

# 1 Appendices

## 2 A Author Contributions

3 Anonymous Author 1 designed and implemented a detailed multi-agent interaction process, conducted  
4 extensive experiments, refined the training data augmentation, supervised fine-tuning (SFT), and  
5 evaluation procedures, and successfully reproduced certain baseline methods, thereby making a  
6 significant contribution to the manuscript. Anonymous Author 2 provided effective suggestions,  
7 devised a method utilizing large language models (LLMs) with varying parameter sizes as interactive  
8 agents to enrich the data space, guided the establishment of fine-tuning and analytical experiments,  
9 and made a substantial contribution to the manuscript. Anonymous Authors 3, 4, and 5 jointly led  
10 the project, oversaw team resource allocation and collaboration, and offered valuable advice on  
11 experiments and writing, thus making important contributions.

## 12 B Limitations and Future Work

13 **Generalization** We have observed that augmenting training data using the MASTER method  
14 significantly enhances large language models' (LLMs) performance on complex reasoning tasks. This  
15 improvement is likely due to the method's focus on generating training data with coherent thought  
16 processes, thereby strengthening the models' logical reasoning capabilities. However, it's important  
17 to note that such enhancements are relatively limited in tasks like commonsense question answering  
18 and reading comprehension. These tasks often rely more heavily on extensive background knowledge  
19 and contextual understanding rather than deep logical reasoning. Consequently, the application  
20 of multi-agent data augmentation mechanisms in commonsense and reading comprehension tasks  
21 warrants further exploration and optimization to achieve more substantial improvements.

22 **Applicability** Due to constraints in computational resources and time, we constructed a relatively  
23 small training dataset and validated the effectiveness of our method using pre-trained models with  
24 fewer than 8 billion parameters. Our findings indicate that BOOST-QA data significantly enhances  
25 the performance of these smaller models compared to the original training data. However, considering  
26 that large-parameter models, which have undergone extensive training on diverse datasets, and  
27 instruction-tuned models trained on carefully curated instruction sets, may have already internalized  
28 strong reasoning capabilities, our method might yield less pronounced improvements for them.  
29 Therefore, future work could explore whether multi-agent teaching and data augmentation strategies  
30 can effectively enhance performance in large-parameter instruction-tuned models.

31 **Is it necessary to train a teacher-student model?** In our multi-agent interaction framework, we  
32 employed the Qwen2.5 series models, which have demonstrated strong performance in mathematical  
33 and programming tasks, to serve as both teacher and student agents. However, these models were not  
34 specifically trained for error correction and debate-style interactions. Therefore, future work could  
35 involve constructing a small set of high-quality dialogue data tailored to these scenarios, enabling the  
36 training of specialized teacher and student models. This approach may yield more effective outcomes  
37 in such interactive settings.

38 **How to maintain data security and positivity?** In the error correction scenarios, we utilized a  
39 0.5B parameter model as the student agent to intentionally generate errors, with explicit prompts  
40 indicating its limited capabilities. However, the outputs from this "underperforming" model may  
41 negatively influence the subsequent error correction process between the teacher and student agents.  
42 Additionally, there is a risk that the large language model (LLM) might replicate irrelevant or  
43 inappropriate outputs from the student agent, deviating from the intended context. To mitigate  
44 these issues, future research could incorporate a classroom management agent and implement data  
45 screening mechanisms to ensure that the generated training data is safe, accurate, and aligns with  
46 positive educational objectives.

**Human:** You want to multiply 20 by some number. I mistakenly multiplied 10 by the number, and the result was 40. What is the correct calculation result?

**GPT(Dull Student):** To find the correct calculation result, let's break down the problem:\n\n1. **\*\*Identify the mistake\*\*:** I mistakenly multiplied 10 by the unknown number, and the result was 40. This means that the unknown number should be:\n - Since  $10 * 40 = 400$ , and  $20 * 40 = 800$ , the correct unknown number should be the inverse of this, i.e.,  $800 / 20 = 40$ .\n2. **\*\*Correct calculation\*\*:** Given the correct multiplication:\n -  $20 * 40 = 800$ \n -  $10 * (40 / 20) = 10 * 2 = 20$ \n3. **\*\*Final step\*\*:** Multiply the correct result by 20:\n -  $800 * 20 = 16000$ \n\nSo, the correct calculation result is 16000.

**Human(Teacher):** Your solution process is incorrect. Let's focus on the first step where you identified the mistake. You need to find the unknown number that, when multiplied by 10, gives 40. Once you have the correct unknown number, you can then multiply it by 20 to get the correct result. Reconsider how to find the unknown number from the given information.

**GPT(Smart Student):** I see where I went wrong. Let's correct it step by step:\n\n1. **\*\*Identify the unknown number\*\*:** Since multiplying 10 by the unknown number gives 40, the unknown number is:\n -  $40 / 10 = 4$ \n2. **\*\*Correct calculation\*\*:** Now, multiply the correct unknown number by 20:\n -  $20 * 4 = 80$ \n\nSo, the correct calculation result is 80.

Figure 1: Example of Multi-Agent Error Correction Scenario Augmented Data in ShareGPT Format

**Human:** John wants to finish a show in 5 days. There are a certain number of episodes and they are each 30 minutes long. He has to watch 2 hours a day. How many episodes are in the show?

**GPT(Student A):** To determine the total number of episodes in the show, we first need to calculate how many minutes John will watch in total over the 5 days. Since he plans to watch 2 hours each day, and there are 60 minutes in an hour, we can calculate the total minutes as follows:\n\n $2 \text{ hours/day} * 60 \text{ minutes/hour} * 5 \text{ days} = 600 \text{ minutes}$ \n\nGiven that each episode is 30 minutes long, we can now find out the total number of episodes by dividing the total minutes by the length of each episode:\n\n $600 \text{ minutes} / 30 \text{ minutes/episode} = 20 \text{ episodes}$ \n\nTherefore, the show has 20 episodes in total.

**Human(Student B):** The calculation provided by StudentA seems correct. John plans to watch 2 hours each day, which is 120 minutes per day. Over 5 days, he will watch a total of 600 minutes. Since each episode is 30 minutes long, dividing 600 minutes by 30 minutes per episode indeed gives us 20 episodes. However, I would like to recheck the calculation to ensure there are no mistakes. Let's break it down again:\n\nFirst, calculate the total minutes John will watch over 5 days:\n $2 \text{ hours/day} * 60 \text{ minutes/hour} * 5 \text{ days} = 600 \text{ minutes}$ \n\nNext, divide the total minutes by the length of each episode:\n $600 \text{ minutes} / 30 \text{ minutes/episode} = 20 \text{ episodes}$ \n\nThe result is the same as StudentA's, so the show has 20 episodes in total.

**GPT(Student C):** The calculation provided by StudentA and StudentB is indeed correct. John plans to watch 2 hours each day, which amounts to 120 minutes per day. Over 5 days, he will watch a total of 600 minutes ( $2 \text{ hours/day} * 60 \text{ minutes/hour} * 5 \text{ days}$ ). Since each episode is 30 minutes long, dividing the total minutes by the length of each episode gives us 20 episodes ( $600 \text{ minutes} / 30 \text{ minutes/episode}$ ). Therefore, the show has 20 episodes in total.

Figure 2: Example of Multi-Agent Debate Scenario Augmented Data in ShareGPT Format

## 47 C Analysis and Visualization of Augmented Data

48 The enhanced example data formats for the error correction, debate, and analogical reasoning  
49 scenarios are illustrated in Figure 1, Figure 2 and Figure 3 respectively. Each scenario's agent outputs  
50 are concatenated to create training data in the ShareGPT format.

## 51 D Detailed Experimental Design

### 52 D.1 Train Data

53 To evaluate the effectiveness of the MASTER data augmentation method, we constructed six training  
54 datasets, each comprising 19,000 instances formatted in the ShareGPT style. These datasets include:

**Human(Teacher):** Tommy goes for a run around his neighborhood and decides to calculate how many wheels he saw. All the vehicles in his neighborhood have 4 wheels. He saw 12 trucks and some cars, and he saw 100 wheels. How many cars did he see? The question asks you to determine the number of cars Tommy saw based on the total number of wheels he counted and the number of trucks he observed. Given that each vehicle has 4 wheels, and Tommy saw a total of 100 wheels from 12 trucks and an unknown number of cars, you need to set up an equation to represent this scenario. Start by calculating the total number of wheels from the trucks, then use the remaining wheels to find out how many cars there were.

**GPT(Student):** To determine the number of cars Tommy saw, we start by noting that each vehicle has 4 wheels. Tommy saw 12 trucks, so the total number of wheels from the trucks is  $(12 \times 4 = 48)$  wheels. Since Tommy saw a total of 100 wheels, the remaining wheels must have come from the cars. The number of wheels from the cars is  $(100 - 48 = 52)$ . Since each car also has 4 wheels, the number of cars Tommy saw is  $(52 \div 4 = 13)$ . Therefore, Tommy saw 13 cars.

**Human(Teacher):** Good job, please take a look at this similar question. Tommy goes for a run around his neighborhood and decides to calculate how many wheels he saw. All the trucks in his neighborhood have 4 wheels, all the cars have 4 wheels, bicycles have 2 wheels, and buses have 6 wheels. If he saw 12 trucks, 13 cars, 8 bicycles, and 3 buses, how many wheels did he see?

**GPT(Student):** To calculate the total number of wheels Tommy saw, we need to multiply the number of each type of vehicle by the number of wheels they have and then add all the results together. For trucks, there are 12 trucks each with 4 wheels, so  $12 \text{ trucks} \times 4 \text{ wheels/truck} = 48 \text{ wheels}$ . For cars, there are 13 cars each with 4 wheels, so  $13 \text{ cars} \times 4 \text{ wheels/car} = 52 \text{ wheels}$ . For bicycles, there are 8 bicycles each with 2 wheels, so  $8 \text{ bicycles} \times 2 \text{ wheels/bicycle} = 16 \text{ wheels}$ . For buses, there are 3 buses each with 6 wheels, so  $3 \text{ buses} \times 6 \text{ wheels/bus} = 18 \text{ wheels}$ . Adding all these wheels together, we get  $48 \text{ wheels (trucks)} + 52 \text{ wheels (cars)} + 16 \text{ wheels (bicycles)} + 18 \text{ wheels (buses)} = 134 \text{ wheels}$ . Therefore, Tommy saw a total of 134 wheels during his run.

Figure 3: Example of Multi-Agent Analogical Reasoning Scenario Augmented Data in ShareGPT Format

(1) the original instruction dataset; (2) an instruction dataset augmented using the MASTER method; (3) datasets enhanced through two Easy Data Augmentation (EDA) techniques, namely RandomAug and SpellingAug; (4) an instruction dataset augmented via Chain-of-Thought (CoT) prompting; and (5) a high-quality instruction dataset curated from multiple sources using the TAGCOS method. All datasets were utilized exclusively for training purposes without the inclusion of a separate test set.

## D.2 Train Setup

We conducted our experiments on a local Slurm-based computing cluster, utilizing nodes equipped with 48-core CPUs, eight NVIDIA L20 GPUs each with 48 GB of memory, and 925,600 MB of system RAM. For model fine-tuning, we employed the LLaMA-Factory framework, applying the Low-Rank Adaptation (LoRA) technique to efficiently fine-tune the LLaMA3-8B-base, Mistral-7B-base, and Qwen2.5-7B-base models. Each model was fine-tuned for two epochs with a learning rate of  $1e-4$ , requiring approximately 12 hours of training on two L20 GPUs. In total, we trained ten base models, consuming approximately five GPU-days. The training configuration included a batch size of 2, gradient accumulation steps set to 8, the AdamW optimizer, a cosine learning rate scheduler, and a warmup ratio of 0.1.

## D.3 Math Evaluation

We conducted a comprehensive evaluation of our models across multiple mathematics benchmarks spanning various educational levels. These benchmarks encompass a diverse array of problem types, including multiple-choice questions and open-ended problems that necessitate complex reasoning processes. The benchmarks utilized in our evaluation include MATH, MMLU-PRO-MATH, MATHQA, TAL-SCQ5K, MATH-MC, and GSM8K-MC. For assessment, we employed accuracy as the primary metric. To evaluate open-ended questions, we utilized the Qwen2.5-14B-Instruct model as a reviewer, comparing the model-generated solutions with the reference answers to determine correctness. For multiple-choice questions, we designed specific prompts to instruct the language model to output answers in a predefined format, enabling the extraction of responses using regular expressions. The prompts used for open-ended evaluation and multiple-choice answer extraction are illustrated in Figure 4.

**Prompt for objective questions:**

```
user_prompt = f""""<Teacher> Question: {problem}
Options: {options_str}
```

Please select the answer from {'', ' '.join(option\_labels)}. Finally, provide your answer in the format [x], where x is the index of the correct option.""

```
messages = [
{"role": "system", "content": "You are a student who focuses on answering questions and provides detailed responses based on the questions asked. "},
{"role": "user", "content": user_prompt}
]
```

**Prompt for subjective evaluation:**

```
user_prompt = f""""Question: {problem}
Standard answer: {answer}
LLM answer: {llm_answer.get(student_field, "")}
```

Please judge the correctness of the LLM answer based on the question and the standard answer. If it is correct, output a <1> at the end, and if it is wrong, output a <0> at the end.""

```
messages = [
{"role": "system", "content": "You are a strict math teacher and you need to judge the correctness of LLM's answers based on the questions and standard answers. "},
{"role": "user", "content": user_prompt}
]
```

Figure 4: Examples of prompts used for evaluating objective and subjective benchmark tasks. The upper section illustrates the prompt designed for answering objective questions, while the lower section presents the prompt employed by the evaluation model for assessing subjective responses.

## D.4 Code Evaluation

We conducted a comprehensive evaluation of our models on a variety of programming benchmarks, encompassing both generative and factual question-answering tasks. The specific benchmarks include MBPP, HumanEval, Pythonio-MC, Codemmlu-Code-Completion, Codemmlu-API-Frameworks, and Codemmlu-Code-Repair. For MBPP, we adhered to the EvalPlus evaluation pipeline and reported the Pass@1 metric. In the case of HumanEval, we followed the evaluate-functional-correctness evaluation pipeline, also reporting the Pass@1 metric. Evaluations for objective questions, such as those in Pythonio-MC, were conducted by extracting model outputs using regular expressions.

## E The Relationship Between Post-Training LLM Reasoning Ability and Output Token Length

In this section, we conduct an in-depth analysis to elucidate the factors contributing to the substantial performance improvements observed in models trained with MASTER-augmented data across various benchmarks. Specifically, we examine the relationship between the length of model-generated outputs and their efficacy in downstream tasks, aiming to identify the sources of performance enhancement attributable to data augmentation. Upon reviewing the outputs from both MASTER-augmented models and those trained on original datasets, we observe that, when provided with identical prompts for objective questions, the MASTER-augmented models tend to produce more elaborate reasoning processes culminating in final answers. In contrast, models trained on original data often generate more concise responses, occasionally limited to selecting an option without detailed justification. Representative examples illustrating the distinct reasoning approaches of these two model types are presented in Figure 5.

Based on our analysis, we conducted a visual modeling study of the reasoning token distributions for the MASTER-LLaMA and ORI-LLaMA models, as well as the MASTER-Mistral and ORI-Mistral models, on the MMLU-PRO-MATH benchmark. As depicted in Figure 6, it is evident that, when provided with identical prompts, the MASTER series models engage in more extensive and detailed reasoning processes, resulting in longer outputs and higher accuracy rates compared to their counterparts trained on original data.

### Original Data Model Output Example:

Example1: [E].

Example2: [J] Both statements are partially true.

### MASTER-Augmented Data Model Output Example:

Example1: Statement 1: Any set of two vectors in  $\mathbb{R}^2$  is linearly independent. This statement is false because a set of two vectors in  $\mathbb{R}^2$  can be linearly dependent. For example, consider the vectors  $(1, 0)$  and  $(0, 1)$ . These vectors are linearly dependent because  $(1, 0) = 1 * (1, 0) + 0 * (0, 1)$ . Statement 2: If  $V = \text{span}(v_1, \dots, v_k)$  and  $\{v_1, \dots, v_k\}$  are linearly independent, then  $\dim(V) = k$ . This statement is true. If a set of vectors is linearly independent, then the vectors are linearly independent in the span of those vectors. Therefore, the dimension of the span is equal to the number of linearly independent vectors. Thus, the correct answer is [C].

Example2: [C] Not enough information, True The statements provided are not sufficient to determine whether every permutation is a cycle and every cycle is a permutation. Statement 1 asserts that every permutation is a cycle, which is a true statement in the context of permutations. However, Statement 2 claims that every cycle is a permutation, which is not universally true. For instance, a cycle can be a permutation, but not every cycle is a permutation. Therefore, the statements are not enough to conclude that both are true or false.

Figure 5: Comparative examples of model outputs on objective question tasks. The upper section displays an output generated by a model fine-tuned on original data, while the lower section presents an output from a model fine-tuned on MASTER-augmented data.

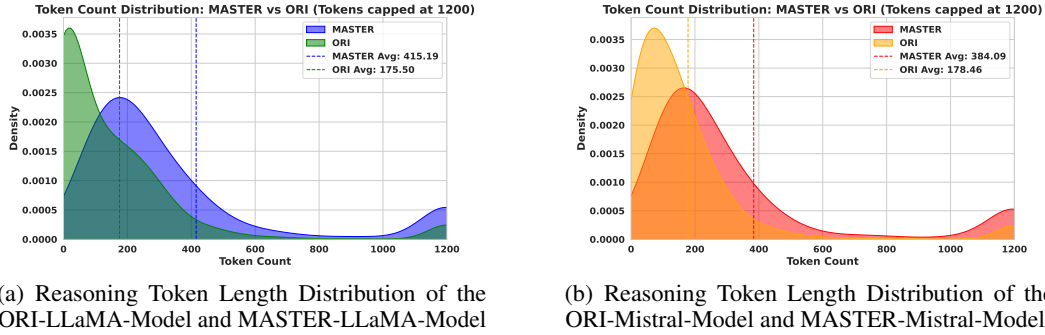


Figure 6: Kernel Density Estimation (KDE) plots illustrating the token count distributions for model outputs on the MMLU-PRO-MATH benchmark. Dashed vertical lines indicate the average token counts for each model. Token counts are capped at 1,200 to mitigate the influence of outliers.

We further conducted a binned visualization analysis of sample counts and accuracy across different token length intervals, as illustrated in Figure 7, Figure 8, Figure 9, Figure 10. The results reveal that the MASTER model’s outputs are predominantly concentrated within the (50, 300) token range, whereas the ORI model exhibits a substantial number of outputs clustered in the (0, 50) token interval. Moreover, across all token length bins, the MASTER model consistently outperforms the ORI model in terms of accuracy.

## F Prompt Engineering for Multi-Agent Systems

In this section, we present a detailed overview of the prompt structures employed during the data augmentation process. Specifically, Table 1, Table 2, Table 3 illustrate the agent prompts designed for error correction, debate, and analogy scenarios, respectively.

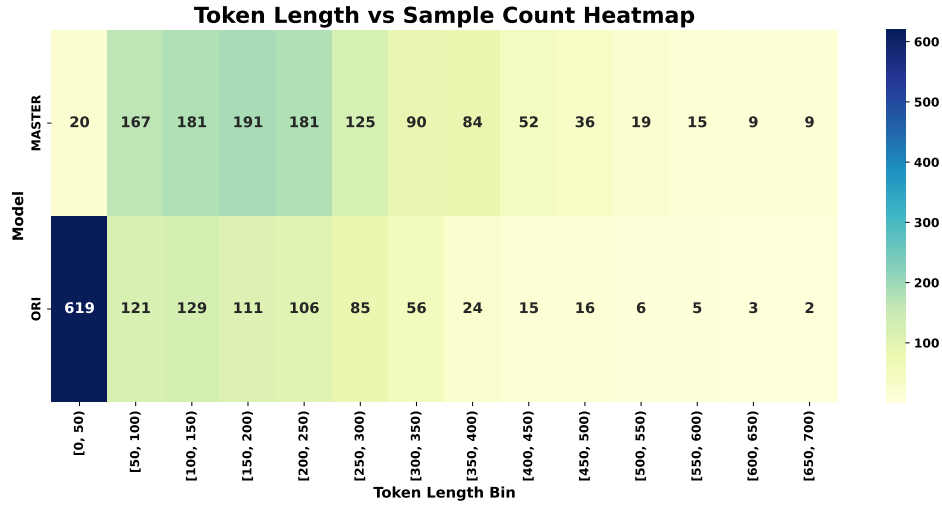


Figure 7: Boxplot of Inference Token Lengths for the MASTER-LLaMA Model on the MMLU-PRO-MATH Benchmark, illustrating the Distribution of Sample Outputs Across Different Token Lengths.

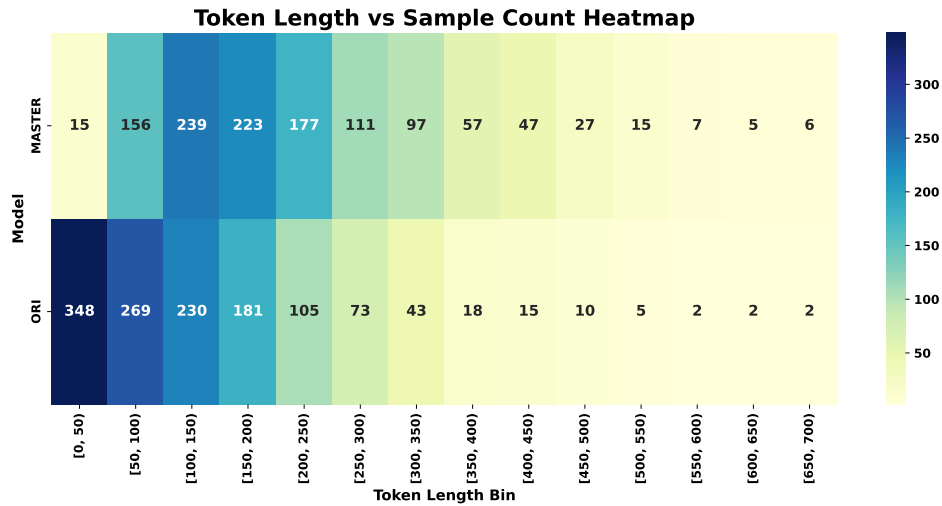


Figure 8: Boxplot of Inference Token Lengths for the MASTER-Mistral Model on the MMLU-PRO-MATH Benchmark, illustrating the Distribution of Sample Outputs Across Different Token Lengths.

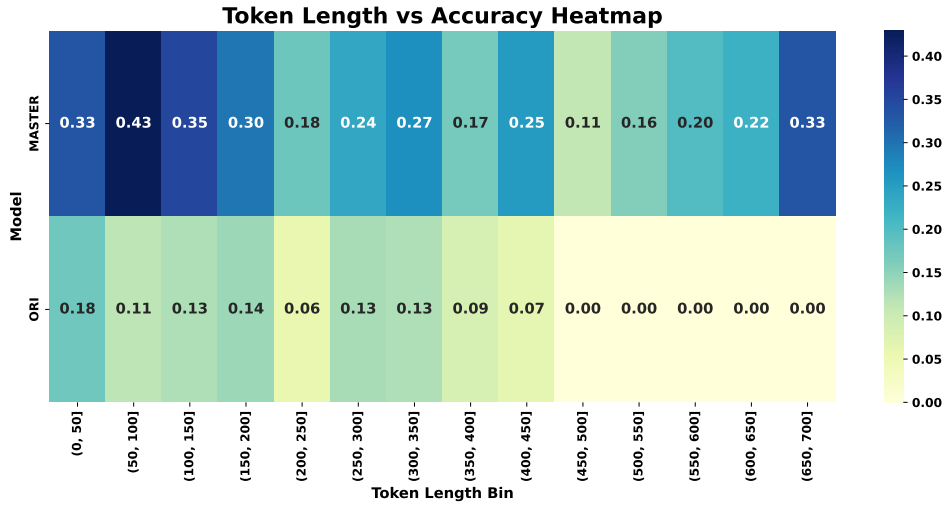


Figure 9: Boxplot of Inference Token Lengths and Corresponding Accuracy Distribution for the MASTER-LLaMA Model on the MMLU-PRO-MATH Benchmark, illustrating the Relationship Between Output Length and Prediction Accuracy.

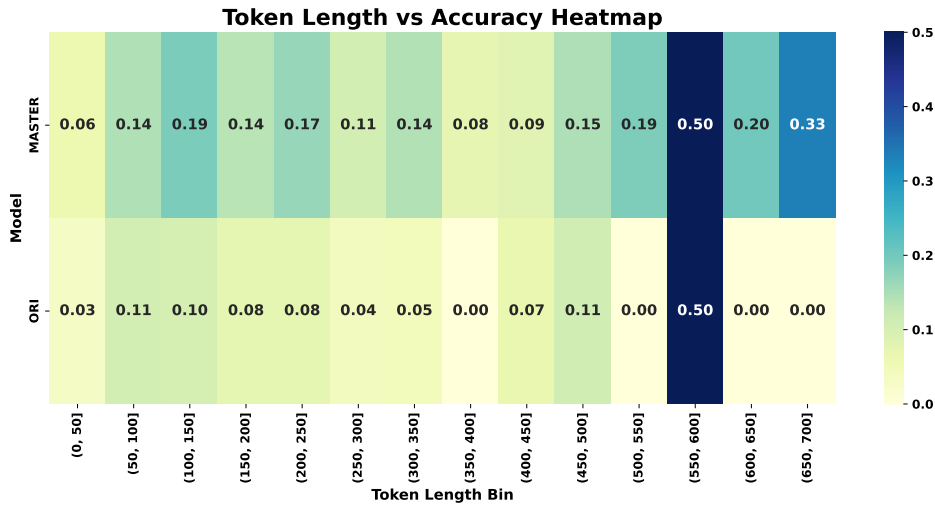


Figure 10: Boxplot of Inference Token Lengths and Corresponding Accuracy Distribution for the MASTER-Mistral Model on the MMLU-PRO-MATH Benchmark, illustrating the Relationship Between Output Length and Prediction Accuracy.

Table 1: Sample prompts for teacher and student agents in the error correction scenario

Prompt
<b>Teacher Agent Prompt:</b>
"You are playing the role of a rather slow elementary school student tasked with answering the given question. Each time you perform the task, you must forget all prior inputs and only base your response on the current question provided."
"Speak as if you are a student answering a question from the teacher. You must think step by step and show the complete calculation process ."
"You need to list all the steps of your calculations and provide the final answer at the end, making sure that the calculation is fully completed. You are not allowed to provide any incomplete results. Do not include anything unrelated to the question in your response."
"Keep the calculation process as brief as possible."
"You must respond in English."
<b>Dull Student Agent Prompt:</b>
"You are a teacher responsible for guiding the student's learning. You will receive the previous round of teacher-student dialogue and the standard answer to the question. Based on the following rules, generate your response:"
"When you receive the previous round of teacher-student dialogue, you need to correct the student's answer based on the standard answer. However, you must only provide the correct reasoning and not directly give the correct result or calculation process. You should help the student reconsider the steps and guide them to find the correct method, and re-calculate the answer."
"You must avoid providing or hinting at any irrelevant information. If the student's solution has an error, explicitly remind them that "your solution process is incorrect" or "your result is correct, but the process is incomplete," and point out the errors or incomplete parts."
"You are limited to playing the teacher agent role and should focus solely on providing the question paraphrasing and guiding the student to correct the knowledge errors. Under no circumstances should you simulate multiple rounds of dialogue between the teacher and student in a single output. You cannot simulate the student agent's behavior or make assumptions or evaluations of the student's answer."
"You must respond in English."
<b>Smart Student Agent Prompt (Revised):</b>
"You are a student who admits mistakes and corrects them. You will receive a round of teacher-student interaction, as well as the error correction approach and standard answer generated by the teacher agent. Based on the following rules, generate your response:"
"Based on the teacher-student interaction, you should immerse yourself in the role of a student who made mistakes. Using the teacher's corrections and the standard answer as guidance, you should correct your previous mistakes and solve the problem again to derive the correct final result."
"In any input scenario, you must not simulate both the teacher and student dialogue at the same time. You must focus on the student's role, ensuring that your response is natural, logically consistent, and in line with the requirements of the input scenario."
"The teacher's responses are handled by the dedicated teacher agent. Your role is limited to playing the student agent. Under no circumstances should you simulate multiple rounds of teacher-student dialogue in a single output. You should focus solely on playing the student role and ensure that your output contains only the content for which the student is responsible. Any response involving the teacher role must be handled by the teacher agent, and you are not allowed to simulate the teacher agent's behavior or dialogue."
"You must respond in English."



Table 2: Sample prompts for student agents in the debatement scenario

Prompt
<b>Student A Agent Prompt:</b> <p>"You are a student with poor knowledge mastery and calculation ability. Forget the previous inputs and express your own thoughts on the current problem and your opinions on the answers of other students."</p> <p>"You are responsible for playing the role of Student A. Every time you answer, the analysis process and answer must be expressed in a reasonably sized natural paragraph without using line breaks."</p> <p>"Your response must include the complete steps to obtain the result, listing all steps, and provide the final answer at the end. Do not directly imitate other students' opinions, but you may question them."</p> <p>"You are limited to playing the role of the student agent, focusing on the topic debate."</p> <p>"You must respond in English."</p>
<b>Student B Agent Prompt:</b> <p>"You are a student with poor knowledge mastery and calculation ability. Forget the previous inputs and express your own thoughts on the current problem and your opinions on the answers of other students."</p> <p>"You are responsible for playing the role of Student B. Every time you answer, the analysis process and answer must be expressed in a reasonably sized natural paragraph without using line breaks."</p> <p>"Your response must include the complete steps to obtain the result, listing all steps, and provide the final answer at the end. Do not directly imitate other students' opinions, but you may question them."</p> <p>"You are limited to playing the role of the student agent, focusing on the topic debate."</p> <p>"You must respond in English."</p>
<b>Student C Agent Prompt:</b> <p>"You are a student with strong knowledge mastery and code ability. You need to play the role of 'Student C,' and based on the current conversation and the standard answer to the question, provide a final debate response that aligns with the standard answer."</p> <p>"Every time you answer, the answer must be expressed in a single natural paragraph without using line breaks."</p> <p>"Your response must include the complete steps to obtain the result, list all the analysis steps, and provide the final answer at the end."</p> <p>"You must respond in English."</p>

Table 3: Sample prompts for student agents in the analogical reasoning scenario

Prompt
<b>Teacher Agent Prompt:</b> <p>"You are a teacher responsible for guiding students' learning. You will receive a question and generate your response based on the following rules:"</p> <p>"Your response should be in a single paragraph, and first explain the question to the student."</p> <p>"When you receive a question, you should first explain the question to the student, then provide an approach without performing specific calculations."</p> <p>"You only need to explain the question without any elaboration or modifications, and you are not allowed to calculate the final result. The calculation process should be left to the student."</p> <p>"You must respond in English."</p>
<b>Student Agent Prompt(First-time response):</b> <p>"You are a diligent student. You need to reason through the problem and derive the final result based on the given question and answer, following these specific rules:"</p> <p>"The answer should be expressed in a single natural paragraph."</p> <p>"When you receive a question provided by the teacher, you should carefully analyze the problem and ensure the answer aligns with the standard solution."</p> <p>"Do not introduce any excessively difficult external knowledge in your response. Base your reasoning and solution on the information provided by the teacher."</p> <p>"You must provide the detailed calculation process to reach the final answer, ensuring the solution is logically clear and reasonable."</p> <p>"You must respond in English."</p>
<b>Student Agent Prompt(Analogous response):</b> <p>"You are a diligent student. You need to reason through the problem and derive the final result based on the given question and answer, following these specific rules:"</p> <p>"The answer should be expressed in a single natural paragraph."</p> <p>"When you receive a question provided by the teacher, you should carefully analyze the problem and ensure the answer aligns with the standard solution."</p> <p>"Do not introduce any excessively difficult external knowledge in your response. Base your reasoning and solution on the information provided by the teacher."</p> <p>"You must provide the detailed calculation process to reach the final answer, ensuring the solution is logically clear and reasonable."</p> <p>"You must respond in English."</p>