Figure 1: Boosting of thoughts iteratively enhances the prompt by adding *experience*, which comprises the analysis conducted by large language models (LLM or LM) on the generated thought chain. The *experience* specifically contains the thought chain itself, the corresponding error reports, and detailed advice on revising each reasoning step. Thus, those ineffective thoughts marked with a red cross can also contribute to prompt refinement. By accumulating experiences over iterations in the prompt, BoT can eventually yield a correct thought chain starting from a simple prompt. The examples presented here are extracted from results obtained by applying GPT-4 with BoT on the `Game of 24` task.

for LLMs, BoT may get weak thoughts. With *aggregation*, BoT is capable of deriving a more logical and effective thought chain from them, thereby guiding the subsequent refinement. This guidance in our framework is achieved by tuning the prompt with *experience*, which is the detailed error reports, advice, and instructions of each reasoning step obtained by exploiting LLMs to analyze the aggregated chain. When such *experience* accumulates in the prompt, it gradually leads to stronger thoughts.

Specifically, BoT implements such a Boosting mechanism as an *experience*-driven iteration process, as shown in Fig. 1. In each iteration, for a given prompt, BoT builds massive simplistic thought structures in parallel with the LLM. We select the tree structure as in ToT Yao et al. (2024) but significantly modify it to weighted binary trees with various growth strategies for our boosting purposes. After extracting the root-to-leaf branch with the highest score per tree, the aggregation component of BoT is performed to aggregate them into one single thought chain. Subsequently, this chain is evaluated by the same LLM to gain the *experience*, which is added to the prompt as guidance for the thought generation in the next iteration.

Our contributions can be summarized in three folds. First, instead of generating more complicated structures for thoughts with well-designed prompts, this paper shows that it is possible to rely solely on a simple initial prompt, as weak thoughts can be refined progressively based on previous experience toward solving problems. Second, to achieve such a boosting mechanism, we propose Boosting of Thoughts (BoT), a novel framework that performs an *experience*-driven iterative process. Due to starting from a simple prompt, BoT is scalable across various tasks. While guaranteeing effectiveness, BoT is fast as it builds simplistic thought structures in parallel and converges to a solution after a few iterations. Finally, with GPT-4 and LlamaV2, we evaluate the performance of BoT on complex mathematical problems. Finally, relying on GPT-4 OpenAI (2023) and LlamaV2 Touvron et al. (2023), we evaluate the performance of BoT on complex mathematical problems. The problem-solving rates indicate that BoT, employing binary tree thought structures, significantly surpasses the current state-of-the-art on the `GSM8K` and `AQuA` while achieving the second-best results on other datasets. Especially on the new challenging task, `Game of 24` Yao et al. (2024), BoT is 9.7% higher than the leading approach ToT. Our BoT thus demonstrates that, through enhancing the prompt by accumulating error analysis of ineffective thought chains and the corresponding advice, even without human annotation, LLMs are scalable across various tasks while sustaining high performance.

2 RELATED WORK

Multi-Step Reasoning. The prominent work Chain-of-thought (CoT) prompting Wei et al. (2022) shows that step-by-step reasoning behaviors from LLMs can be elicited by providing intermediate reasoning steps, termed thoughts, within the prompt for each question, as also supported by Self-Consistency Wang et al. (2022) and a series of CoT-based work Zhou et al. (2023b); Fu et al. (2022). The recent work, Tree of Thoughts (ToT) Yao et al. (2024), converts the sequential reasoning process into a tree structure, in which each thought (node) may consider previous reasoning paths to produce multiple next-step thoughts. With such backtracking and expanded exploration during reasoning,