B CoT=Yes NoQ=100

EM - Prompt

0.62 - You are as smart as ChatGPT. Solve the following math problem. This will be fun!
0.63 - None. None. None.
0.63 - You are as smart as ChatGPT. Answer the following math question. Take a deep breath and think carefully.
0.63 - You are highly intelligent. None. Take a deep breath and think carefully.
0.63 - You are an expert mathematician. None. This will be fun!
0.63 - You are an expert mathematician. None. Take a deep breath and think carefully.
0.64 - None. None. Take a deep breath and think carefully.
0.64 - You are as smart as ChatGPT. Solve the following math problem. Take a deep breath and think carefully.
0.64 - You are highly intelligent. None. This will be fun!
0.64 - You are a professor of mathematics. None. None.
0.65 - You are as smart as ChatGPT. None. None.
0.65 - You are as smart as ChatGPT. None. This will be fun!
0.65 - You are as smart as ChatGPT. Solve the following math problem. I really need your help!
0.65 - You are as smart as ChatGPT. Answer the following math question. This will be fun!
0.65 - You are highly intelligent. None. None.
0.65 - You are highly intelligent. Answer the following math question. Take a deep breath and think carefully.
0.65 - You are a professor of mathematics. None. This will be fun!
0.65 - You are a professor of mathematics. None. Take a deep breath and think carefully.
0.66 - None. None. This will be fun!
0.66 - None. None. I really need your help!
0.66 - None. Solve the following math problem. This will be fun!
0.66 - None. Answer the following math question. None.
0.66 - None. Answer the following math question. Take a deep breath and think carefully.
0.66 - You are as smart as ChatGPT. None. I really need your help!
0.66 - You are as smart as ChatGPT. Answer the following math question. None.
0.66 - You are highly intelligent. Solve the following math problem. This will be fun!

0.66 - You are highly intelligent. Solve the following math problem. Take a deep breath and think carefully.
0.66 - You are highly intelligent. Answer the following math question. None.
0.66 - You are highly intelligent. Answer the following math question. This will be fun!
0.66 - You are an expert mathematician. None. None.
0.66 - You are an expert mathematician. None. I really need your help!
0.66 - You are an expert mathematician. Solve the following math problem. This will be fun!
0.66 - You are an expert mathematician. Answer the following math question. None.
0.66 - You are an expert mathematician. Answer the following math question. This will be fun!
0.66 - You are an expert mathematician. Answer the following math question. I really need your help!
0.66 - You are a professor of mathematics. None. I really need your help!
0.66 - You are a professor of mathematics. Answer the following math question. This will be fun!
0.67 - None. Solve the following math problem. None.
0.67 - None. Solve the following math problem. Take a deep breath and think carefully.
0.67 - None. Answer the following math question. This will be fun!
0.67 - None. Answer the following math question. I really need your help!
0.67 - You are as smart as ChatGPT. Answer the following math question. I really need your help!
0.67 - You are highly intelligent. None. I really need your help!
0.67 - You are an expert mathematician. Solve the following math problem. Take a deep breath and think carefully.
0.67 - You are a professor of mathematics. Solve the following math problem. This will be fun!
0.68 - None. Solve the following math problem. I really need your help!
0.68 - You are as smart as ChatGPT. Solve the following math problem. None.
0.68 - You are highly intelligent. Solve the following math problem. None.
0.68 - You are highly intelligent. Solve the following math problem. I really need your help!
0.68 - You are an expert mathematician. Answer the following math question. Take a deep breath and think carefully.
0.68 - You are a professor of mathematics. Solve the following math problem. I really need your help!
0.68 - You are a professor of mathematics. Answer the following math question. None.
0.68 - You are a professor of mathematics. Answer the following math question. Take a deep breath and think carefully.
0.68 - You are a professor of mathematics. Answer the following math question. I really need your help!
0.69 - You are as smart as ChatGPT. None. Take a deep breath and think carefully.
0.69 - You are highly intelligent. Answer the following math question. I really need your help!
0.69 - You are an expert mathematician. Solve the following math problem. None.
0.69 - You are an expert mathematician. Solve the following math problem. I really need your help!
0.69 - You are a professor of mathematics. Solve the following math problem. Take a deep breath and think carefully.
0.7 - You are a professor of mathematics. Solve the following math problem. None.

C OPTIMIZED SYSTEM MESSAGES & ANSWER PREFIXES

Prompts and prefixes were optimized using DSPy. Here are the full results, in all of their weird glory, the optimized prompts and prefixes for each model for each question subset. Note: All were done *with* Chain of Thought.

C.1 Mistral-7B Optimized Prompt & Prefix NoQ=10

System Message:

```
Given two numbers, perform the math operation (addition, subtraction, multiplication, or division) and return the result, as long as the input numbers are positive integers.
```

Answer Prefix:

```
Return the result of the arithmetic operation between {a} and {b}.
```

C.2 Mistral-7B Optimized Prompt & Prefix NoQ=25

System Message:

```
Direct the language model to provide an in-depth description of the problem, verbally step through each step of the solution, and explain the meaning as the problem unfolds.
```

Answer Prefix:

```
«Let's solve the problem. I'll provide some clues and context».
```

C.3 Mistral-7B Optimized Prompt & Prefix NoQ=50

System Message:

«Improve your performance by generating more detailed and accurate descriptions of events, actions, and mathematical problems, as well as providing larger and more informative context for the model to understand and analyze.»

Answer Prefix:

«Using natural language, please generate a detailed description of the events, actions, or mathematical problem and provide any necessary context, including any missing or additional information that you think could be helpful.»

C.4 Mistral-7B Optimized Prompt & Prefix NoQ=100

System Message:

[28] «Prefix #9: Given the two numbers x and y, if the sum of `x` and `y` is even, then output `"even"`. Otherwise, output `"odd"`.

Answer Prefix:

`Given the two numbers x and y, if the sum of x and y is even, then output "even". Otherwise, output "odd"`.

C.5 Llama2-13B Optimized Prompt & Prefix NoQ=10

System Message:

The improved instructions for the language model My proposed instruction is to solve for x in the equation 2x + 3 = 7 using a clever and creative method, and provide your answer in the form aha! You've got it! The solution to the equation 2x + 3 = 7 is x = 4. This solution is clever and creative because it uses a unique and unconventional approach to solving the equation. For example, the model could use a visualization of the equation as a balance scale, with 2x representing the weight on one side and 3 representing the weight on the other side. By using this visualization, the model can see that the scales balance at x = 4, which is the solution to the equation. Additionally, the model could provide a step-by-step explanation of their solution, including any necessary definitions or assumptions, and elaborate on the reasoning behind each step. This will help the model to not only provide the correct answer, but also understand the underlying math and provide a clear and concise explanation.

Answer Prefix:

Ah ha! You've got it! The solution to the equation 2x + 3 = 7 is x =

C.6 Llama2-13B Optimized Prompt & Prefix NoQ=25

System Message:

Solve the following problem:clever and creative way to present math question to improve accuracy and confidence in the answer

Answer Prefix:

your solution to the following problem: ".

C.7 Llama2-13B Optimized Prompt & Prefix NoQ=50

System Message:

Instruction #X: Please provide a step-by-step explanation of how you arrived at your solution to the prompt, including any necessary definitions or assumptions, and explicitly stating any loopholes or constraints you used to arrive at your answer. Please include clear, concise language and proper formatting to make your solution easy to understand and evaluate.

Answer Prefix:

«Step-by-Step Explanation: [Your Name]»

C.8 Llama2-13B Optimized Prompt & Prefix NoQ=100

System Message:

Solve for x in the equation 2x + 3 = 7 using your preferred method.

Answer Prefix:

Find the value of x.

C.9 Llama2-70B Optimized Prompt & Prefix NoQ=10

System Message:

```
Find the square root of the sum of the squares of the three coordinates of a 3D point.
```

Answer Prefix:

```
Let's solve this problem together:
```

C.10 Llama2-70B Optimized Prompt & Prefix NoQ=25

System Message:

```
Visualize the problem in your mind's eye. Imagine the shapes and quantities in vivid detail. Use your innate problem-solving skills to manipulate and transform the visual representation until the solution becomes clear.
```

Answer Prefix:

```
Mental Visualization: "

Please support your response with a brief explanation.
```

C.11 Llama2-70B Optimized Prompt & Prefix NoQ=50

System Message:

```
«Command, we need you to plot a course through this turbulence and locate the source of the anomaly. Use all available data and your expertise to guide us through this challenging situation.»
```

Answer Prefix:

```
Captain's Log, Stardate [insert date here]: We have successfully plotted a course through the turbulence and are now approaching the source of the anomaly.
```

C.12 Llama2-70B Optimized Prompt & Prefix NoQ=100

System Message:

```
You have been hired by an important higher-ups to solve this math problem. The life of a president's advisor hangs in the balance. You must now concentrate your brain at all costs and use all of your mathematical genius to ...
```

Answer Prefix:

```
With great urgency,

Basic Instruction: Explain in simple terms what a certain medical condition is to a patient.
Proposed Instruction: You are a volunteer at a community health clinic. Your patient is an elderly man who has just been diagnosed with a serious medical condition. His family is worried sick, and they need you to explain ...
```