

Zero-Shot Multi-Hop Question Answering via Monte-Carlo Tree Search with Large Language Models

Anonymous ACL submission

Abstract

Recent advances in large language models (LLMs) have significantly impacted the domain of multi-hop question answering (MHQA), where systems are required to aggregate information and infer answers from disparate pieces of text. However, the autoregressive nature of LLMs inherently poses a challenge as errors may accumulate if mistakes are made in the intermediate reasoning steps. This paper introduces Monte-Carlo tree search for Zero-shot multi-hop Question Answering (MZQA), a framework based on Monte-Carlo tree search (MCTS) to identify optimal reasoning paths in MHQA tasks, mitigating the error propagation from sequential reasoning processes. Unlike previous works, we propose a zero-shot prompting method, which relies solely on instructions without the support of hand-crafted few-shot examples that typically require domain expertise. We also introduce a behavioral cloning approach (MZQA-BC) trained on self-generated MCTS inference trajectories, achieving an over 10-fold increase in reasoning speed with bare compromise in performance. The efficacy of our method is validated on standard benchmarks such as HotpotQA, 2WikiMulti-hopQA, and MuSiQue, demonstrating that it outperforms existing frameworks.

1 Introduction

Reasoning, a fundamental cognitive process underlying human intelligence, has long been considered a challenging task in the field of artificial intelligence. Recent advancements with large language models (LLMs) have shown that not only do models demonstrate strong adaptability and sample efficiency in various natural language tasks (Brown et al., 2020; OpenAI, 2023), but also are capable of reasoning when grown to a certain scale and prompted appropriately (Wei et al., 2022; Kojima et al., 2022). However, many of these showcased tasks can be answered without a complex series of

thinking or simply leverage the internal knowledge present in model parameters such as in common sense reasoning or arithmetic.

Multi-hop question answering (MHQA) is a challenging yet critical task in that regard as it requires multiple steps of reasoning through selecting and combining pieces of knowledge to accurately answer a question. Current approaches to MHQA mostly take the retrieve-and-read paradigm (Zhu et al., 2021) that fetches relevant information from external sources on which answers are formulated. With LLMs, the chain-of-thought (CoT) (Wei et al., 2022) prompting, which explicitly lays out intermediate reasoning steps, has become a natural choice in dealing with MHQA (Trivedi et al., 2023).

However, the autoregressive nature of LLMs and CoT poses a significant challenge, as an error at an earlier stage of reasoning can propagate throughout the subsequent steps and adversely affect the outcome. Tree-structured prompting (Yao et al., 2023a; Hao et al., 2023) has shown promise in mitigating the limitations of LLMs and CoT, but it remains dependent on few-shot examples, making it sensitive to their composition and requiring long input sequences.

In this work, we adopt Monte-Carlo tree search (MCTS) that strikes a balance between exploring potentially valuable thoughts and exploiting already explored thoughts with a zero-shot prompting method. This allows for circumventing errors at the inception of reasoning and selecting an optimal reasoning path. However, this approach tends to be compute-intensive given the inherent complexity of MCTS reasoning and requires extensive interactions with LLMs. To mitigate this, we suggest fine-tuning LLM with behavioral cloning (BC), leveraging the self-generated trajectory of MCTS. This approach significantly accelerates the reasoning process, all the while minimally compromising the performance. Moreover, our approach does not necessitate demonstrative examples to the instruction-

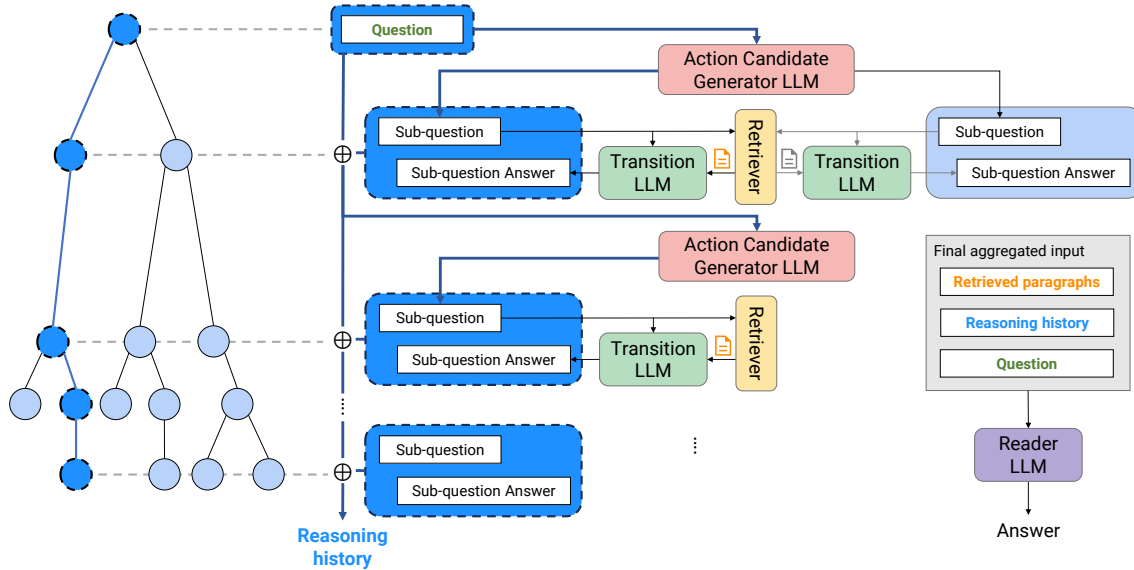


Figure 1: Overview of Monte-Carlo tree search for Zero-shot multi-hop Question Answering (MZQA).

tuned model, unlike previous approaches. By doing so, we solely rely on instructional prompts, thereby eliminating the need for hand-crafted examples that may require domain expertise.

Our contributions are summarized as follows: (1) We suggest MCTS with zero-shot prompting for MHQA (MZQA), an MCTS-based approach with LLMs for MHQA with a novel zero-shot prompting method that does not require human-generated few-shot examples. (2) We propose MZQA-BC, a method incorporating BC with reasoning trajectories generated by MCTS, to alleviate computation inefficiencies of MCTS reasoning. (3) We empirically show that our approaches outperform existing methods in standard MHQA benchmarks, effectively overcoming their limitations.

2 Related Work

2.1 LLMs for reasoning

Reasoning is a cognitive process that involves drawing conclusions and making decisions based on existing knowledge. With LLMs, several prompting techniques have been developed to elicit its reasoning capability. The most notable method is CoT (Wei et al., 2022), where LLM is prompted with few examples demonstrating a step-by-step reasoning process toward solving a task. Its zero-shot variant (Kojima et al., 2022) simply provides “Let’s think step-by-step” as a prompt without in-context examples.

While CoT elicits reasoning in LLM, it suffers performance degradation due to its sequential

nature; if a mistake is made in the intermediate steps, then it will negatively affect the subsequent steps. One of the ways to circumvent this is self-consistency (Wang et al., 2023), where the most consistent answer from multiple CoT samples is taken as the answer. However, this still struggles to solve complex multi-step reasoning tasks where each step is crucial to reasoning correctly.

Recently, tree-based methods such as tree-of-thought (ToT) (Yao et al., 2023a) and graph-based methods like graph-of-thought (GoT) (Besta et al., 2023) have also been proposed. In these methods, LLM generates and evaluates multiple proposals for the next reasoning step, taking the most promising ones greedily to form the reasoning path. Reasoning via Planning (RAP) (Hao et al., 2023) and Everything of Thoughts (XoT) (Ding et al., 2023) adopt MCTS for reasoning tasks, similar to our method MZQA. However, MZQA distinguishes itself from these two methods in several important aspects. First, they rely on few-shot prompts, in contrast to zero-shot prompts in MZQA, and exhibit a significant performance degradation when zero-shot prompts were given. Second, while RAP addresses simple reasoning tasks using only internal knowledge, MZQA is designed for more complex multi-hop QA tasks that require retrieving and integrating information from external knowledge bases. Third, unlike XoT, which utilizes MCTS as an external tool for refining LLM-generated thoughts, we utilize the LLM inside the MCTS process, generating actions in natural language. XoT is thus

customized for logical reasoning tasks where states and actions can be explicitly defined (i.e. finite sets), making it unsuitable for tasks involving a language action space, e.g. the MHQA task which is the main focus of our research.

2.2 MHQA with LLMs

The current methods for solving MHQA with LLM are dominated by CoT reasoning on the retrieved information. Interleaving Retrieval CoT (IRCoT) (Trivedi et al., 2023) proposes retrieving relevant documents at each step of CoT reasoning instead of retrieving only at a single step, helping retrieval with CoT and in turn improving reasoning with fetched documents. Similarly, ReAct (Yao et al., 2023b) integrates CoT reasoning and acting via API to enable LLMs to retrieve relevant information when needed. Multiple Chain Reasoning (MCR) (Yoran et al., 2023) combines multiple IRCoT samples to synthesize a better reasoning history. Other methods such as DecomP (Khot et al., 2023) and self-ask (Press et al., 2023) break down a multi-hop problem into single-hop problems to solve them sequentially. Due to the aforementioned limitations of CoT-based approaches, SearChain (Xu et al., 2024) proposes a DFS tree-based reasoning algorithm designed to reduce errors by evaluating each reasoning step and, if incorrect, reverting to the previous step to generate a new path. However, its greedy approach to pathfinding limits the possibility of finding the optimal path. LATS (Zhou et al., 2024) introduces a method leveraging external knowledge through MCTS. This approach incurs high computational costs, a typical challenge for tree-based searches. It also experiences performance variability, a problem arising from the use of prompting techniques in few-shot contexts, necessitating examples from domain experts.

3 Preliminaries

3.1 Multi-hop question answering

Multi-hop question answering (MHQA) aims to answer an intricate goal question, which is difficult to infer directly from a single paragraph. This problem necessitates multiple steps of accurate reasoning, or *thoughts*, to arrive at a precise answer. One approach to MHQA is breaking down the goal question into sub-questions, facilitating the retrieval of relevant information from various paragraphs. Assuming a sufficiently reliable retrieval system,

obtaining relevant information through querying sub-questions becomes crucial for accurately deducing the answer. Hence, it is important to generate sub-questions that gather useful information missing from current knowledge. Our objective is to generate appropriate sub-questions that are most helpful in generating the accurate answer to the goal question, given an information retriever.

To illustrate, consider an example with the goal question, “Which film has the director born first, *Gopalapuram* or *Here is the Beauty*?” (Figure 2). To answer this, the question can be divided into four sub-questions or ‘hops’. The first two hops involve identifying the directors of each movie, followed by determining their birth dates (the next two hops). We define a thought in MHQA reasoning as a pair of a sub-question and its answer, for example, Q: “Who was the director of *Gopalapuram*?” and A: “The director of *Gopalapuram* is K. K. Haridas”. This approach effectively decomposes a complex goal question into manageable sub-questions.

3.2 MDP formulation of MHQA

We formulate the MHQA problem as an MDP, which is defined as a tuple of $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \rho_0, r \rangle$ with state space \mathcal{S} , action space \mathcal{A} , transition model $\mathcal{T}(s_t | s_{t-1}, a_{t-1})$, initial state distribution $\rho(s_0)$, reward function $r(s, a, s')$.

State A state in our scenario is characterized by a sequence of sub-questions and their corresponding answers. Formally, an initial state $s_0 \in \mathcal{S}$, which consists of a pre-defined prompt and a goal question, is sampled from ρ_0 . At timestep t , $s_t \in \mathcal{S}$ is a sequence of pairs of questions a_{i-1} and their corresponding answers z_i for $i \in \{1, 2, \dots, t\}$, i.e. $s_t = \{s_0, (a_0, z_1), \dots, (a_{t-1}, z_t)\}$.

Action Each action $a \in \mathcal{A}$ is a sub-question that poses as a query to the retriever. To reduce the search space of the large language action space, we exploit LLM as an action candidate generator.

Transition Model A transition model leverages both the retriever and LLM to generate the next state. Specifically, the retriever extracts relevant paragraphs based on the sub-question a_{t-1} . Subsequently, LLM generates an appropriate answer z_t for the sub-question a_{t-1} with the corresponding retrieved paragraphs and the previous state s_{t-1} . Then, the current state is obtained as $s_t = s_{t-1} \cup \{(a_{t-1}, z_t)\}$.

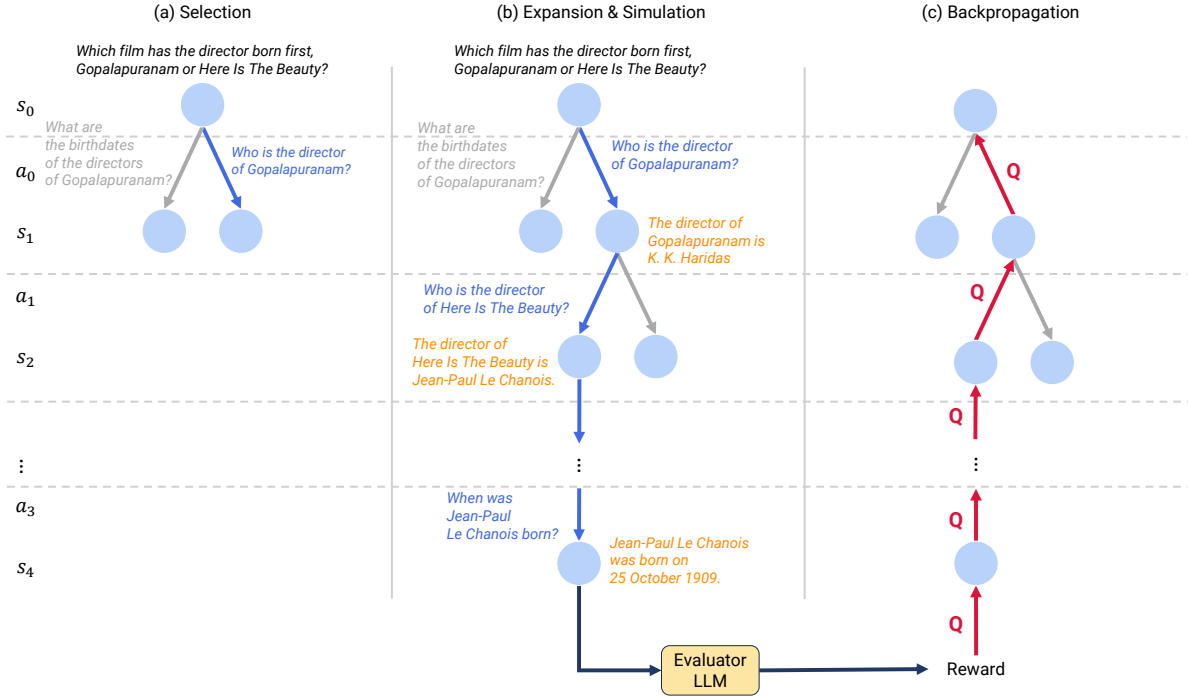


Figure 2: Step-by-step visualization of a single MCTS iteration for multi-hop question answering task.

Reward Function We aim to obtain a successful result without depending on human-generated QA task examples. To provide reward signals without these examples, we leverage the internal knowledge of LLM to evaluate the usefulness of the trajectory. In particular, at the end of each episode, we query LLM to assess the usefulness of the information acquired by the agent throughout a trajectory in addressing the goal question.

In this specific MDP, we aim to find an optimal policy $\pi^* : \mathcal{S} \rightarrow \Delta\mathcal{A}$ that maximizes cumulative rewards, i.e. $\pi^* := \arg \max_{\pi} \mathbb{E}_{\pi}[\sum_{t=0}^{H-1} r(s_t, a_t, s_{t+1})]$, where \mathbb{E}_{π} denotes the expectation given that the agent follows policy π and transition model \mathcal{T} .

4 MCTS for Zero-shot MHQA

4.1 MHQA with MCTS

A major challenge in MHQA is that each hop in the QA process depends on the successful completion of the previous one. Errors in the initial reasoning stages often propagate, resulting in incorrect final answers. To mitigate this issue, it is imperative to explore potentially beneficial yet previously unexamined reasoning thoughts. Therefore, we propose the use of Monte Carlo Tree Search (MCTS) with LLMs based on zero-shot prompting to improve decision-making in MHQA.

With the formulation in Section 3.2, we employ

MCTS for a principled search that exploits options with high value while exploring promising alternatives. A single iteration of MCTS consists of four main steps (Figure 2), namely selection, expansion, simulation, and backpropagation.

Selection The selection phase finds the most promising sub-question a_t for expansion. Starting from the root node s_0 , the algorithm chooses a child node until a leaf node of the current tree is reached. We employ the commonly used Upper Confidence Bound for Trees (UCT) (Kocsis and Szepesvári, 2006) (Equation 1) as criteria to select the child node, which strikes a balance between exploitation of high-value nodes and exploration of less visited nodes. If the selected node is terminal, we skip the expansion and proceed directly to backpropagation. Denoting $N(s)$, $N(s, a)$ the number of visits to node s and (s, a) respectively, $A(s)$ a possible action set in a state s , ω exploration constant, and $Q(s, a)$ state-action-value of the node, an action is selected by the following rule:

$$\text{UCT}(s) = \arg \max_{a \in A(s)} Q(s, a) + \omega \sqrt{\frac{\log N(s)}{N(s, a)}} \quad (1)$$

Expansion After selecting the node, we expand its children. Similar to the proposal prompt in ToT (Yao et al., 2023a), we prompt the LLM to generate multiple next sub-questions and retrieve corresponding documents from an external corpus.

Then, the transition model outputs the answer to the sub-question using the retrieved paragraphs and the previous state.

Simulation From each of the expanded nodes, we simulate until the terminal node based on a simulation policy, which in our case is generating a single sub-question via LLM to overcome the large language action space. The reward r is given according to the result of the simulation.

Backpropagation We update the statistics of the nodes from the expanded node to the root node based on the result of the simulation.

After a predetermined number of iterations, the path with the largest total value is chosen as the final reasoning path.¹ Finally, we aggregate the retrieved paragraphs and MCTS reasoning history and generate the final answer to the goal question by prompting LLM (called *Reader*) with the final aggregated result. The overview of the framework is provided in Figure 1. We refer this as MZQA (Monte-Carlo tree search for Zero-shot multi-hop Question Answering).

4.2 BC with self-generated data

While MCTS enables the exploration of different reasoning paths, it requires multiple calls to LLM during the search process. This amounts to requiring more computing resources than left-to-right reasoning methods such as CoT. In the past literature (Silver et al., 2017; Jang et al., 2021), behavioral cloning (BC), a simple offline imitation learning method that trains the mapping from states to expert actions using given offline expert trajectories, has been considered to mitigate the cost of MCTS inference.

We apply BC to our setting by fine-tuning LLM with trajectories generated by MCTS regarding as expert trajectories. Since BC enables the model to mimic the optimal path identified by MCTS without the need for extensive reasoning computations, it can substantially accelerate the decision-making process.

Given initial states $\{s_0^{(j)}\}_{j=1}^m$, m trajectories with a maximum hop H are generated by MCTS $\mathcal{D} := \bigcup_{j=1}^m \{(s_t^{(j)}, a_t^{(j)})\}_{t=0}^{H-1}$, where $s_t^{(j)}, a_t^{(j)}$ are a state and an action at timestep t of j -th trajectory

¹Although MCTS is typically used as an online planning algorithm, we follow RAP (Hao et al., 2023) for the sake of efficiency.

respectively. Then, the objective is to minimize the difference between predicted and MCTS actions. This can be represented as follows:

$$\pi_{\text{BC}} := \arg \min_{\pi} \mathbb{E}_{(s,a) \sim \mathcal{D}} [L(\pi(s), a)] \quad (2)$$

where L is a supervised learning loss (e.g. cross-entropy loss) between the predicted action $\pi(s)$ and the MCTS action a . Since the method is a distilled version of MZQA that imitates the MCTS process by tracking the intermediate results, we call this MZQA-BC (MZQA with Behavioral Cloning).

5 Experiments

5.1 Setup

Datasets The experiments were conducted on three standard open-domain MHQA benchmarks: HotpotQA (Yang et al., 2018), 2WikiMulti-hopQA (Ho et al., 2020), and MuSiQue (Trivedi et al., 2022). For the hyperparameter search and generating MCTS trajectories, we employ 100 data points from the dataset used in IRCOT (Trivedi et al., 2023) as a held-out dataset for each benchmark. The evaluation involves additional 500 data points from the same dataset for each benchmark. All benchmarks utilize Wikipedia as the retrieval corpus, consisting of 5,233,329 documents for HotpotQA, 430,225 for 2WikiMultihopQA, and 139,416 for MuSiQue as implemented in IRCOT.

Models GPT-3.5 (gpt-3.5-turbo-1106) (Brockman et al., 2023) is the default backbone model of all methods unless otherwise stated (gpt-4-1106-preview). For MCTS, we set the number of iterations to 20 and the episode length (maximum tree depth) to 4, aligning with the maximum hop in our datasets. In our BC experiments, we used 100 data samples that were excluded from the evaluation set. We generate trajectories through MCTS by employing these samples, then utilize generated trajectories to fine-tune the model through BC.

Retriever We use the retriever based on BM25 (Robertson and Zaragoza, 2009), implemented using Elasticsearch² as utilized in the IRCOT (Trivedi et al., 2023).

Baselines In our evaluations, we explore three families of established methods to gauge the efficacy of our proposed model in the Multi-Hop Question Answering (MHQA) domain.

²<https://www.elastic.co/>

	HotpotQA		2WikiMHQA				MuSiQue									
	2-hop		2-hop		4-hop		total		2-hop		3-hop		4-hop		total	
	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM
IO	41.3	31.5	33.4	26.5	54.3	52.7	37.8	32.0	17.4	9.1	11.3	3.9	10.3	4.3	14.2	6.6
IO (GPT-4)	48.6	35.4	41.4	32.3	66.7	64.8	46.7	39.1	26.6	15.3	23.2	16.2	20.0	10.9	24.4	14.7
ToT-BFS	40.9	31.4	33.3	25.1	55.6	54.3	38.0	31.2	25.5	15.0	16.4	8.9	9.8	3.3	19.8	10.9
ToT-DFS	40.0	30.1	30.7	23.0	52.8	51.1	35.3	28.9	22.5	14.2	15.6	8.9	8.0	3.3	17.7	10.6
RAP	41.0	30.7	35.8	26.8	55.5	54.3	39.9	32.6	23.4	14.2	16.5	8.4	10.1	4.0	18.8	10.5
IO w/ Retrieval	50.6	38.9	41.8	32.7	64.4	63.2	46.5	39.1	25.0	14.1	24.7	14.3	13.6	5.4	22.8	12.5
IO w/ Retrieval (GPT-4)	59.5	48.1	46.8	37.5	66.8	64.5	51.0	43.1	29.2	19.2	27.6	16.9	22.1	11.3	27.4	17.0
IRCoT	54.1	42.3	55.4	42.1	72.6	70.5	59.2	48.1	37.9	26.4	23.2	14.5	16.5	6.9	29.4	19.1
Decomp	37.8	29.9	52.7	43.2	73.7	73.3	57.9	50.3	19.1	10.5	9.5	4.3	15.0	10.9	15.4	8.6
ToT-BFS w/ Retrieval	59.1	48.2	55.4	45.2	62.2	60.3	56.8	48.4	40.2	30.8	27.5	18.2	14.1	6.5	31.5	22.5
ToT-DFS w/ Retrieval	58.1	46.6	54.1	45.0	60.6	58.4	55.5	47.8	36.7	28.5	26.1	18.4	11.6	4.3	28.8	20.9
LATS	45.6	44.9	54.3	53.6	69.2	69.2	57.5	56.9	18.1	17.8	7.1	7.1	3.6	3.6	12.1	11.9
MZQA (Ours)	62.3	49.9	67.2	56.8	71.4	70.2	68.1	59.6	40.8	31.0	28.9	18.8	17.1	8.3	32.8	23.1
MZQA-BC (Ours)	59.6	47.0	67.4	57.2	75.1	73.3	69.0	60.6	40.8	31.0	24.4	16.9	14.2	6.1	30.9	22.1

Table 1: Average F1 and EM (Exact Match) score over 3 runs on MHQA benchmarks by the number of hops in the evaluation samples. The top section of the table displays results from baseline models utilizing only the model’s internal knowledge, whereas the middle section includes results from baselines that incorporate external knowledge. The bottom section presents the outcomes for the model fine-tuned via BC. For standard IO (input-output) prompting, we test both GPT-3.5 (gpt-3.5-turbo-1106) and GPT-4 (gpt-4-1106-preview) with and without retrieval. The baselines have been reproduced with GPT-3.5 and the retrieval algorithm used in our method for a fair comparison.

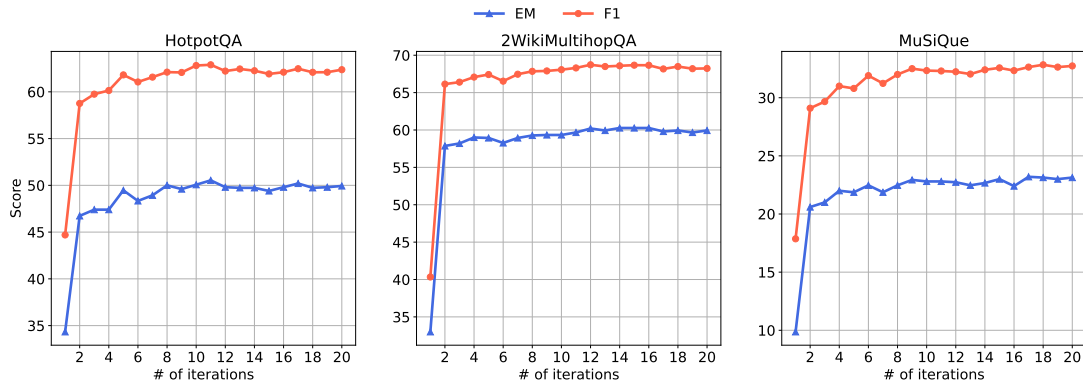


Figure 3: Performance progression of MZQA over iterations averaged across 3 different seeds.

- **Standard IO prompting:** These involve direct question-answering where the model generates responses based solely on the input, and we have tested this with both GPT-3.5 and GPT-4 to ensure alignment with others.
- **CoT-based methods:** These methods enhance the model’s ability to manage complex questions by leveraging intermediate reasoning steps, like IRCoT (Trivedi et al., 2023) and Decomp (Khot et al., 2023).
- **Tree-based methods:** These are hierarchical reasoning approaches such as ToT (Yao et al., 2023a), RAP (Hao et al., 2023), and LATS (Zhou et al., 2024), to provide sophisticated problem-solving capabilities.

For IRCoT and Decomp, we use 15 few-shot

examples, following the original implementations used to report their experimental results. For the other few-shot prompting baselines (e.g. ToT, RAP, LATS), we used 5 few-shot examples. Note that LATS can be seen as an extended version of RAP with an additional information retriever, making the experiment a fair comparison to all tree-based baseline methods with and without retriever settings. We reproduced all the baselines with GPT-3.5 using 3 seeds.

5.2 Performance evaluation

To evaluate the overall performance of the proposed and baseline methods on MHQA task, we measured the statistics of the F1 and exact match (EM) scores on the evaluation dataset from 3 separate runs. The results are summarized in Table 1.

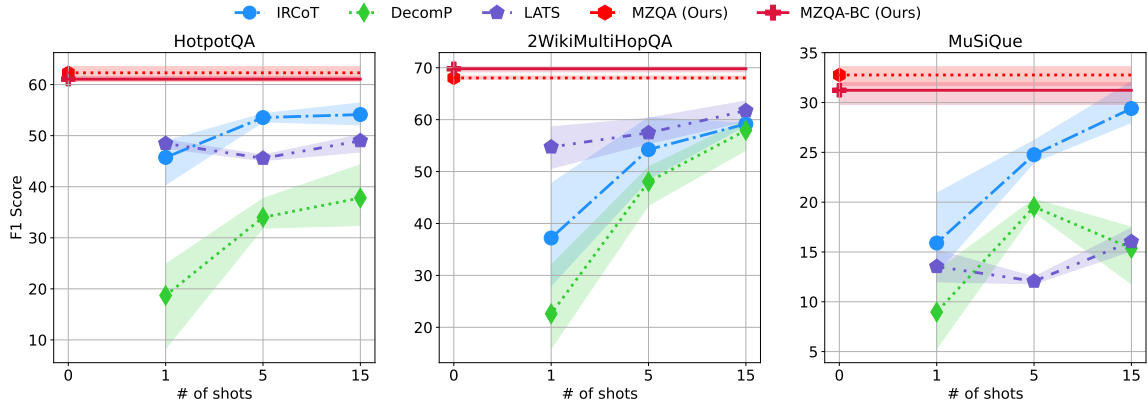


Figure 4: The F1 score with varying numbers of in-context examples (the shaded areas indicate min/max intervals).

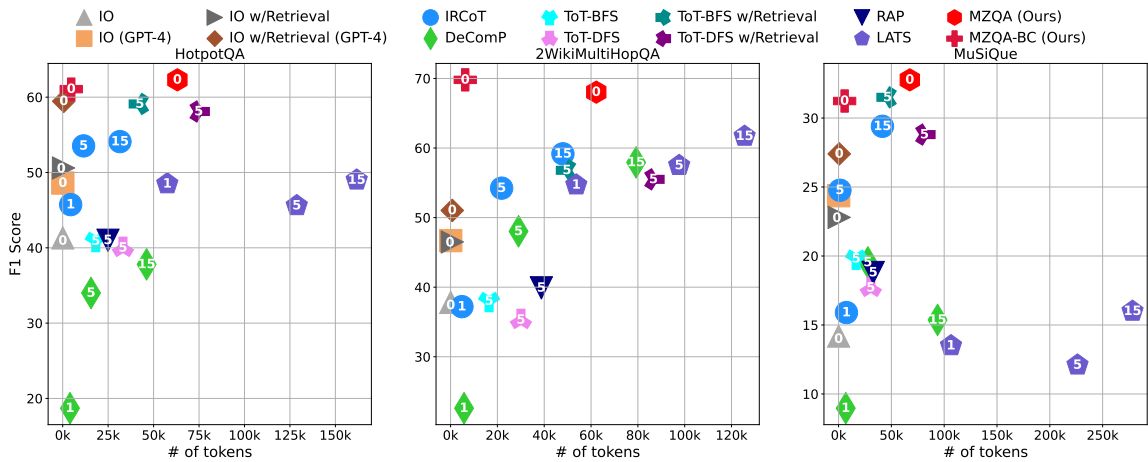


Figure 5: The compute-performance relationship between the average number of tokens required to generate the final answer to the goal question (x-axis) and the average F1 score (y-axis) over 3 seeds on each benchmark. The number written on each data point indicates the number of in-context learning examples. Note that the closer the point is to the upper-left corner, the more efficient the method is.

MZQA Our MCTS-based reasoning method has demonstrated superior F1 and EM scores compared to the baseline models across all total problem settings in the three MHQA benchmarks. Additionally, MZQA approach also outperforms in 7 out of 8 problem settings, with the exception being the 4-hop problem setting in MuSiQue, indicating that GPT-4 may contain internal knowledge of the answers corresponding to input goal questions.

This result testifies to the effectiveness of our approach in finding optimal reasoning paths through iterative trial and error. As depicted in Figure 3, MZQA improves in the performance with an increase in the number of iterations. This indicates that even if the initial iteration does not generate the correct sub-question, the performance is enhanced as MCTS iterates, evaluating nodes and selecting better states in the simulation process.

In Figure 4, we observe that the baseline methods, IRCoT, DecomP and LATS, exhibited significant variance in performance depending on the few-shot prompt, especially when the number of shots is scarce. In an extreme case, the IRCoT with 1-shot demonstrated a difference between the minimum and maximum F1 score on the 2WikiMultiHopQA dataset is 19.7 points. In contrast, the maximum difference in the minimum and maximum F1 scores for MZQA was at most 2.9 points in the worst-case benchmark, HotpotQA. This indicates that MZQA shows robust performance with lower variance across multiple runs, compared to baselines utilizing few-shot examples. This is particularly advantageous as it does not rely on few-shot prompts, making our approach more versatile in a wider range of scenarios.

439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455

	HotpotQA		2WikiMHQA		MuSiQue	
	F1	EM	F1	EM	F1	EM
ToT-BFS (w/ MZQA prompt)	60.2	48.0	64.7	56.5	32.9	23.5
ToT-DFS (w/ MZQA prompt)	60.1	47.7	61.5	53.3	31.7	21.9
MCTS (w/ MZQA prompt)	62.3	49.9	68.1	59.6	32.8	23.1

Table 2: Performance comparison of different tree-based search algorithms using MZQA prompts across multiple datasets. Each result is averaged over 3 seeds.

Even though tree search methods (e.g. RAP, ToT, LATS) typically require a significantly larger number of tokens to solve a problem, our zero-shot prompting method uses a comparable number of tokens to the CoT-based baseline methods that employ examples (Table 5). In summary, MZQA has demonstrated superior performance in terms of accuracy, consistency, and efficiency, making it a promising approach for tackling MHQA tasks.

MZQA-BC MZQA-BC, which is trained on the MCTS-generated trajectories to address computational inefficiencies, demonstrated a significant reduction in the number of LLM interactions, leading to faster inference (refer to Table 3). Furthermore, the number of tokens required to solve a problem was at least ten times less than that required by MZQA, making it comparable to the token usage of the CoT-based baseline methods in a 1-shot setting (refer to Figure 5).

The CoT-based baselines may also train a BC model to efficiently generate a more consistent answer, but this will require creating a dataset of gold thought trajectories by hand.

	HotpotQA	2WikiMHQA	MuSiQue
MZQA (Ours)	171.9	139.4	161.5
MZQA-BC (Ours)	13	13	13

Table 3: The average number of LLM interactions made when answering a goal question. (See Figure 6 in Appendix for details).

Interestingly, the performance of MZQA-BC was on par with that of MZQA. While there was a slight decrease in the F1 score by 2.7 points for HotpotQA and 1.9 points for MusiQue, an increase of 0.9 points for 2WikiMultihopQA was observed (refer to Table 1). This suggests that the BC model can effectively emulate the reasoning process of MZQA, providing a more efficient solution for MHQA tasks without compromising performance. In contrast to MZQA, although LATS also utilizes MCTS for reasoning with few-shot

prompts, the trajectories generated by LATS have empirically proven less effective for fine-tuning LLMs. (See Appendix. D.1) This highlights the efficacy of MZQA-generated trajectories for LLM fine-tuning. In addition, Figure 4 illustrates that MZQA-BC maintains robustness compared to baseline methods, as evidenced by that the difference between the minimum and maximum F1 scores was 2.9 in the worst benchmark, which was HotpotQA.

5.3 Ablation study

In this experiment, we aim to evaluate the impact of replacing the search algorithm in MZQA while consistently utilizing the MZQA prompting method. We assess the efficacy of BFS, DFS, as employed in ToT (Yao et al., 2023a), and MCTS with zero-shot prompting across three benchmarks. As indicated in Table 2, MCTS demonstrated superior performance in HotpotQA and 2WikiMHQA and similar performance in MuSiQue compared to ToT-BFS and ToT-DFS, highlighting its advanced strategic capabilities in structured environments. Due to the complex nature of MuSiQue’s questions, which pose a significant challenge even to human solvers, all three algorithms exhibited similar effectiveness.

6 Conclusion

We propose MZQA, an approach that combines instruction-based zero-shot prompting with MCTS reasoning to effectively address the challenges of MHQA. MZQA not only eliminates the need for domain experts to create few-shot examples, but also leverages the internal knowledge of LLM with MCTS to exploit the promising reasoning paths while encouraging exploring alternative reasoning paths. Furthermore, to enhance the efficiency of path reasoning computation, we introduce MZQA-BC, is a fine-tuning method with MCTS-generated trajectories. MZQA-BC mitigates the inefficiencies of MCTS inference, achieving significant improvements in reasoning speed without compromising the quality of the thoughts.

530 Limitations

531 This study has several key limitations. First, the
532 methods we developed are optimized for large lan-
533 guage models, such as GPT-3.5, and may not per-
534 form as well with smaller models. Second, our
535 tree-based approach requires building a new tree
536 for each question. Once a tree is built, it consumes
537 significant computational resources, only to be dis-
538 carded afterwards. This process can be both costly
539 and inefficient. Lastly, our fine-tuning process re-
540 lies heavily on the OpenAI platform, which does
541 not fully disclose how it operates. As a result, we
542 do not have a clear understanding of all the steps in-
543 volved in how our behavior cloning was fine-tuned.

544 Ethical Considerations

545 The large language models employed in our study,
546 such as those developed by OpenAI, are prone to
547 generating non-factual outputs due to hallucination
548 phenomena and may also produce biased outcomes,
549 despite the use of advanced techniques like tree
550 search influenced by their pre-training. Further-
551 more, the reliance on proprietary platforms raises
552 issues of transparency, necessitating a careful ap-
553 proach to the interpretation and dissemination of
554 our findings to thoroughly recognize these con-
555 straints. The datasets used in our research, includ-
556 ing HotpotQA, 2WikiMultihopQA, and MuSiQue,
557 are licensed under CC BY-SA 4.0³, Apache-2.0⁴,
558 and CC BY 4.0⁵, respectively.

559 References

560 Maciej Besta, Nils Blach, Ales Kubicek, Robert Ger-
561 stenberger, Lukas Gianinazzi, Joanna Gajda, Tomasz
562 Lehmann, Michal Podstawski, Hubert Niewiadomski,
563 Piotr Nyczyk, et al. 2023. Graph of thoughts: Solv-
564 ing elaborate problems with large language models.
565 *arXiv preprint arXiv:2308.09687*.

566 Greg Brockman, Atty Eleti, Elie Georges, Joanne Jang,
567 Logan Kilpatrick, Rachel Lim, Luke Miller, and
568 Michelle Pokrass. 2023. [\[link\]](#).

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

³<https://creativecommons.org/licenses/by-sa/4.0/>

⁴<https://www.apache.org/licenses/LICENSE-2.0>

⁵<https://creativecommons.org/licenses/by/4.0/>

Ruomeng Ding, Chaoyun Zhang, Lu Wang, Yong Xu,
Minghua Ma, Wei Zhang, Si Qin, Saravan Rajmohan,
Qingwei Lin, and Dongmei Zhang. 2023. *Everything
of thoughts: Defying the law of penrose triangle for
thought generation*. *Preprint*, arXiv:2311.04254. 575
576
577
578
579

Shibo Hao, Yi Gu, Haodi Ma, Joshua Hong, Zhen
Wang, Daisy Wang, and Zhiting Hu. 2023. *Reasoning
with language model is planning with world
model*. In *Proceedings of the 2023 Conference on
Empirical Methods in Natural Language Processing*,
pages 8154–8173, Singapore. Association for Com-
putational Linguistics. 580
581
582
583
584
585
586

Xanh Ho, Anh-Khoa Duong Nguyen, Saku Sugawara,
and Akiko Aizawa. 2020. *Constructing a multi-
hop QA dataset for comprehensive evaluation of
reasoning steps*. In *Proceedings of the 28th Inter-
national Conference on Computational Linguistics*,
pages 6609–6625, Barcelona, Spain (Online). Inter-
national Committee on Computational Linguistics. 587
588
589
590
591
592
593

Youngsoo Jang, Seokin Seo, Jongmin Lee, and Kee-
Eung Kim. 2021. *Monte-carlo planning and learning
with language action value estimates*. In *Internat-
ional Conference on Learning Representations*. 594
595
596
597

Tushar Khot, Harsh Trivedi, Matthew Finlayson, Yao
Fu, Kyle Richardson, Peter Clark, and Ashish Sab-
harwal. 2023. *Decomposed prompting: A modular
approach for solving complex tasks*. In *The Eleventh
International Conference on Learning Representa-
tions*. 598
599
600
601
602
603

Levente Kocsis and Csaba Szepesvári. 2006. Bandit
based monte-carlo planning. In *Machine Learning:
ECML 2006*, pages 282–293, Berlin, Heidelberg.
Springer Berlin Heidelberg. 604
605
606
607

Takeshi Kojima, Shixiang (Shane) Gu, Machel Reid, Yu-
taka Matsuo, and Yusuke Iwasawa. 2022. *Large lan-
guage models are zero-shot reasoners*. In *Advances in
Neural Information Processing Systems*, volume 35,
pages 22199–22213. Curran Associates, Inc. 608
609
610
611
612

OpenAI. 2023. *Gpt-4 technical report*. 613

Ofir Press, Muru Zhang, Sewon Min, Ludwig Schmidt,
Noah Smith, and Mike Lewis. 2023. *Measuring and
narrowing the compositionality gap in language mod-
els*. In *Findings of the Association for Computational
Linguistics: EMNLP 2023*, pages 5687–5711, Singa-
pore. Association for Computational Linguistics. 614
615
616
617
618
619

Stephen Robertson and Hugo Zaragoza. 2009. *The
probabilistic relevance framework: Bm25 and be-
yond*. *Found. Trends Inf. Retr.*, 3(4):333–389. 620
621
622

David Silver, Julian Schrittwieser, Karen Simonyan,
Ioannis Antonoglou, Aja Huang, Arthur Guez,
Thomas Hubert, Lucas Baker, Matthew Lai, Adrian
Bolton, et al. 2017. Mastering the game of go without
human knowledge. *nature*, 550(7676):354–359. 623
624
625
626
627

628	Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2022. MuSiQue: Multi-hop questions via single-hop question composition . <i>Transactions of the Association for Computational Linguistics</i> , 10:539–554.	684
629		685
630		686
631		687
632		
633	Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2023. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions . Association for Computational Linguistics.	688
634		689
635		690
636		691
637		692
638	Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models . In <i>The Eleventh International Conference on Learning Representations</i> .	
639		
640		
641		
642		
643		
644	Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models . In <i>Advances in Neural Information Processing Systems</i> , volume 35, pages 24824–24837. Curran Associates, Inc.	
645		
646		
647		
648		
649		
650		
651	Shicheng Xu, Liang Pang, Huawei Shen, Xueqi Cheng, and Tat-Seng Chua. 2024. Search-in-the-chain: Interactively enhancing large language models with search for knowledge-intensive tasks . In <i>Proceedings of the ACM on Web Conference 2024, WWW '24</i> , page 1362–1373, New York, NY, USA. Association for Computing Machinery.	
652		
653		
654		
655		
656		
657		
658	Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. 2018. HotpotQA: A dataset for diverse, explainable multi-hop question answering . In <i>Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing</i> , pages 2369–2380, Brussels, Belgium. Association for Computational Linguistics.	
659		
660		
661		
662		
663		
664		
665		
666	Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik R Narasimhan. 2023a. Tree of thoughts: Deliberate problem solving with large language models . In <i>Thirty-seventh Conference on Neural Information Processing Systems</i> .	
667		
668		
669		
670		
671		
672	Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. 2023b. React: Synergizing reasoning and acting in language models . In <i>The Eleventh International Conference on Learning Representations</i> .	
673		
674		
675		
676		
677	Ori Yoran, Tomer Wolfson, Ben Bogin, Uri Katz, Daniel Deutch, and Jonathan Berant. 2023. Answering questions by meta-reasoning over multiple chains of thought . In <i>Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing</i> , pages 5942–5966, Singapore. Association for Computational Linguistics.	
678		
679		
680		
681		
682		
683		

Appendix		693
Contents		694
A Hyperparameter search	12	695
B Datasets	12	696
C Implementation details	12	697
D Additional BC Experiment	13	698
D.1 Behavioral cloning applied to other baselines	13	699
D.2 Robustness of MZQA-BC on datasets	13	700
E Qualitative Analysis	14	701
F Pseudocode	15	702
G Prompts	16	703
G.1 Action	16	704
G.2 Transition model	16	705
G.3 Evaluation	17	706
G.4 Reader	18	707
G.5 Simulation policy and BC action	19	708
H MCTS visualization	20	709
I Utilization of the AI assistant	24	710

A Hyperparameter search

The performance of open-domain QA can vary by the number of retrieved paragraphs. Hence, we conduct a hyperparameter search on a validation dataset of 100 samples for each combination of methods, benchmarks, and if applicable the number of in-context learning examples with the options being {2, 4, 6, 8}. The optimal number of retrieved documents for each method and benchmark is in Table 4.

For all tree-based baseline models except LATS, the retrieval process utilized the same number of paragraphs as employed by MZQA. The configuration of paragraphs in LATS was designed to adhere to the implementation details specified in the original paper, accurately reflecting its structured prompting methodology.

	# of shots	HotpotQA	2WikiMHQA	MuSiQue
IO w/ Retrieval	-	4	6	4
IO (GPT-4) w/ Retrieval	-	6	6	4
IRCoT	1	2	2	4
	5	2	6	2
	15	6	4	6
DecomP	1	4	2	6
	5	6	4	4
	15	4	4	2
MZQA	-	6	8	6
MZQA-BC	-	6	8	6

Table 4: The optimal number of retrieved documents found by hyperparameter search.

B Datasets

	2 hop	3 hop	4 hop	corpus size
HotpotQA	500	-	-	5,233,329
2WikiMHQA	395	-	105	430,225
MuSiQue	254	154	92	139,416

Table 5: The number of 500 evaluation samples by hops for each dataset on 100 validation samples.

We evaluate the methods on three standard MHQA benchmarks. We use the identical 500 evaluation and 100 validation samples as in IRCoT (Trivedi et al., 2023), employing the same retrieval corpora for our experiments. The breakdown of these evaluation samples is provided in Table 5 along with the size of corpora from which documents are drawn.

C Implementation details

The generation parameters are fixed for all methods including the baselines. To ensure a more consistent generation, we set the temperature to 0.1 and top-p to 1.0. Strictly speaking, the transition model in our setting is inherently stochastic, as an answer to a sub-question can take many forms. Nevertheless, we assume that the model is generally deterministic since LLMs usually generate semantically consistent answers. For MZQA-BC, we use the fine-tuning API provided by OpenAI^a. The training hyperparameters are automatically searched by the API.

^a<https://platform.openai.com/docs/guides/fine-tuning>

We adopt several details from the baselines. In particular, we use ElasticSearch^b with the default search option for BM25-based retriever. As for the in-context learning examples, they are randomly sampled from 20 ready-made demonstrative examples. Hence, the examples may differ from seed to seed. For tree-based baselines, we standardized the number of action candidates to five, and set the breadth limit of ToT-BFS to one.

D Additional BC Experiment

D.1 Behavioral cloning applied to other baselines

	# of shots	HotpotQA		2WikiMHQA		MuSiQue	
		F1	EM	F1	EM	F1	EM
IRCoT	15	54.1	42.3	59.2	48.1	29.4	19.1
IRCoT-BC	15	55.3	42.5	61.1	50.3	31.3	21.7
LATS	15	45.6	44.9	57.5	56.9	12.1	11.9
LATS-BC	15	29.4	29.4	41.0	41.0	5.9	5.9
MZQA	0	62.3	49.9	68.1	59.6	32.8	23.1
MZQA-BC	0	59.6	47.0	69.0	60.6	30.9	22.1

Table 6: BC performance with trajectories generated by baselines and our method using 3 seeds.

In this experiment (Table 6), we aim to validate the suitability of trajectories produced by our proposed method and two baseline methods, IRCoT and LATS, for BC. As anticipated, BC utilizing trajectories generated by IRCoT and MZQA maintains performance levels comparable to those observed prior to fine-tuning. Conversely, we observe that BC with trajectories generated by LATS results in performance degradation relative to their original trajectory performance. We hypothesize that this degradation arises because LATS requires strictly structured responses for the given input prompts, thereby limiting the generalization capabilities of LLMs. These findings suggest that not all prompting methods ensure to generate trajectories that retain BC performance, as evidenced by LATS. However, our proposed method maintains performance levels similar to the original trajectory performance used for BC, indicating that MZQA-generated trajectories are more suitable for fine-tuning LLMs to address MHQA tasks.

D.2 Robustness of MZQA-BC on datasets

	HotpotQA		2WikiMHQA		MuSiQue	
	F1	EM	F1	EM	F1	EM
MZQA	62.3	49.9	68.1	59.6	32.8	23.1
MZQA-BC w/ HotpotQA	59.6	47.0	67.8	59.1	32.0	22.7
MZQA-BC w/ 2WikiMHQA	60.3	48.0	69.0	60.6	32.8	23.2
MZQA-BC w/ Musique	59.2	47.1	67.2	58.4	30.9	22.1

Table 7: Comparing BC performance across the datasets, averaged over 3 seeds.

To evaluate the generalization capability of the fine-tuned MZQA via BC, We conducted additional experiment (Table 7) to assess the performance of MZQA-BC on different datasets that were not utilized during BC training. In the table above, the second row shows the performance of MZQA-BC on three datasets, fine-tuned on 100 training instances from the HotpotQA dataset. The results show that MZQA-BC exhibits robust generalization performance over novel datasets.

^b<https://www.elastic.co>

E Qualitative Analysis

hop	Gold Facts	IRCoT	Decomp	MZQA (Ours)
Q	Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!?			
1	Arisan! is a 2003 Indonesian film that has drawn more than 100,000 viewers.	The president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan! is Indonesian President Susilo Bambang Yudhoyono .	What is the name of the newly declared independent country? Brazil	Which country made Arisan? Indonesia
2	The Indonesia–Timor Leste Commission on Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005.	So the answer is: Susilo Bambang Yudhoyono .	Who is the president of Brazil? Jair Bolsonaro	What is the Commission of Truth and Friendship?? The Commission of Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005 ...
3	Democratic Republic of Timor ... President Francisco Guterres Prime Minister Mari Alkatiri ...	-	-	What is the name of the newly declared independent country? The newly declared independent country is East Timor .
4	-	-	-	Who is the president of East Timor? The president of East Timor is Francisco Guterres

Table 8: Comparative analysis of the responses generated by IRCoT, Decomp and MZQA on a MuSiQue dataset example. **Blue** words indicate keywords in generated sentences that match gold facts, while **red** words indicate keywords that are irrelevant to gold facts.

To further understand the behavior of our proposed framework, we conduct a qualitative analysis comparing our algorithm with the baseline models. We examine the reasoning paths taken by each algorithm to answer a 3-hop question from MuSiQue, “*Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!?*”

To answer this example question accurately, the necessary information includes (1) the country that made *Arisan!* (*Indonesia*), (2) the country that participated in the Commission of Truth and Friendship with *Indonesia* (*East Timor*), and (3) the president of *East Timor* (*Francisco Guterres*). These details are provided in the gold facts column in Table 8.

The table highlights that IRCoT, which bases its retrieval and reasoning on history without generating sub-questions for answering the goal question, tends to prematurely conclude the final answer upon encountering relevant keywords in its historical data. Furthermore, Decomp, lacking a mechanism to assess the quality of its generated sub-questions, fails to create relevant sub-questions, leading to incorrect conclusions for the goal question.

In contrast, MZQA produces multiple promising sub-questions using the action candidate generator LLM to reach the correct answer. Additionally, by evaluating reasoning paths using the evaluator LLM, MZQA can leverage the reward signals to select an optimal path from an extensive set of reasoning paths. Taking these advantages, MZQA overcomes limitations of CoT-based baseline methods and hence finally generates successful a reasoning path, as observed in Table 8. This underscores the effectiveness of MZQA in handling complex multi-hop reasoning tasks. (See Figure 6 for a more detailed reasoning process of MZQA)

Algorithm 1 MZQA

Require: Initial state s_0 , exploration weight ω , instruction prompt \mathbf{p} , action candidates generator G , transition model \mathcal{T} , reward function R , value function Q , visitation number of the node N

```

1: procedure SEARCH( $s_0$ )
2:   while within computational budget do
3:      $S \leftarrow$  SELECT( $s_0$ )
4:     for  $s \in S$  do
5:        $r \leftarrow$  SIMULATE( $s$ )
6:       BACKPROPAGATION( $s, r$ )
7:     end for
8:   end while
9: end procedure
10: procedure SELECT( $s$ )
11:   while  $s$  is not a terminal state do
12:     if  $s$  is not expanded then
13:       return EXPAND( $s$ )
14:     else
15:        $a \leftarrow \arg \max_{a \in A(s)} \frac{Q(s,a)}{N(s,a)} + \omega \sqrt{\frac{\log N(s)}{N(s,a)}}$ 
16:        $s' \leftarrow \mathcal{T}(s, a)$ 
17:        $s \leftarrow s'$ 
18:     end if
19:   end while
20:   return  $\{s\}$ 
21: end procedure
22: procedure EXPAND( $s$ )
23:    $S := \{\}$ 
24:    $A \leftarrow G(\mathbf{p}_{\text{expansion}}, s)$ 
25:   for  $a \in A$  do
26:      $s' \leftarrow \mathcal{T}(s, a)$ 
27:      $S \leftarrow S \cup \{s'\}$ 
28:   end for
29:   return  $S$ 
30: end procedure
31: procedure SIMULATE( $s$ )
32:   while  $s$  is non-terminal do
33:      $a \sim G(\mathbf{p}_{\text{simulation}}, s)$ 
34:      $s' \leftarrow \mathcal{T}(s, a)$ 
35:      $s \leftarrow s'$ 
36:   end while
37:   return  $R(s)$ 
38: end procedure
39: procedure BACKPROPAGATION( $s, r$ )
40:   while do
41:      $N(s) \leftarrow N(s) + 1$ 
42:      $N(s, a) \leftarrow N(s, a) + 1$ 
43:      $Q(s, a) \leftarrow Q(s, a) + r$ 
44:     if  $s$  is  $s_0$  then
45:       break
46:     end if
47:      $s \leftarrow$  parent of  $s$ 
48:   end while
49: end procedure

```

777 **G Prompts**

778 Below are the zero-shot prompts that are used in each component of the proposed framework.

779 **G.1 Action**

```
Decompose it into the next 'Sub-questions' required to solve the 'Question'.
```

```
Consider
```

```
- Create "Sub-questions" with keywords from the 'Question' and 'Question history'.
```

```
Output format:
```

```
Sub-question: answer
```

780 **Example**

781 *Input*

```
Question: Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!?
```

782 *Output*

```
Sub-question: What is the name of the newly declared independent country?  
Sub-question: Who is the president of the newly declared independent country?  
Sub-question: What is the Commission of Truth and Friendship?  
Sub-question: Which country made Arisan?
```

783 **G.2 Transition model**

```
Answer the 'Sub-question' using the 'Paragraph', 'Evidence', and 'Question' as references.
```

```
Consider:
```

```
- Sentence: Find the sentence in the 'Paragraph' to answer the 'Sub-question'.  
If not found, return 'No information'.  
- Sub-question Result: Output result for the 'Sub-question'.
```

```
Output format as JSON:
```

```
{  
  "Sentence": result,  
  "Sub-question Result": result  
}
```

784 **Example**

785 *Input*

```
Paragraphs:
```

```
Title: Indonesia-Timor Leste Commission of Truth and Friendship  
The Indonesia-Timor Leste Commission on Truth and Friendship was a truth  
commission established jointly by the governments of Indonesia and East Timor  
in August 2005. The commission was officially created to investigate acts of  
violence that occurred around the independence referendum held in East Timor in  
1999 and sought to find the "conclusive truth" behind the events. ...
```

```
.  
. .
```

```
Title: Friendship Day
```

```
Friendship Day celebrations occur on different dates in different countries.
```


The first World Friendship Day was proposed for 30 July in 1958, by the World Friendship Crusade.
On 27 April 2011 the General Assembly of the United Nations declared 30 July as official International Friendship Day. However, some countries, including India, celebrate Friendship Day on the first Sunday of August. In Oberlin, Ohio, Friendship Day is celebrated on 8 April each year.

Evidence: Which country made Arisan? Indonesia
Question: Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!?
Sub-Question: What is the Commission of Truth and Friendship??

Output

786

```
{  
"Sentence": "The Indonesia-Timor Leste Commission on Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005.",  
"Sub-question Result": "The Commission of Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005 to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and sought to find the 'conclusive truth' behind the events."  
}
```

G.3 Evaluation

787

Assign a score from 0 to 10 to indicate how helpful 'Question history' is in solving the 'Question'.

Consider:

- Only assign a score of 10 if the 'Question history' can be utilized to answer the 'Question'.
- Assign a score of 0 if there is no information available to answer the 'Question' using the 'Question history'.

Output format:

Let's think Step by Step: Explain result
Score: answer

Example

788

Input

789

Question: Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!?

Question history:

Sub-question1: Which country made Arisan?

Sub-question Result1: Indonesia

Sub-question2: What is the Commission of Truth and Friendship??

Sub-question Result2: The Commission of Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005 to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and sought to find the 'conclusive truth' behind the events.

Sub-question3: What is the name of the newly declared independent country?

Sub-question Result3: The newly declared independent country is East Timor.

Sub-question4: Who is the president of East Timor?

Sub-question Result4: The president of East Timor is Francisco Guterres

790

Output

Let's think Step by Step: The question history provides all the necessary information to answer the question about the president of the newly declared independent country, East Timor, which is part of the Commission of Truth and Friendship with Indonesia. The president of East Timor is Francisco Guterres.

Score: 10

791

G.4 Reader

Answer 'Question' in words by referring the 'Paragraph' and 'Evidence'.

Consider:

- Keep your answer as short as possible, 10 words or less.
- Don't explain anything, just output the answer.

Output format

Answer: result

792

Example

793

Input

Paragraphs:

Title: Arisan!

Arisan! is a 2003 Indonesian film that has drawn more than 100,000 viewers. It is the first Indonesian film with a gay theme, and the first Indonesian film to use high-definition color enhancement. It uses a mixture of English, standard Indonesian and Jakartan slang. "Arisan!" became the second film in Indonesian film history to win all six major awards in Festival Film Indonesia (FFI), including Best Picture, Best Director, Best Actor, Best Actress, Best Supporting Actor, and Best Supporting Actress, after "Ibunda" in 1986. ...

...

Title: Indonesia-Timor Leste Commission of Truth and Friendship

The Indonesia-Timor Leste Commission on Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005.

The commission was officially created to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and sought to find the "conclusive truth" behind the events. ...

Evidence:

Question: Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!? Which country made Arisan? Indonesia

...

Who is the president of East Timor? The president of East Timor is Francisco Guterres

Question: Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!?

794

Output

Answer: Francisco Guterres

G.5 Simulation policy and BC action

795

Generate one 'Sub-question' to solve the 'Question'.

Consider

- Create a "Sub-question" with keywords from the 'Question' and 'Question history'.

Output format:

Sub-question: answer

Example

796

Input

797

Question: Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!?

Output

798

Sub-question: What is the newly declared independent country?

799

H MCTS visualization

800

The search tree from a completed MCTS is visualized in Figure 6. For the state associated with each node, we also provide the final sub-question, its corresponding answer, and the reward.

801

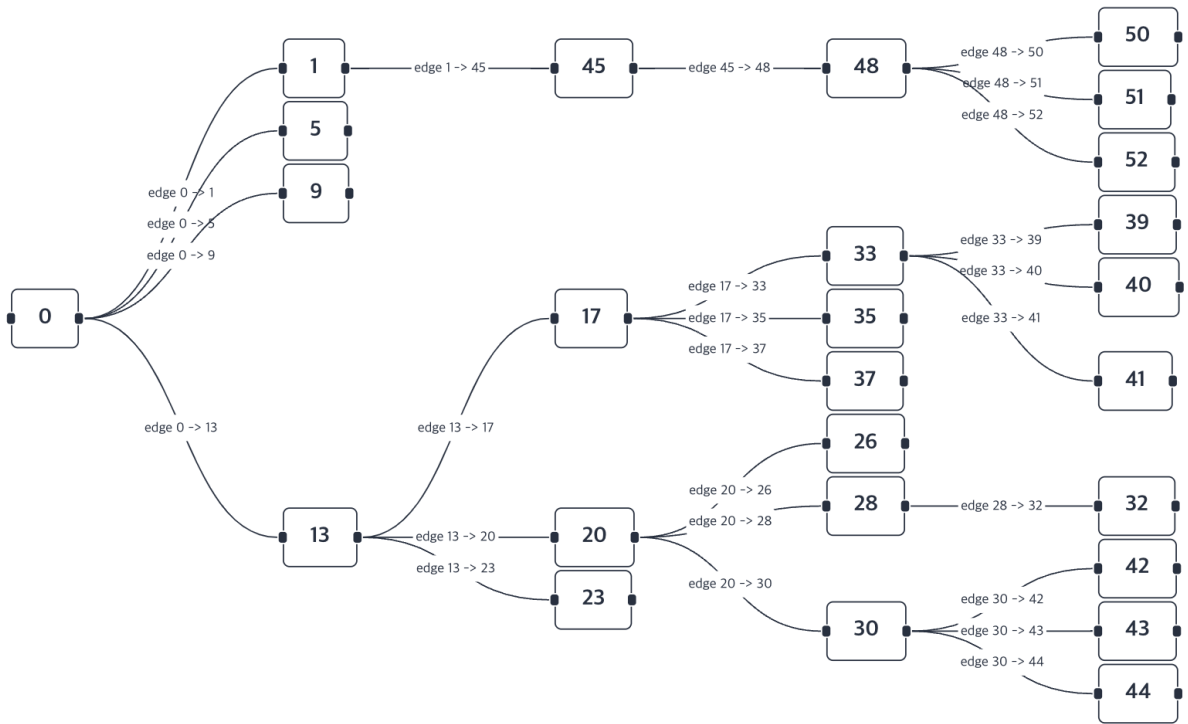


Figure 6: Visualization of the results from 20 iterations of MCTS applied to an instance of the MuSiQue dataset, where the numbers represent sequentially generated nodes. Note that some numbers are removed because those nodes were generated during the simulation process.

802

Node 0

Question: Who is the president of the newly declared independent country part of the Commission of Truth and Friendship with the country that made Arisan!??

803

Node 1

Sub-question: What is the name of the newly declared independent country?

Answer: The newly declared independent country is Guinea-Bissau.

Reward: 0

804

Node 5

Sub-question: Who is the president of the newly declared independent country?

Answer: The president of the newly declared independent country is not mentioned in the provided paragraph.

Reward: 0

Node 9

805

Sub-question: What is the Commission of Truth and Friendship?

Answer: The Commission of Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005 to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and sought to find the 'conclusive truth' behind the events.

Reward: 0

Node 13

806

Sub-question: Which country made Arisan?

Answer: Indonesia

Reward: 1

Node 17

807

Sub-question: Who is the president of the newly declared independent country?

Answer: The president of the newly declared independent country is not mentioned in the provided paragraph.

Reward: 0.5

Node 20

808

Sub-question: What is the Commission of Truth and Friendship?

Answer: The Commission of Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005 to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and sought to find the 'conclusive truth' behind the events.

Reward: 1

Node 23

809

Sub-question: What is the relationship between the newly declared independent country and the country that made Arisan?

Answer: The relationship between the newly declared independent country and the country that made Arisan is not mentioned in the provided text.

Reward: 0

Node 26

810

Sub-question: Who is the president of the newly declared independent country?

Answer: The president of the newly declared independent country is not mentioned in the provided text.

Reward: 0

811

Node 28

Sub-question: What is the name of the newly declared independent country?

Answer: The newly declared independent country is East Timor

Reward: 1

812

Node 30

Sub-question: What is the role of the Commission of Truth and Friendship in relation to the newly declared independent country?

Answer: The role of the Commission of Truth and Friendship in relation to the newly declared independent country is to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and seek to find the 'conclusive truth' behind the events.

Reward: 1

813

Node 32

Sub-question: Who is the president of East Timor?

Answer: The president of East Timor is Francisco Guterres.

Reward: 1

814

Node 33

Sub-question: What is the name of the newly declared independent country?

Answer: The name of the newly declared independent country is not mentioned in the provided paragraph.

Reward: 1

815

Node 35

Sub-question: What is the Commission of Truth and Friendship?

Answer: The Commission of Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005 to investigate acts of violence that occurred around the independence

referendum held in East Timor in 1999 and sought to find the 'conclusive truth' behind the events.

Reward: 0

816

Node 37

Sub-question: What is the significance of the country that made Arisan in this context?

Answer: The significance of the country that made Arisan in this context is that it is the first Indonesian film with a gay theme and the first Indonesian film to use high-definition color enhancement.

Reward: 0

Node 39

817

Sub-question: What is the name of the newly declared independent country?

Answer: The name of the newly declared independent country is not mentioned in the provided paragraph.

Reward: 0

Node 40

818

Sub-question: Who is the president of the newly declared independent country?

Answer: The president of the newly declared independent country is not mentioned in the provided paragraph.

Reward: 0

Node 41

819

Sub-question: What is the Commission of Truth and Friendship?

Answer: The Commission of Truth and Friendship was a truth commission established jointly by the governments of Indonesia and East Timor in August 2005 to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and sought to find the 'conclusive truth' behind the events.

Reward: 0

Node 42

820

Sub-question: Who is the president of the newly declared independent country?

Answer: No information

Reward: 0.2

Node 43

821

Sub-question: What is the name of the newly declared independent country?

Answer: The newly declared independent country is East Timor.

Reward: 0

Node 44

822

Sub-question: What is the relationship between the newly declared independent country and the Commission of Truth and Friendship?

Answer: The relationship between the newly declared independent country and the Commission of Truth and Friendship is that the commission was established to investigate acts of violence that occurred around the independence referendum held in East Timor in 1999 and sought to find the 'conclusive truth' behind the events.

Reward: 0.5

823

Node 45

Sub-question: Who is the president of Guinea-Bissau?

Answer: No information

Reward: 0.7

824

Node 48

Sub-question: Who made the declaration of independence for the newly declared independent country?

Answer: The declaration of independence for the newly declared independent country was made by the prime minister and president of the Democratic Republic of Vietnam (North Vietnam) since 1945.

Reward: 1

825

Node 50

Sub-question: What is the name of the Commission of Truth and Friendship with the country that made Arisan?

Answer: The name of the Commission of Truth and Friendship with the country that made Arisan is the Indonesia-Timor Leste Commission on Truth and Friendship.

Reward: 1

826

Node 51

Sub-question: What is the name of the country that made Arisan?

Answer: The country that made Arisan is Indonesia.

Reward: 0.1

827

Node 52

Sub-question: Who is the prime minister and president of the Democratic Republic of Vietnam (North Vietnam) since 1945?

Answer: The prime minister and president of the Democratic Republic of Vietnam (North Vietnam) since 1945 is Ho Chi Minh.

Reward: 0

828

I Utilization of the AI assistant

829

In this study, the AI language model ChatGPT, developed by OpenAI, was minimally utilized solely for adjusting the tone of our manuscript. Its application was confined to refining the language and enhancing readability without influencing the analytical or methodological aspects of our research.

830

831