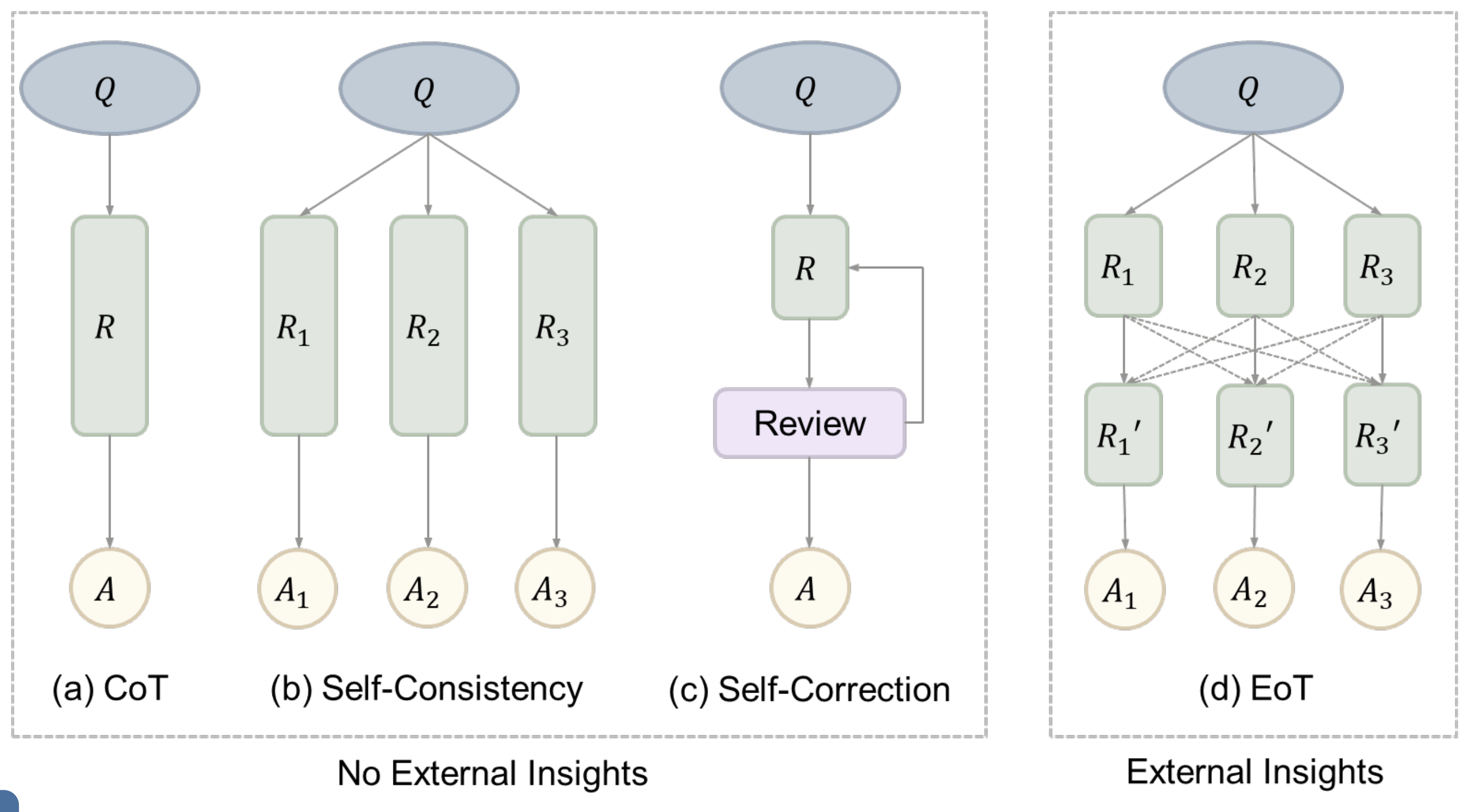
Exchange-of-Thought: Enhancing Large Language Model Capabilities through Cross-Model Communication

Zhangyue Yin, Qiushi Sun, Cheng Chang, Qipeng Guo, Junqi Dai,
Xuanjing Huang, Xipeng Qiu

Code & Paper

Motivation

- ❑ Chain-of-Thought and Self-Consistency in the reasoning process rely solely on the model's own understanding and perspective of the question, lacking external insights.
- ❑ Current research has found that the self-correction method, which amends responses through the model’s inherent capabilities, also struggles to enhance reasoning performance without external feedback.
- ❑ We propose **Exchange-of-Thought**, which allows models to exchange their analyses and problem-solving strategies during the reasoning process. Through role-playing, models incorporate the thoughts of their counterparts as external insights.

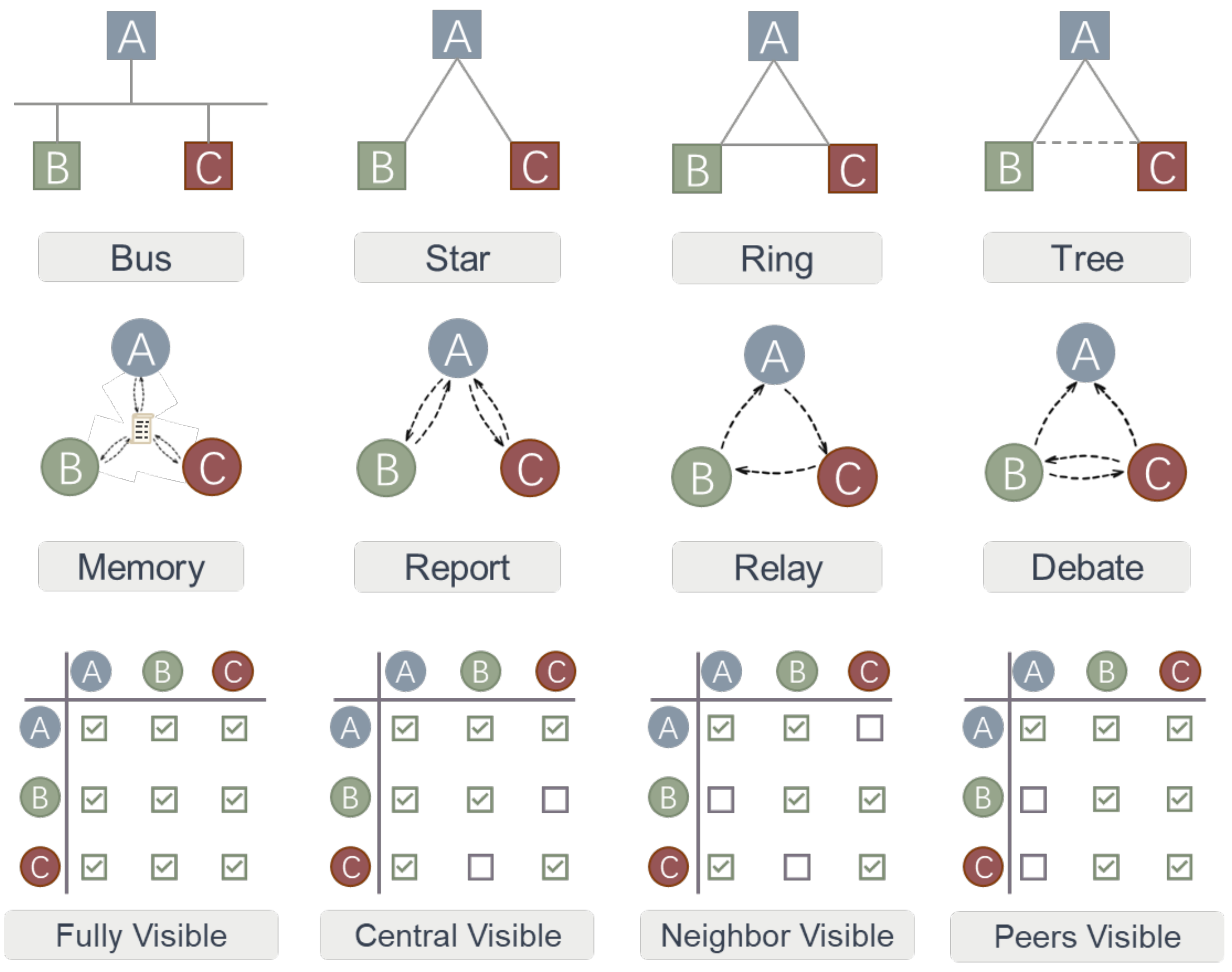


Methodology

Communication Paradigms

Inspired by network topology structures, we propose four communication paradigms:

- ❑ **Memory** (bus topology), where the thinking processes of all models are recorded in Memory and shared.
- ❑ **Report** (star topology), where the thinking processes of models are collected at a central node, and the central node's thought process is transmitted to each model.
- ❑ **Relay** (ring topology), where nodes are connected end-to-end to form a ring, with each node receiving information from the preceding node and sending its information to the following node.
- ❑ **Debate** (tree topology), where leaf nodes can exchange information, and parent nodes aggregate the information from leaf nodes, illustrating a bottom-up flow of information.



Confidence Evaluation

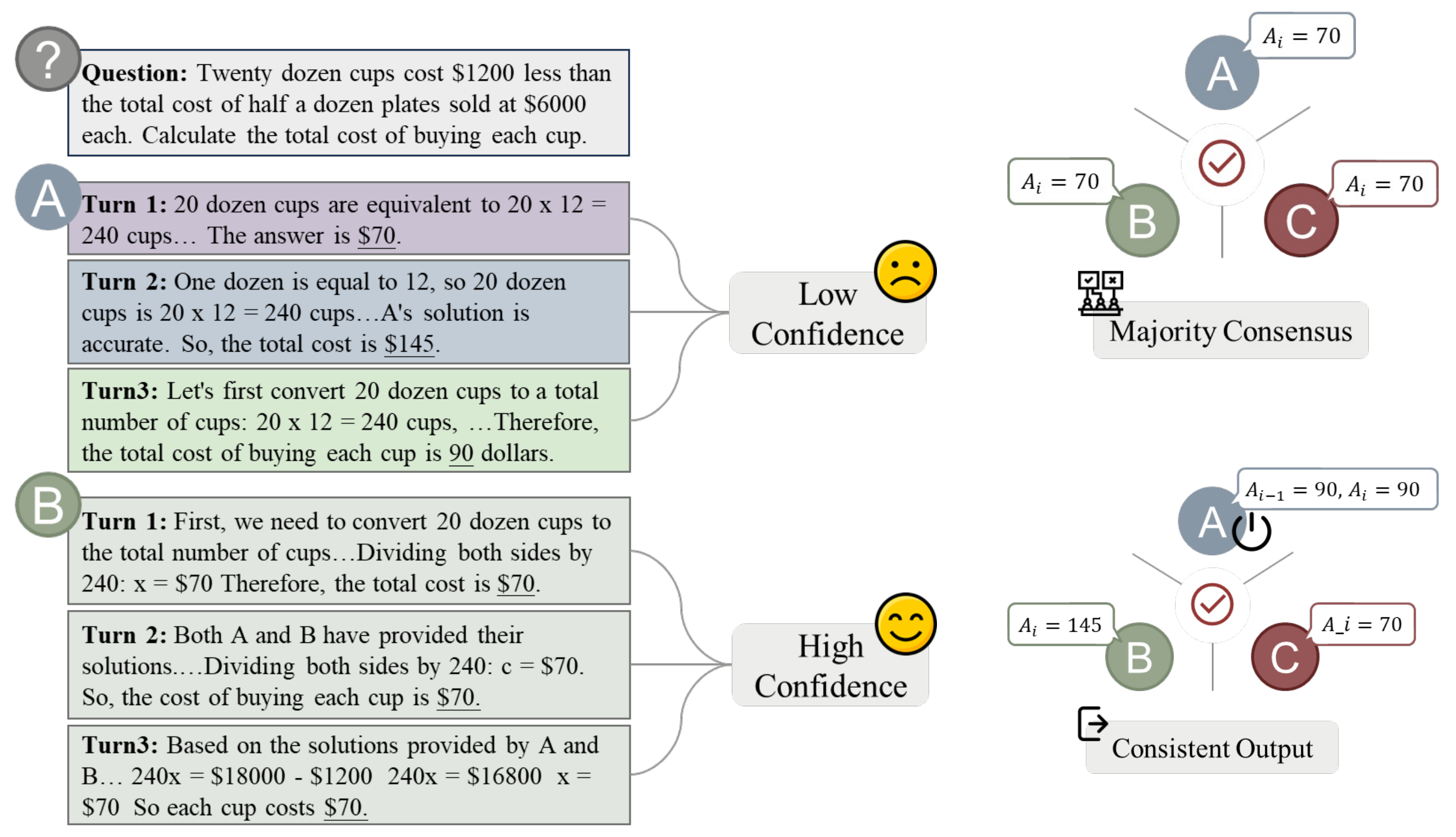
Confidence Evaluation: Evaluating the confidence by observing the changes in answers during the communication process.

- ❑ **Low confidence**: Frequently changing the answer.
- ❑ **High confidence**: Consistently sticking to one answer.

Termination Condition

Termination Condition: Stopping criteria for model communication.

- ❑ **Consistent Output**: Model exits the communication when its outputs are consistent between two consecutive interactions.
- ❑ **Majority Consensus**: Terminate when the majority of models reach a consensus on the answer.



Question: Twenty dozen cups cost $1200 less than the total cost of half a dozen plates sold at $6000 each. Calculate the total cost of buying each cup.

A

Turn 1: 20 dozen cups are equivalent to 20 x 12 = 240 cups… The answer is $70.

Turn 2: One dozen is equal to 12, so 20 dozen cups is 20 x 12 = 240 cups…A's solution is accurate. So, the total cost is $145.

Turn3: Let's first convert 20 dozen cups to a total number of cups: 20 x 12 = 240 cups, …Therefore, the total cost of buying each cup is 90 dollars.

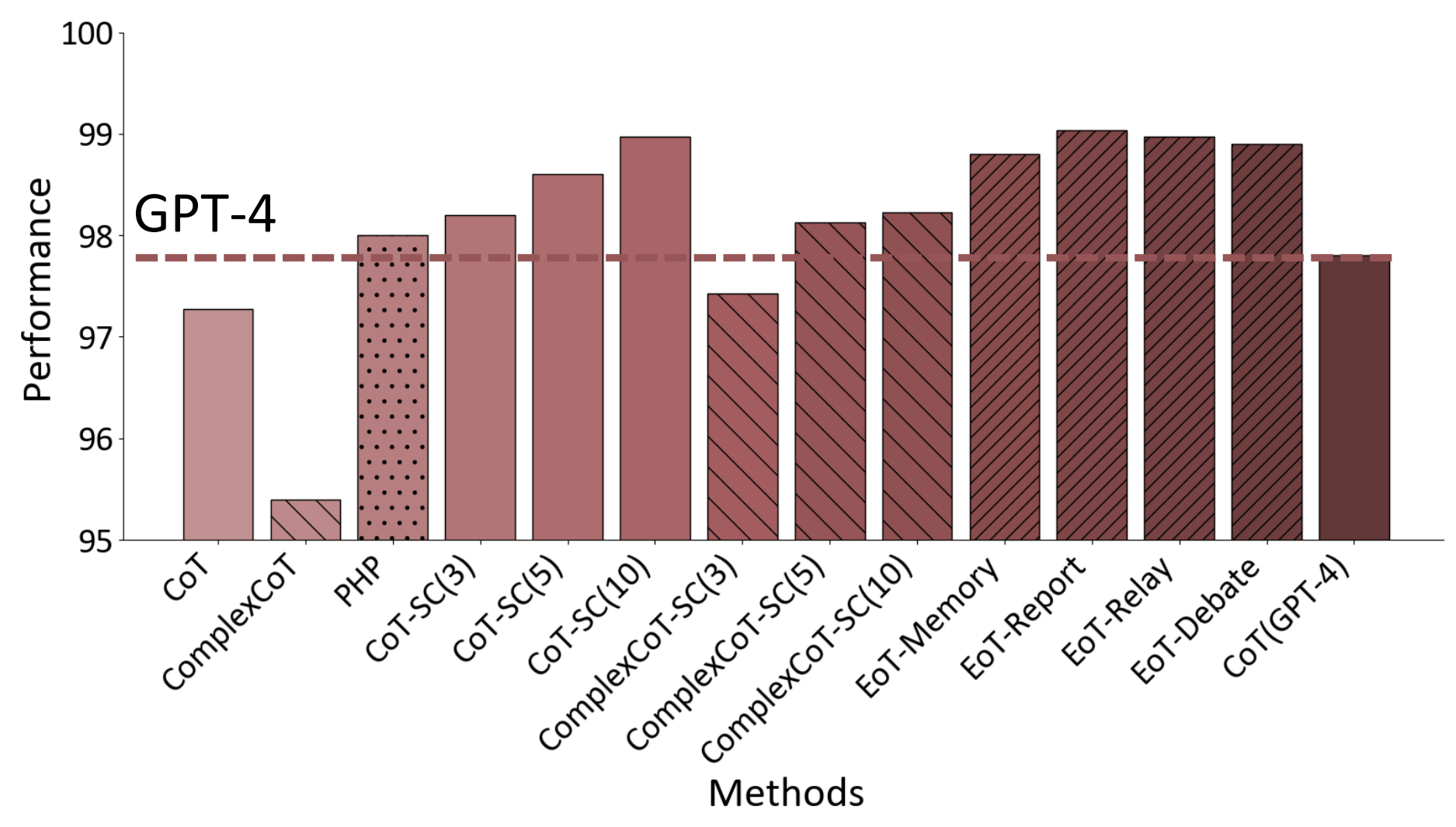
B

Turn 1: First, we need to convert 20 dozen cups to the total number of cups…Dividing both sides by 240: x = $70 Therefore, the total cost is $70.

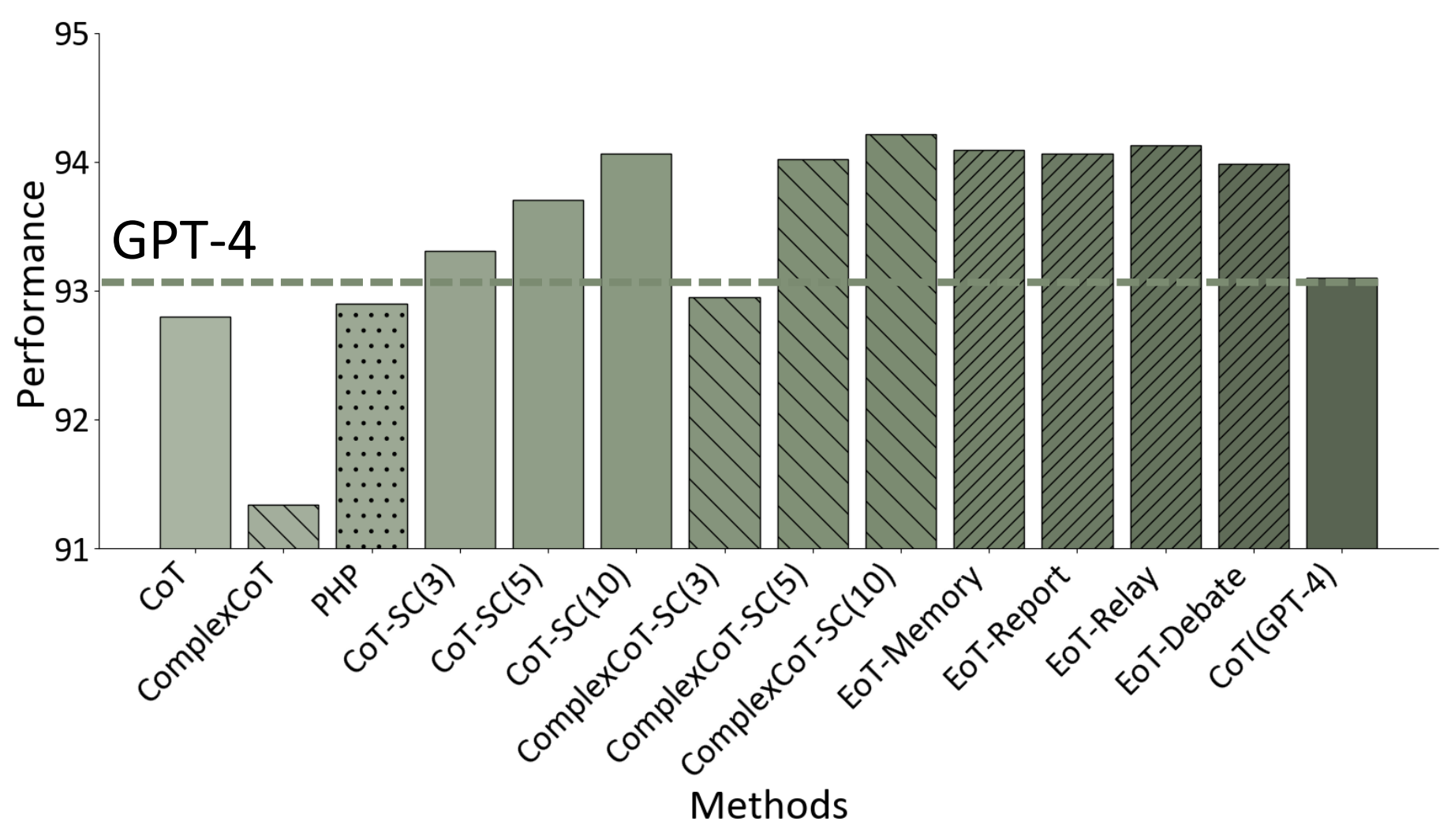
Turn 2: Both A and B have provided their solutions.…Dividing both sides by 240: c = $70. So, the cost of buying each cup is $70.

Turn3: Based on the solutions provided by A and B… 240x = $18000 - $1200 240x = $16800 x = $70 So each cup costs $70.

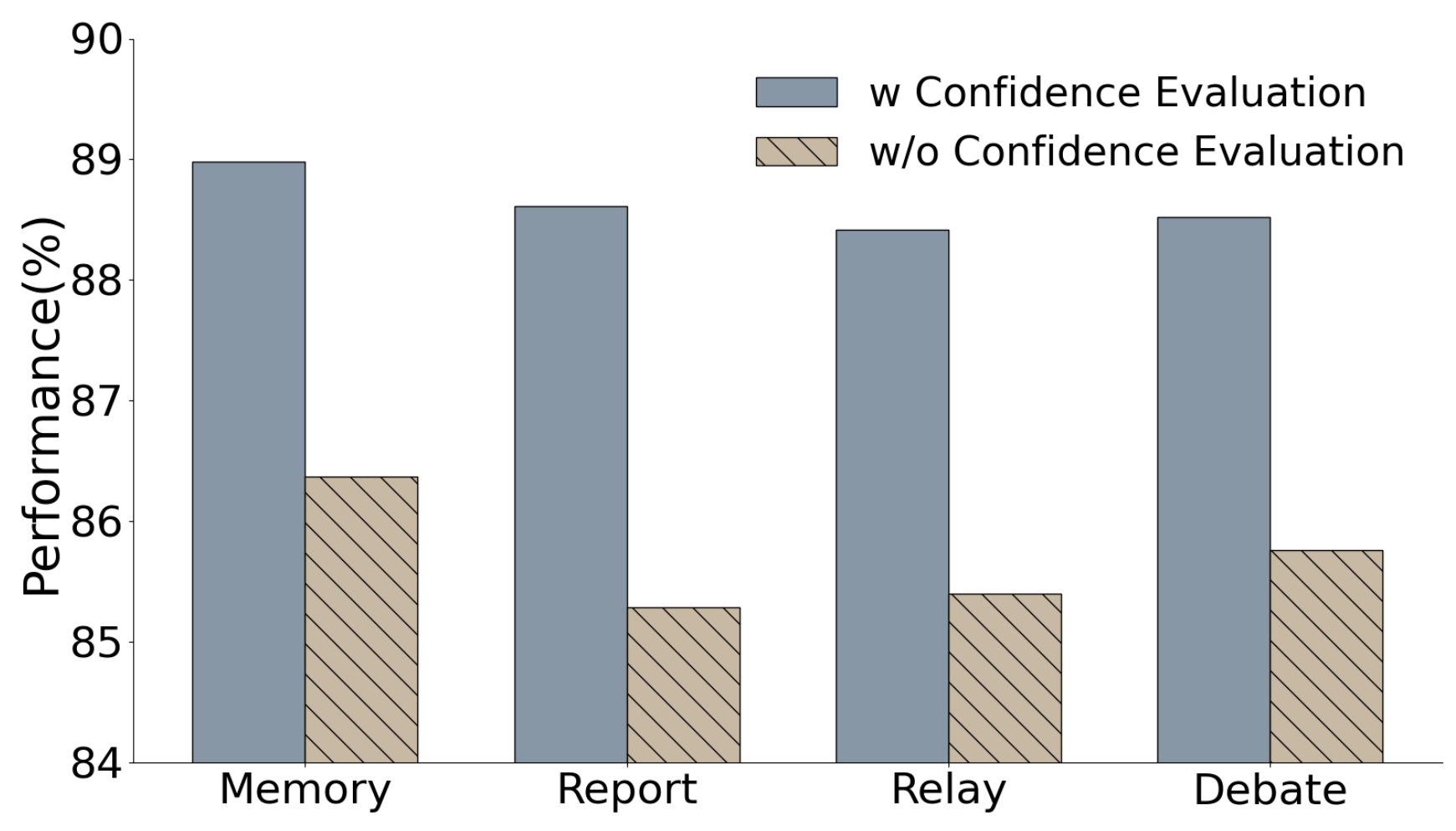
Experiment



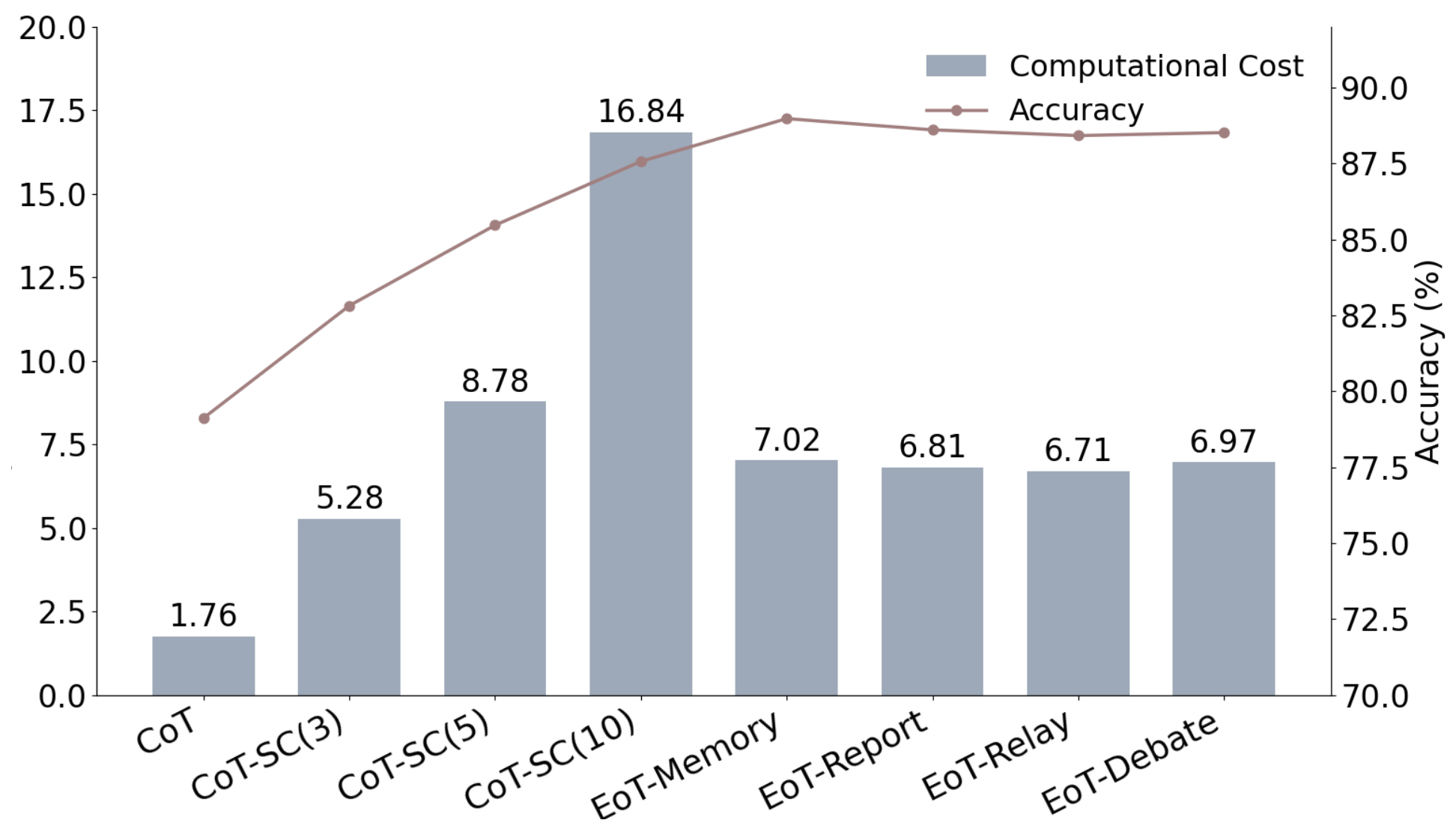
Performance on MultiArith



Performance on SingleEQ



Effectiveness of confidence evaluation on GSM8K



Cost & Performance Analysis on GSM8K

- ❑ **EoT vs. Self-Consistency:** EoT significantly outperforms voting-based methods in complex reasoning tasks, demonstrating superior effectiveness.
- ❑ **Performance Gains:** EoT enables three GPT-3.5-Turbo models to surpass a GPT-4 with CoT in some reasoning tasks, illustrating how EoT empowers weaker models to outperform stronger counterparts. Two heads are better than one!
- ❑ **More Reliable Answers :** EoT improves reasoning performance by scoring the reliability of information from other models, effectively managing information quality by confidence evaluations.
- ❑ **Cost-Effectiveness:** Comparing to Self-Consistency, EoT reaches notable performance enhancements while reducing costs by 20%, making it a more accessible choice for players with limited budgets.

EMNLP 2023 Poster