CreDes: Causal Reasoning Enhancement and Dual-End Searching for Solving Long-Range Reasoning Problems using LLMs

Anonymous Author(s)
Affiliation
Address
email

Abstract

Large language models (LLMs) have demonstrated limitations in handling combinatorial optimization problems involving long-range reasoning, partially due to causal hallucinations and huge search space. As for causal hallucinations, i.e., the inconsistency between reasoning and corresponding state transition, this paper introduces the Causal Relationship Enhancement (CRE) mechanism combining cause-effect interventions and the Average Treatment Effect (ATE) to guarantee the solid causal rightness between each step of reasoning and state transition. As for the long causal range and huge search space limiting the performances of existing models featuring single-direction search, a Dual-End Searching (DES) approach is proposed to seek solutions by simultaneously starting from both the initial and goal states on the causal probability tree. By integrating CRE and DES (CreDes), our model has realized simultaneous multi-step reasoning, circumventing the inefficiencies from cascading multiple one-step reasoning like the Chain-of-Thought (CoT). Experiments demonstrate that CreDes significantly outperforms existing State-Of-The-Art (SOTA) solutions in long-range reasoning tasks in terms of both accuracy and time efficiency.

1 Introduction

Reasoning aims to realize the causal transfer from the initial state to the goal state through several intermediate steps, which widely exists in the domains of Societal Simulation[1, 2, 3], Economic Simulation[4, 5, 6], Game Theory[7, 8, 9] and Gaming[10, 11, 12], etc. LLMs like GPT-3 have shown competitive performances in many reasoning tasks[13, 14, 15]. However, their performances and efficiency are limited when dealing with complex combinatorial optimization problems that require multi-step long-range reasoning[16].

The first challenge is causal hallucinations, i.e., causality between one-step reasoning (OSR) and state transition in LLMs is not always guaranteed. Similar to pre-trained LLMs that are prone to produce hallucinations when processing certain factual information, causal hallucinations reflect the fact that LLMs lack rigor due to inherent randomness in accomplishing complex mathematical[17, 18, 19], logical[20, 21], or common-sense reasoning[22, 23, 24], which is somehow entrenched in statistical inevitability and independent of the Transformer architecture or data quality[25]. For example, CoT-based finite-step reasoning methods[26, 27] suffer from causal hallucinations, which cannot effectively ensure the causality between OSR and state transition in LLMs, resulting in unreliable reasoning and relatively low success rates (especially for long-range reasoning problems with significant error accumulation effects). The reasonableness between OSR and state transition can be summarized as follows: There is a causal relationship between reasonable OSR and state transition. At the same time, there is only a correlation or no relationship between unreasonable OSR and state transition, which

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

suggests that training with cross-entropy loss alone does not enable the model to have sufficient causal rigor. Inspired by this, we designed the CRE mechanism to make each step of reasoning correct and *causally sound* by including the causality measure between OSR and state transition as part of the training objective, thus more closely modeling the rigor, adaptability, and comprehensiveness of human reasoning[28].

The second challenge is that long-range reasoning problems have a huge search space. Although complex architectures such as CoT, Tree of Thought (ToT)[29], and Program of Thought (PoT)[30] can effectively improve the reasoning accuracy of LLMs through external guidance, they are limited when handling long-range reasoning processes and task decomposition. A crucial reason is that long-range reasoning has a huge state space, i.e., each branch in the state transition process expands the search space approximately exponentially. Most of the existing LLM-based methods, e.g., Monte Carlo search[31], are based on unidirectional reasoning, making them time-inefficient and easy to fall into local optima when dealing with reasoning problems with large search spaces. In this paper, a bi-directional Dual-End Searching method is developed, which first decomposes a long-range reasoning problem into a combination of short-range reasoning problems and then searches for the intersection of two causal probability trees starting from the initial and goal states, respectively.

A structured and generalized reasoning framework, CreDes, is developed for long-range reasoning with LLMs in this paper, and the contributions can be summarized as follows:

First, the CRE mechanism is introduced to improve the rigor of LLM-based long-range reasoning methods: Structural Causal Modeling (SCM) is exploited to enhance the causality between OSR and state transitions, involving performing causal interventions and optimizing the absolute value of ATE during training, which has effectively alleviated causal hallucinations in long-range reasoning of LLMs.

Second, the DES method is developed to improve the search efficiency for long-range reasoning: After constructing causal probability trees starting from the initial states and ending at the goal states, long-range reasoning (e.g., 12 steps) is transformed into more manageable combinations of smaller segments (e.g., 2 or 4 steps) by minimizing the distances between leaves of the tree and employing end-matching techniques. By avoiding long-range sequential search from scratch, the DES method has greatly lowered the complexity when solving long-range reasoning problems.

Third, simultaneous multi-step reasoning is realized to improve the time-efficiency of long-range reasoning: By integrating CRE and DES, CreDes can perform simultaneous multi-step reasoning within the model, i.e., avoiding the inefficiency of *cascading single-step reasoning* in frameworks such as CoT. While ensuring the accuracy of the reasoning process, CreDes can significantly reduce the time required for multi-step reasoning in LLMs.

Fourth, adequate and rigorous testing of CreDes: CreDes has been extensively tested in the Blocksworld, GSM8K, and Hanoi Tower scenarios, respectively, and the experimental results show that CreDes outperforms existing SOTA regarding reasoning accuracy and time efficiency.

2 Related Work

Decision-Making Capabilities in LLMs: The core of intelligence partially lies in planning, which encompasses generating a sequence of actions aimed at accomplishing a predefined objective[32, 33]. Classical planning methods have found extensive application in robotics and embodied environments, where they are commonly employed to guide decision-making processes externally[34, 35]. Recent advancements, such as the Chain-of-Thought model[26, 36, 37], have significantly bolstered the LLMs' capability to perform detailed reasoning[38, 39, 40]. This model breaks down intricate queries into a series of manageable steps, thereby enhancing the LLMs' decision-making ability. Subsequent initiatives like ReACT[41] have modified this approach to improve reasoning ability in decision contexts using a CoT-based framework. Additionally, Reflexion[42] provides a corrective mechanism that enables LLMs to recognize their errors during the decision-making process, reflect on these mistakes, and make accurate decisions in subsequent attempts. Further developments have led to the creation of tree-based decision-making frameworks that tailor LLM capabilities to specific scenarios. The Tree-of-Thought[29] utilizes Breadth First Search (BFS) and Depth First Search (DFS) algorithms to facilitate decision-making in activities such as the Game of 24, Creative Writing, and Mini Crosswords. Meanwhile, Reasoning via Planning (RAP)[43] employs the Monte Carlo

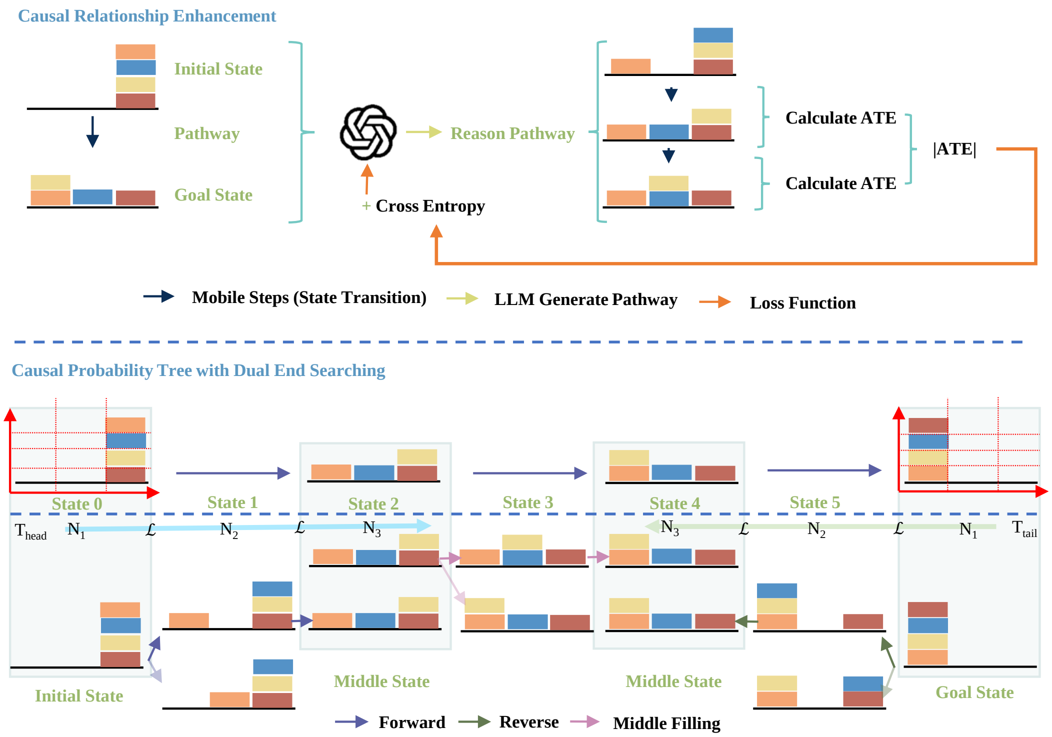

Figure 1: Integrating Causal Relationship Enhancement (CRE) and Dual-End Searching (DES).

Tree Search technique to optimize solutions across tasks like Blocksworld[44], Math Reasoning[45]. DFSDT[46] proposed an efficient version of DFS for LLMs to make decisions, but it lacks the judgment ability to evaluate different decisions. JUDEC[47] utilizes an Elo rating system to enable LLMs to develop self-assessment capabilities, thereby enabling them to generate optimal solutions for a wide range of real-world tasks, independent of any task-specific expertise. Lastly, Graph-of-Thought[48] represents the thoughts as nodes in a graph, combining thoughts non-sequentially. Encouraged by the studies above, we leveraged LLMs to solve long-range reasoning problems.

Integrating Causal Analysis in LLMs for Multi-step Decision-Making: The causal analysis aims to discern and elucidate the causal relationships between actions, circumstances, or decisions. This method entails investigating the origins or causes leading to an event and the potential consequences that follow[49, 50, 51]. Although various causal models may produce identical observational distributions, they can yield distinct distributions when interventions are applied[52]. Therefore, using interventions allows for the distinction of possible causal frameworks that align with the observed data[53, 54]. Previous work suggests that, while CoT has been lauded for its potential to improve task performance, its application does not always lead to enhanced outcomes[36, 55]. Also, research has shown that the statistical pretraining of LLMs encourages models to achieve high empirical performance but not necessarily to reason[56, 57, 58, 59]. Inspired by this, we designed the CRE mechanism to control the causal hallucinations of LLMs to solve long-range reasoning problems.

Solving Multi-step Problems with LLMs: Recent studies have shown that with substantial design, LLMs are capable of performing not only basic arithmetic tasks but also complex multi-step reasoning[60, 61]. For instance, increasing computational resources significantly enhances the accuracy of datasets like GSM8K[62]. Concurrently, Research[63] demonstrated that a 2B parameter LLM could achieve 89.9% accuracy in 5x5 multiplication tasks using curriculum learning with 50 million training instances. This evidence suggests that adequately scaled LLMs can process multiple reasoning steps effectively internally. While trees are frequently used to represent games (especially extensive-form games[64, 65]) and sequential reasoning problems[66], it was Shafer's groundbreaking work[67] that initially established a framework for understanding causality through the use of probability trees. Inspired by Shafer's approach, we recognized that LLMs tend to struggle with long-range reasoning problems involving multiple steps but excel in short-range reasoning tasks.

By integrating causal probability trees, we can enhance search efficiency. This insight led to the development of DES.

3 Method

The pipeline of CreDes is illustrated in Fig. 1. It comprises two main components: CRE and DES. In CRE, the inputs of LLMs for training are the initial state, goal state, and pathway (containing a series of OSRs), while for testing, the inputs are the initial and goal states only. The DES starts from the initial and goal states of the probability tree, expands them into two intermediate states, and uses the CRE-trained model to infer the pathway between them, ultimately producing the complete pathway.

3.1 Problem Definition

To further improve the capability of LLMs in solving combinatorial optimization problems that involve a finite number of discrete intermediate steps, we conducted experiments using the Blocksworld and Hanoi Tower datasets with 7B parameter models. The Blocksworld dataset includes 602 test cases categorized by the minimum number of required actions, ranging from 2 to 12 steps. For Hanoi Tower, cases are grouped based on the complexity related to the number of disks and poles, which directly influences the solution steps.

For each category, our model is trained on 80 samples without common instructions. In the reasoning process, the following elements are included: initial state, OSR, state transition, next state, and goal state, as shown in Fig. 2. During testing, the model was tested on new, categorically similar samples from different datasets, assessing its ability to transform the initial state to the goal state successfully.

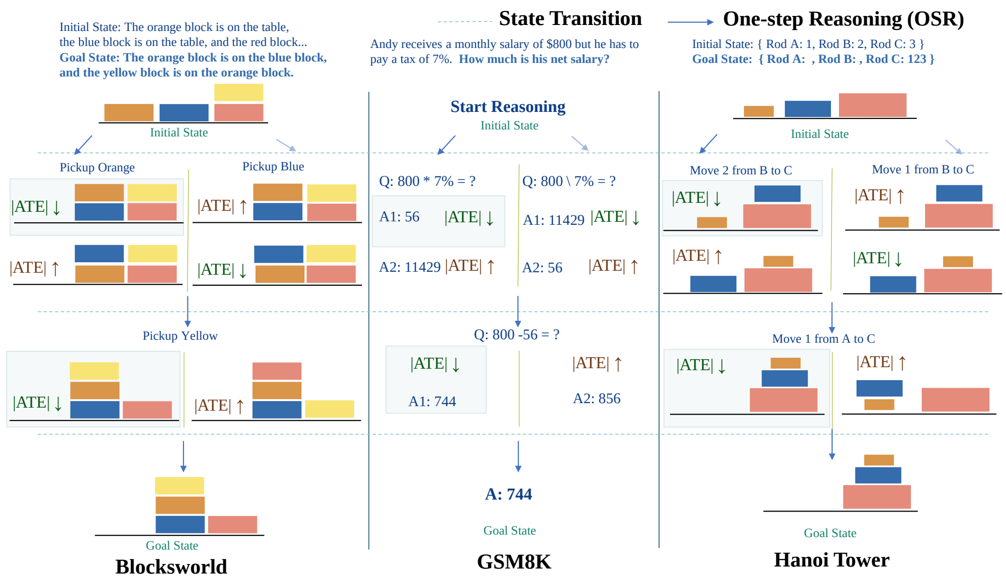

Figure 2: Schematic illustration of Causal Relationship Enhancement(CRE).

3.2 Causal Relationship Enhancement (CRE)

Firstly, all the samples are classified into two categories: Correct and Incorrect. Within the Incorrect category, three scenarios exist, i.e., a correct OSR leading to an incorrect state transition, an incorrect OSR leading to an incorrect state transition, and an incorrect OSR resulting in a correct state transition. Given this, it is evident that we need to strengthen the causal connection between the OSR and the transition, and reduce the occurrence of samples where the OSR and the transition are non-causal. In CRE, we first use the ATE to estimate the causality between OSR and state transition quantitatively, and then embed the $|\mathrm{ATE}|$ into the loss function in the training process (the remaining is cross-entropy), enhancing the causality of state transitions. As is shown in Fig. 2, we leave the reasoning path selection to be controlled by the cross-entropy loss, while the suppression of hallucinations is handled by the |ATE| loss. Perplexity (PPL) is a metric used to evaluate the performance of a LLM, indicating how well the model predicts the next word in a sequence, and lower values signify better predictive accuracy. The estimation of ATE is detailed as the follows:

Given binary variables X and Y indicating the correctness of OSR and next state (state transition), respectively, i.e., $X, Y \sim B(0,1)$, and $X = 1$ (or $Y = 1$) means correctness. First, we calculate the cause-effect interventions between X and Y, then subsequently modify the distribution of Y by intervening in X. From a statistical correlation perspective, if X and Y are correlated, Y can be predicted using X. However, if there is no causal relationship between X and Y, intervening in X will not alter the distribution of Y. Hence, if X and Y are correlated but not causally linked, then manipulating or intervening in X would not lead to any changes in the distribution of Y. This distinction is crucial in statistical analysis and experimental design because it addresses the potential fallacy that correlation inherently means causation.

$$P_Y^{do(X)} = P(Y|do(X=1)) - P(Y|do(X=0)) \tag{1}$$

In (1) and (2), $do(\cdot)$ refers to Do-calculus[68], which denotes an external intervention on the value of X without affecting the actual state of Y. Using interventions independent from other variables, we can obtain whether the treated variable X causes the target variable Y. Consequently, we can use ATE[69] to estimate the effect of the intervention, which compares the distributions of the target variable Y with and without the treatment. Y_x is the potential outcome of Y under the intervention $X = x$. Then ATE is defined as follows:

$$\mathrm{ATE} = E(Y|do(X)) - E(Y) = E[Y_1 - Y_0] \tag{2}$$

Based on (2), under the intervention, the proportion of positive and negative cases (hallucinations) in the model output samples remains roughly unchanged; the more robust the causal relationship between different OSRs and corresponding positive and negative cases, the lower the |ATE|. The reason is that cross-entropy basically ensures the majority of positive cases. At the same time, |ATE| reduces the occurrence of negative cases, making the distribution of positive and negative cases more stable. Consequently, we incorporate the ATE into the loss function, as is shown in (3) and (4), $p_{1|X}$ and $p_{0|X}$ denote the conditional probabilities of Y being *1* and *0*, respectively, given the state of X.

$$\mathcal{L}_{CrossEntropyLoss} = -\left[Y \log(p_{1|X}) + (1-Y)\log(p_{0|X})\right] \tag{3}$$

$$\mathcal{L}_{Loss} = \mathcal{L}_{CrossEntropyLoss} + |\mathrm{ATE}| = \ln(\mathrm{PPL}) \tag{4}$$

We estimated the probabilities of correct and incorrect (hallucinations) samples in each path separately, and take |ATE| as part of the loss function for each case based on the sampling results between different paths, where |ATE| is smaller for the category with strong causal effects. This process allows the model to internalize the logical judgment between OSR and the next state during training, i.e., correct answers with strong causal effects and low |ATE|, and wrong answers with weak causal effects and high |ATE|. Therefore, with the loss function composed of cross-entropy and ATE, we can realize the synergistic optimization of path selection and hallucination elimination simultaneously.

3.3 Causal Probability Trees with Dual End Searching (DES)

In this section, we improve the success rate of LLMs when solving long-range reasoning problems, like the 12-step scenarios in Blocksworld, by leveraging its higher success rates in simpler 2-step and 4-step scenarios. We construct two causal probability trees from the initial and goal states. Each node represents a state in the reasoning process, with arrows showing causal relationships. These trees outline possible reasoning outcomes within a limited number of intermediate steps. By matching the leaves of both trees, we identify several end-to-end permutation schemes to form a continuous and feasible path, as shown in the bottom of Fig. 1.

DES first calculates the ATE between tree unfolding and distance reduction, which in turn clarifies the causal relationship between tree unfolding and distance reduction and then infers a better unfolding direction and pruning process. At this point, the ATE calculation formula is:

$$\mathrm{ATE}(A) = E(A|do(B)) - E(A) \tag{5}$$

Where A is the decrease in distance D of N_i relative to N_{i-1} and B is the number of unfolded layers where the current leaf is located N_i. We utilize the spatial positioning numbers of the blocks (disks) to calculate the distance D, and estimate the ATE of the reduction in distance, denoted as δD, relative to the number of layers N_i of tree expansion. The distance is obtained by calculating the Euclidean distance for the current position of the block and the coordinates of the target position.

In each layer of the tree expansion, we calculate the distance by comparing the current position with their target positions, ensuring that the reduction in distance and the direction of the tree's expansion have a strong causal effect. To avoid the expansion direction falling into local optimum, we conduct counterfactual assessments, hypothesizing alternative expansion routes that might have been taken during the random expansion process, and incorporating the causal impacts of these hypothetical routes into consideration. Both these values are summed up to form the loss function, taking into account both the head T_{head} and tail T_{tail} trees.

$$\mathcal{L} = |\mathrm{ATE}(\delta D_{T_{head}}^{N_i - N_{i-1}})| + |\mathrm{ATE}(\delta D_{T_{tail}}^{N_i - N_{i-1}})| + D \tag{6}$$

During the expansion process of the probability trees at both ends, we intervene by minimally altering the $\mathcal{L}$, directing the expansion toward our desired outcome. Minimizing $\mathcal{L}$ realizes the pruning and unfolding direction judgment, prioritizing the direction with the lowest $\mathcal{L}$ as the unfolding direction. The whole process of DES is in **Algorithm 1**.

Algorithm 1 DES (Taking the 12-step Blocksworld as an example)

```
Input: State_init and State_goal, denoting the initial and goal states, respectively
Output: Complete 12-step solution process
Construct T_head and T_tail from State_init and State_goal
Match leaves of T_head and T_tail to form paths
for every four steps do
    Determine intermediate steps and fill in details
end for
for expanding T_head and T_tail do
    Calculate distance D
    Minimize L
    if local optimum detected then
        Assess alternative routes
    end if
end for
```

4 Experiment

In this section, we validated the effectiveness of CreDes compared to baseline approaches.

4.1 Setup

Blocksworld: There are n blocks initially placed randomly on a table[44]. The LLM's goal is to stack these blocks in a specified order. The LLM can perform four actions: pick up a block from the table, put down a block it is holding onto the table, unstack a block from another to hold it, and stack the block in its hand onto another block. The LLM can only manipulate one block at a time, and blocks with others on top are immovable.

GSM8K: The GSM8K dataset[62] includes 1,319 diverse grade school math word problems curated by human problem writers. These tasks typically begin with a description and culminate in a final question requiring multi-step mathematical calculations contextual to the problem. To effectively tackle the final question, our approach involves decomposing it into a sequential series of smaller sub-questions, allowing for a structured solution process.

Hanoi Tower: The Hanoi Tower problem[70], a classic puzzle involving three pegs and a set of discs of varying sizes, serves as a key component of our experimental setup. The challenge requires moving the entire stack of discs from one peg to another, obeying the rules that only one disc can be moved at a time, and no disc may be placed on top of a smaller one. This task, structured around sequential and strategic disc placement, tests the model's ability to plan and execute a series of actions based on simple yet strict rules.

4.2 Dataset and Basemodel

Dataset: The datasets we used are the open source datasets Blocksworld[44], GSM8K[62], AQUA[71], QASC[72], and our own production of Hanoi Tower. where the experiments for AQUA and QASC are in the Table 4.

Basemodel: The pre-trained models used in our study include: LLAMA-2-7B[73], Phi-2-7B[74], Mistral-7B[75] and Mixtral-8x7B[76], Qwen1.5-7B[[77]], TAIDE-LX-7B[1], Mpt-7B[[78]], Baichuan2-7B[[79]],The model test results not mentioned in the main text will be supplemented in the Appendix.

4.3 Benchmark

Train Parameter: In this paper, we primarily utilize the 7B models for training on a single NVIDIA A100 GPU and models are loaded in 4-bit.

RAP: A technique that employs Monte Carlo Tree Search (MCTS) for exploration[43]. RAP transforms LLMs into both reasoning agents and world models, utilizing MCTS for strategic exploration and decision-making. This approach significantly enhances the LLM's ability to generate action plans and solve mathematical and logical problems, outperforming traditional methods and establishing new benchmarks in LLM's capabilities.

CoT[26]: A technique having enhanced the reasoning capabilities of LLMs. By providing models with intermediate reasoning steps as examples, CoT demonstrates notable improvements across various complex reasoning tasks, including arithmetic, commonsense, and symbolic reasoning. CoT requires the model to generate a reasoning chain to improve the reasoning ability. We used all basemodels to carry out CoT in the experiment.

RoT: A framework[80] to enhance the performance of tree-search-based prompting methods used in LLMs. This innovative approach leverages guidelines derived from past tree search experiences, allowing LLMs to avoid repeating errors and significantly improving their reasoning and planning capabilities across various tasks. We not only used the same basemodel as the original RoT, but also introduced other 7B models as a comparison.

4.4 Results

Blocksworld: We conducted ablation experiments on the Blocksworld dataset. Our methodology, detailed in Section 3, particularly focuses on scenarios with more than 6 steps. As is shown in Table 1 and Table 5, for tasks up to 6 steps, results with our 7B models closely matched those with the benchmark's 70B models, suggesting robust inference capabilities even with reduced model size. For more complex tasks of 8 steps or more, DES improved its success rates by breaking down tasks into simpler segments, though it slightly lagged behind in performance compared to shorter tasks. This approach underlines the potential of our modified strategies in handling varying task complexities. By comparison, our CRE method not only outperforms benchmarks in terms of success rates on the 7B scale, but also achieves a higher success rate than the 70B+RAP method using the 7B model. For the arithmetic cases that use the full CreDes architecture, CreDes helps to improve the performance of the LLMs for long-range reasoning tasks.

GSM8K: We further independently verified the capabilities of CRE based on the GSM8K dataset without introducing DES, to confirm that it helps to enhance the inference capabilities of large models. We found that our CRE is superior to the baseline methods RAP, RoT, and CoT, further demonstrating that completing multi-step reasoning in one go has more advantages than completing multiple single-step reasoning. See Table 2. This example shows that CRE can not only help LLM solve highly structured problems, such as Blocksworld, but also has the ability to assist in solving some abstract mathematical problems.

Hanoi Tower: Unlike the Blocksworld case, the longest reasoning steps for the Hanoi Tower have a fixed quantitative relationship with the number of rods and disks. Therefore, when training the model, we used combinations within 7 steps, i.e., 3 rods and 3 disks. For evaluation, we used problems within 15 steps, i.e., combinations of 3 rods and 4 disks, to test the reasoning ability. From this

[1]http://taide.tw

perspective, our reasoning process is based on a zero-shot setting. Due to the time complexity of the search-based method for long-range reasoning, we did not conduct experiments for too many reasoning steps, and its success rate can be recorded as '-.' As Table 3 shows, CreDes performed best among all the models. By comparing the Hanoi Tower scenario with the Blocksworld scenario, we find that the success rate under Hanoi Tower is lower than that of Blocksworld, and that the reasoning ability of the 7B+CRE group is slightly lower than that of the 70B+RAP group. We believe that this phenomenon occurs because Hanoi Tower has a stricter stacking order qualification relative to Blocksworld, and some of the intermediate steps may not hold at all, see Fig. 2. From the results, the complexity of the Hanoi Tower problem is higher than that of Blocksworld.

Time Efficiency: Using the CRE and DES architecture has significantly shortened the time to complete long-range reasoning tasks compared to benchmarks, as is shown in Fig.3. This is because CreDes can perform simultaneous multi-step reasoning, which is more efficient than other methods that generate answers multiple times and then cascade them together, which is more evident in longer-range reasoning.

Table 1: Succcess Rate under Blocksworld

Model	2-step	4-step	6-step	8-step	10-step	12-step
Llama-2-70B + RAP	0.67	0.76	0.74	0.48	0.17	0.09
Llama-2-7B + RAP	0.39	0.41	0.37	0.11	0.00	0.00
Llama-2-7B + CoT	0.50	0.63	0.40	0.27	0.07	0.00
Llama-2-7B + RoT	0.52	0.67	0.27	0.06	0.00	0.00
Llama-2-7B + CRE	**0.95**	**0.80**	**0.76**	0.22	0.09	0.00
Llama-2-7B + CreDes	-	-	-	**0.68**	**0.51**	**0.34**
Phi-2-7B + RAP	0.40	0.44	0.33	0.00	0.00	0.00
Phi-2-7B + CoT	0.43	0.05	0.01	0.00	0.00	-
Phi-2-7B + RoT	0.54	0.16	0.01	0.01	0.00	-
Phi-2-7B + CRE	**0.91**	**0.86**	**0.79**	0.19	0.05	0.00
Phi-2-7B + CreDes	-	-	-	**0.46**	**0.31**	**0.19**
Mistral-7B + RAP	0.49	0.41	0.35	0.07	0.00	0.00
Mistral-7B + CoT	0.84	0.41	0.24	0.05	0.08	-
Mistral-7B + RoT	0.81	0.49	0.21	0.10	0.12	-
Mistral-7B + CRE	**0.97**	**0.94**	**0.82**	0.24	0.12	0.03
Mistral-7B + CreDes	-	-	-	**0.54**	**0.37**	**0.21**
Mixtral-8x7B + RAP	0.49	0.44	0.35	0.15	0.04	0.00
Mixtral-8x7B + CoT	0.81	0.63	0.55	0.18	0.20	-
Mixtral-8x7B + RoT	0.87	0.71	0.55	0.29	0.27	-
Mixtral-8x7B + CRE	**0.99**	**0.97**	**0.93**	0.34	0.22	0.13
Mixtral-8x7B + CreDes	-	-	-	**0.75**	**0.57**	**0.40**

Table 2: Accuracy under GSM8K

Model	RAP	RoT	CoT	**CRE**
Llama-2-7B	0.51	0.54	0.47	**0.92**
Phi-2-7B	0.45	0.48	0.48	**0.89**
Mistral-7B	0.39	0.32	0.31	**0.85**
Mixtral-8x7B	0.48	0.50	0.49	**0.90**

4.5 Discussion

This study introduced the CreDes framework, which combines CRE and DES to improve LLMs' ability to handle long-range reasoning tasks. CRE ensures robust causal relationships between reasoning steps, and DES can lower the complexity of long-range reasoning by using a bidirectional search approach. Our experiments, particularly in the Blocksworld and Hanoi Tower scenarios,

Table 3: Succcess Rate under Hanoi Tower

Model	3-step	5-step	7-step	9-step	11-step	13-step
Llama-2-70B + RAP	0.57	0.42	0.22	0.07	-	-
Llama-2-7B + RAP	0.29	0.21	0.11	0.00	-	-
Llama-2-7B + CoT	0.34	0.23	0.10	0.02	0.00	0.00
Llama-2-7B + RoT	0.41	0.27	0.13	0.04	-	-
Llama-2-7B + CRE	**0.45**	**0.39**	**0.24**	0.12	0.01	0.00
Llama-2-7B + CreDes	-	-	-	**0.27**	**0.14**	**0.07**
Phi-2-7B + RAP	0.27	0.21	0.14	0.01	-	-
Phi-2-7B + CoT	0.33	0.0.22	0.10	0.02	0.00	0.00
Phi-2-7B + RoT	0.24	0.12	0.02	0.00	-	-
Phi-2-7B + CRE	**0.40**	**0.25**	**0.17**	0.03	0.00	0.00
Phi-2-7B + CreDes	-	-	-	**0.33**	**0.20**	**0.09**
Mistral-7B + RAP	0.34	0.25	0.14	0.04	-	-
Mistral-7B + CoT	0.40	0.32	0.21	0.09	0.00	0.00
Mistral-7B + RoT	0.35	0.22	0.17	0.02	-	-
Mistral-7B + CRE	**0.49**	**0.37**	**0.26**	0.15	0.03	0.00
Mistral-7B + CreDes	-	-	-	**0.37**	**0.19**	**0.11**
Mixtral-8x7B + RAP	0.40	0.24	0.15	0.06	-	-
Mixtral-8x7B + CoT	0.45	0.27	0.14	0.02	0.00	0.00
Mixtral-8x7B + RoT	0.37	0.22	0.10	0.00	-	-
Mixtral-8x7B + CRE	**0.50**	**0.35**	**0.22**	0.11	0.01	0.00
Mixtral-8x7B + CreDes	-	-	-	**0.42**	**0.25**	**0.12**

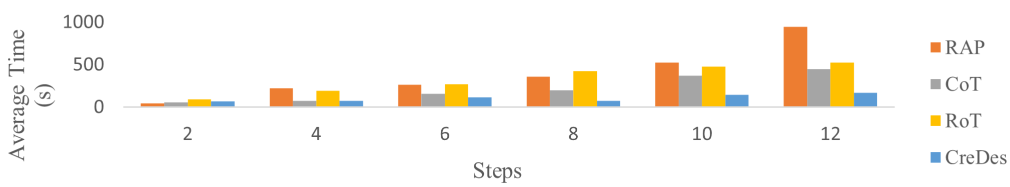

Figure 3: Improvement in reasoning speed for long-range tasks (based on a single A100 GPU).

demonstrated significant improvements in accuracy and efficiency over existing methods, implying that CreDes can effectively address the problem of causal hallucinations and huge search spaces.

4.6 Limitation

In scenarios with strict order of precedence, such as the Hanoi Tower, the accuracy is significantly lower compared to tasks like Blocksworld. The DES approach, while effective for moderate-length tasks, struggles with very long reasoning steps, leading to a decline in performance. Additionally, maintaining causal logic through CRE and DES introduces computational overhead, which may limit the framework's scalability and applicability in real-world scenarios with limited resources. Finally, our approach pays insufficient attention to the sequential ordering of steps, and the ATE can only determine whether the causal logic makes sense, rather than recognizing, for example, the assumption encountered in the Hanoi Tower problem that the larger disk must be placed under the smaller disk.

5 Conclusion

By integrating CRE and DES, the CreDes framework has significantly advanced LLMs' capabilities in long-range reasoning tasks. This combined approach enhances the accuracy and efficiency of multi-step reasoning and maintains the problem-solving and reasoning abilities of pre-trained models across different tasks. Future work will focus on refining the framework to improve scalability and efficiency in various complex problem-solving scenarios.

References

[1] Kanishk Gandhi, Jan-Philipp Fränken, Tobias Gerstenberg, and Noah Goodman. Understanding social reasoning in language models with language models. *Advances in Neural Information Processing Systems*, 36, 2024.

[2] Hainiu Xu, Runcong Zhao, Lixing Zhu, Jinhua Du, and Yulan He. Opentom: A comprehensive benchmark for evaluating theory-of-mind reasoning capabilities of large language models. *arXiv preprint arXiv:2402.06044*, 2024.

[3] Wenyue Hua, Lizhou Fan, Lingyao Li, Kai Mei, Jianchao Ji, Yingqiang Ge, Libby Hemphill, and Yongfeng Zhang. War and peace (waragent): Large language model-based multi-agent simulation of world wars. *arXiv preprint arXiv:2311.17227*, 2023.

[4] Yang Li, Yangyang Yu, Haohang Li, Zhi Chen, and Khaldoun Khashanah. Tradinggpt: Multi-agent system with layered memory and distinct characters for enhanced financial trading performance. *arXiv preprint arXiv:2309.03736*, 2023.

[5] Qinlin Zhao, Jindong Wang, Yixuan Zhang, Yiqiao Jin, Kaijie Zhu, Hao Chen, and Xing Xie. Competeai: Understanding the competition behaviors in large language model-based agents. *arXiv preprint arXiv:2310.17512*, 2023.

[6] Tian Xia, Zhiwei He, Tong Ren, Yibo Miao, Zhuosheng Zhang, Yang Yang, and Rui Wang. Measuring bargaining abilities of llms: A benchmark and a buyer-enhancement method. *arXiv preprint arXiv:2402.15813*, 2024.

[7] Lin Xu, Zhiyuan Hu, Daquan Zhou, Hongyu Ren, Zhen Dong, Kurt Keutzer, See-Kiong Ng, and Jiashi Feng. Magic: Investigation of large language model powered multi-agent in cognition, adaptability, rationality and collaboration. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*, 2023.

[8] Shaoguang Mao, Yuzhe Cai, Yan Xia, Wenshan Wu, Xun Wang, Fengyi Wang, Tao Ge, and Furu Wei. Alympics: Language agents meet game theory. *arXiv preprint arXiv:2311.03220*, 2023.

[9] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Yan Xia, Man Lan, and Furu Wei. K-level reasoning with large language models. *arXiv preprint arXiv:2402.01521*, 2024.

[10] Gabriel Mukobi, Hannah Erlebach, Niklas Lauffer, Lewis Hammond, Alan Chan, and Jesse Clifton. Welfare diplomacy: Benchmarking language model cooperation. *arXiv preprint arXiv:2310.08901*, 2023.

[11] Chenghao Huang, Yanbo Cao, Yinlong Wen, Tao Zhou, and Yanru Zhang. Pokergpt: An end-to-end lightweight solver for multi-player texas hold'em via large language model. *arXiv preprint arXiv:2401.06781*, 2024.

[12] Xiao Shao, Weifu Jiang, Fei Zuo, and Mengqing Liu. Swarmbrain: Embodied agent for real-time strategy game starcraft ii via large language models. *arXiv preprint arXiv:2401.17749*, 2024.

[13] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[14] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240):1–113, 2023.

[15] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf*, 2(3):8, 2023.

[16] Jean Kaddour, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, and Robert McHardy. Challenges and applications of large language models. *arXiv preprint arXiv:2307.10169*, 2023.

[17] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

[18] Shima Imani, Liang Du, and Harsh Shrivastava. Mathprompter: Mathematical reasoning using large language models. *arXiv preprint arXiv:2303.05398*, 2023.

[19] Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, et al. Solving quantitative reasoning problems with language models. *Advances in Neural Information Processing Systems*, 35:3843–3857, 2022.

[20] Hanmeng Liu, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Zhou, and Yue Zhang. Evaluating the logical reasoning ability of chatgpt and gpt-4. *arXiv preprint arXiv:2304.03439*, 2023.

[21] Fangzhi Xu, Qika Lin, Jiawei Han, Tianzhe Zhao, Jun Liu, and Erik Cambria. Are large language models really good logical reasoners? a comprehensive evaluation from deductive, inductive and abductive views. *arXiv preprint arXiv:2306.09841*, 2023.

[22] Zirui Zhao, Wee Sun Lee, and David Hsu. Large language models as commonsense knowledge for large-scale task planning. *Advances in Neural Information Processing Systems*, 36, 2024.

[23] SP Sharan, Francesco Pittaluga, Manmohan Chandraker, et al. Llm-assist: Enhancing closed-loop planning with language-based reasoning. *arXiv preprint arXiv:2401.00125*, 2023.

[24] Alexandros Xenos, Niki Maria Foteinopoulou, Ioanna Ntinou, Ioannis Patras, and Georgios Tzimiropoulos. Vllms provide better context for emotion understanding through common sense reasoning. *arXiv preprint arXiv:2404.07078*, 2024.

[25] Adam Tauman Kalai and Santosh S Vempala. Calibrated language models must hallucinate. *arXiv preprint arXiv:2311.14648*, 2023.

[26] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

[27] Huaixiu Steven Zheng, Swaroop Mishra, Xinyun Chen, Heng-Tze Cheng, Ed H Chi, Quoc V Le, and Denny Zhou. Take a step back: Evoking reasoning via abstraction in large language models. *arXiv preprint arXiv:2310.06117*, 2023.

[28] Guangsheng Bao, Hongbo Zhang, Linyi Yang, Cunxiang Wang, and Yue Zhang. Llms with chain-of-thought are non-causal reasoners. *arXiv preprint arXiv:2402.16048*, 2024.

[29] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36, 2024.

[30] Wenhu Chen, Xueguang Ma, Xinyi Wang, and William W Cohen. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. *arXiv preprint arXiv:2211.12588*, 2022.

[31] Zirui Zhao, Wee Sun Lee, and David Hsu. Large language models as commonsense knowledge for large-scale task planning. *Advances in Neural Information Processing Systems*, 36, 2024.

[32] John McCarthy et al. *Situations, actions, and causal laws*. Comtex Scientific, 1963.

[33] Tom Bylander. The computational complexity of propositional strips planning. *Artificial Intelligence*, 69(1-2):165–204, 1994.

[34] Eduardo F. Camacho and Carlos Bordons, editors. *Model predictive control*. Springer-Verlag, Berlin Heidelberg, 1999.

[35] Yu-qian Jiang, Shi-qi Zhang, Piyush Khandelwal, and Peter Stone. Task planning in robotics: an empirical comparison of pddl-and asp-based systems. *Frontiers of Information Technology & Electronic Engineering*, 20:363–373, 2019.

[36] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213, 2022.

[37] Zheng Chu, Jingchang Chen, Qianglong Chen, Weijiang Yu, Tao He, Haotian Wang, Weihua Peng, Ming Liu, Bing Qin, and Ting Liu. A survey of chain of thought reasoning: Advances, frontiers and future. *arXiv preprint arXiv:2309.15402*, 2023.

[38] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608*, 2022.

[39] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. Progprompt: Generating situated robot task plans using large language models. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11523–11530. IEEE, 2023.

[40] Yan Ding, Xiaohan Zhang, Chris Paxton, and Shiqi Zhang. Task and motion planning with large language models for object rearrangement. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2086–2092. IEEE, 2023.

[41] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

[42] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.

[43] Shibo Hao, Yi Gu, Haodi Ma, Joshua Jiahua Hong, Zhen Wang, Daisy Zhe Wang, and Zhiting Hu. Reasoning with language model is planning with world model. *arXiv preprint arXiv:2305.14992*, 2023.

[44] Karthik Valmeekam, Matthew Marquez, Sarath Sreedharan, and Subbarao Kambhampati. On the planning abilities of large language models-a critical investigation. *Advances in Neural Information Processing Systems*, 36, 2024.

[45] Xinyu Zhu, Junjie Wang, Lin Zhang, Yuxiang Zhang, Yongfeng Huang, Ruyi Gan, Jiaxing Zhang, and Yujiu Yang. Solving math word problems via cooperative reasoning induced language models. *arXiv preprint arXiv:2210.16257*, 2022.

[46] Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, et al. Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789*, 2023.

[47] Yining Ye, Xin Cong, Yujia Qin, Yankai Lin, Zhiyuan Liu, and Maosong Sun. Large language model as autonomous decision maker. *arXiv preprint arXiv:2308.12519*, 2023.

[48] Yao Yao, Zuchao Li, and Hai Zhao. Beyond chain-of-thought, effective graph-of-thought reasoning in large language models. *arXiv preprint arXiv:2305.16582*, 2023.

[49] David R Heise. *Causal analysis.* John Wiley & Sons, 1975.

[50] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.

[51] Amir Feder, Katherine A Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E Roberts, et al. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *Transactions of the Association for Computational Linguistics*, 10:1138–1158, 2022.

[52] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

[53] York Hagmayer, Steven A Sloman, David A Lagnado, and Michael R Waldmann. Causal reasoning through intervention. *Causal learning: Psychology, philosophy, and computation*, pages 86–100, 2007.

[54] Judea Pearl. *Causality*. Cambridge university press, 2009.

[55] Eric Nichols, Leo Gao, and Randy Gomez. Collaborative storytelling with large-scale neural language models. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pages 1–10, 2020.

[56] Honghua Zhang, Liunian Harold Li, Tao Meng, Kai-Wei Chang, and Guy Van den Broeck. On the paradox of learning to reason from data. *arXiv preprint arXiv:2205.11502*, 2022.

[57] Miles Turpin, Julian Michael, Ethan Perez, and Samuel Bowman. Language models don't always say what they think: unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems*, 36, 2024.

[58] Matej Zečević, Moritz Willig, Devendra Singh Dhami, and Kristian Kersting. Causal parrots: Large language models may talk causality but are not causal. *arXiv preprint arXiv:2308.13067*, 2023.

[59] Tamera Lanham, Anna Chen, Ansh Radhakrishnan, Benoit Steiner, Carson Denison, Danny Hernandez, Dustin Li, Esin Durmus, Evan Hubinger, Jackson Kernion, et al. Measuring faithfulness in chain-of-thought reasoning. *arXiv preprint arXiv:2307.13702*, 2023.

[60] Alethea Power, Yuri Burda, Harri Edwards, Igor Babuschkin, and Vedant Misra. Grokking: Generalization beyond overfitting on small algorithmic datasets. *arXiv preprint arXiv:2201.02177*, 2022.

[61] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[62] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

[63] Zhen Yang, Ming Ding, Qingsong Lv, Zhihuan Jiang, Zehai He, Yuyi Guo, Jinfeng Bai, and Jie Tang. Gpt can solve mathematical problems without a calculator. *arXiv preprint arXiv:2309.03241*, 2023.

[64] Robert J. Leonard. Theory of games and economic behavior. *History of Political Economy*, page 189–190, Mar 2006.

[65] Kevin Leyton-Brown and Yoav Shoham. Essentials of game theory: A concise multidisciplinary introduction. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, page 1–88, Jan 2008.

[66] Stuart Russell and Peter Norvig. Artificial intelligence: A modern approach. *Choice Reviews Online*, pages 33–1577–33–1577, Nov 1995.

[67] Glenn Shafer. The art of causal conjecture. Jan 1996.

[68] Judea Pearl and James M Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *UAI*, volume 95, pages 444–453. Citeseer, 1995.

[69] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[70] Colin Gerety and Paul Cull. Time complexity of the towers of hanoi problem. *ACM SIGACT News*, 18(1):80–87, 1986.

[71] Wang Ling, Dani Yogatama, Chris Dyer, and Phil Blunsom. Program induction by rationale generation: Learning to solve and explain algebraic word problems. *arXiv preprint arXiv:1705.04146*, 2017.

[72] Tushar Khot, Peter Clark, Michal Guerquin, Peter Jansen, and Ashish Sabharwal. Qasc: A dataset for question answering via sentence composition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8082–8090, 2020.

[73] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[74] Yuanzhi Li, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. Textbooks are all you need ii: phi-1.5 technical report. *arXiv preprint arXiv:2309.05463*, 2023.

[75] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023.

[76] Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. Mixtral of experts. *arXiv preprint arXiv:2401.04088*, 2024.

[77] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

[78] MN Team et al. Introducing mpt-7b: A new standard for open-source, commercially usable llms, 2023.

[79] Aiyuan Yang, Bin Xiao, Bingning Wang, Borong Zhang, Ce Bian, Chao Yin, Chenxu Lv, Da Pan, Dian Wang, Dong Yan, et al. Baichuan 2: Open large-scale language models. *arXiv preprint arXiv:2309.10305*, 2023.

[80] Wenyang Hui, Yan Wang, Kewei Tu, and Chengyue Jiang. Rot: Enhancing large language models with reflection on search trees. *arXiv preprint arXiv:2404.05449*, 2024.

A Appendix

A.1 Validation Results of Model's Inherent Capabilities

To verify the success rate of our CRE method on other baseline tasks, we designed a control experiment to ensure that our approach does not impair the model's inherent problem-solving and reasoning abilities. Since DES is specifically designed for Blocksworld, a task with longer reasoning steps, the control experiments listed do not involve such lengthy reasoning steps; therefore, DES's performance is not tested in this section. The experimental results indicate that the CRE method can, to some extent, enhance the model's problem-solving capabilities on other baseline tasks without causing any reduction in performance. See Table 4.

Table 4: Results of model's inherent capabilities

Model	AQUA	QASC
Llama-2-7B	0.25	0.17
Llama-2-7B + CRE	**0.74**	**0.62**
Baichuan-7B	0.31	0.07
Baichuan-7B + CRE	**0.85**	**0.31**
Mpt-7B	0.11	0.05
Mpt-7B + CRE	**0.65**	**0.27**
TAIDE-LX-7B	0.27	0.21
TAIDE-LX-7B + CRE	**0.89**	**0.72**
Qwen1.5-7B	0.57	0.09
Qwen1.5-7B + CRE	**0.75**	**0.37**

A.2 A Note on the Hanoi Tower Dataset

We generated and produced the Hanoi Tower dataset in the paper. The production method is to randomly generate several states conforming to the placement rules of the Hanoi Tower based on a given number of rods and disks, e.g., three rods and three disks, and randomly select one of these states as the starting and target states for a single sample. For a single sample, the classical partition algorithm is used to derive the pathway, and according to the length of the pathway, the sample is categorized into different number of steps groups, e.g., 3-steps, 5-steps, 7-steps, and so on. An odd number is chosen for the allocation because the most complex solving step of Hanoi Tower in the case of three rods and n disks is $2^n - 1$ steps. We generated the dataset Hanoi Tower using exactly the same storage format and Prompt structure as Blocksworld and GSM8K.

A.3 Prompt Templates Used During Training and Testing of CRE

Prompt 1 Prompt Templates Used During **Training**

1: **Input:** Initial State || Goal State #### Pathway
2: **Output:** #### Pathway
3: **Pathway:** <Step1><Step2><Step3><step4>

Prompt 2 Prompt Templates Used During **Testing**

1: **Input:** Initial State || Goal State
2: **Output:** #### Pathway
3: **Pathway:** <Step1><Step2><Step3><step4>

A.4 Full Experimental Results under The Blocksworld Dataset

Table 5: Succcess Rate under Blocksworld (Cont'd Table)

Model	2-step	4-step	6-step	8-step	10-step	12-step
Baichuan-7B + RAP	0.61	0.72	0.70	0.43	0.09	0.01
Baichuan-7B + CRE	**0.93**	**0.74**	**0.71**	0.25	0.05	0.00
Baichuan-7B + CreDes	-	-	-	**0.63**	**0.47**	**0.29**
Mpt-7B + RAP	0.25	0.06	0.00	0.00	0.00	0.00
Mpt-7B + CRE	0.32	0.11	0.04	0.00	0.00	0.00
Mpt-7B + CreDes	-	-	-	0.05	0.00	0.00
TAIDE-LX-7B + RAP	0.62	0.67	0.65	0.52	0.07	0.00
TAIDE-LX-7B + CRE	**0.99**	**0.89**	**0.81**	0.34	0.04	0.00
TAIDE-LX-7B + CreDes	-	-	-	**0.70**	**0.54**	**0.35**
Qwen1.5-7B + RAP	0.57	0.64	0.61	0.28	0.02	0.00
Qwen1.5-7B + CRE	**0.92**	**0.77**	**0.73**	0.34	0.08	0.02
Qwen1.5-7B + CreDes	-	-	-	**0.61**	**0.46**	**0.36**

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer:[NA]

Justification: This thesis is not concerned with theoretical research, theoretical assumptions and proof.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification:

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The source code and homemade Hanoi Tower dataset are not available during the review period for our paper, and the other datasets and pre-training models used are open-source acquired versions. The code is expected to be finalized and open-sourced after the review period.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification:

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: The experiments in this paper did not involve error analysis and statistical significance. The data obtained from the experiments in this paper were averaged under independently repeated experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: Our paper deals only with a generalized framework for solving a class of mathematical problems using LLMs, independent of societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification:

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: In this paper, we have produced our own new dataset, Hanoi Tower.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.