On the Diagram of Thought

Yifan Zhang[1] **Yang Yuan**[1,2] **Andrew Chi-Chih Yao**[1,2]
[1]IIIS, Tsinghua University
[2]Shanghai Qi Zhi Institute

Abstract

We introduce *Diagram of Thought* (DoT), a framework that models iterative reasoning in large language models (LLMs) as the construction of a directed acyclic graph (DAG) within a single model. Unlike conventional approaches that represent reasoning as linear chains or tree structures, DoT organizes propositions, critiques, refinements, and verifications into a unified DAG, enabling the exploration of complex reasoning pathways while preserving logical consistency. In this framework, each node encapsulates a proposition at various stages of evaluation, thereby facilitating iterative self-improvement through detailed natural language feedback. By leveraging auto-regressive next-token prediction augmented with role-specific tokens, DoT seamlessly transitions between generating ideas and engaging in critical evaluation, offering richer, context-aware feedback than binary signals. Moreover, we establish a rigorous mathematical foundation for DoT through Topos Theory, ensuring soundness and consistency in the reasoning process. This integrated approach not only simplifies both training and inference by eliminating the need for multiple models or external control mechanisms but also provides a principled framework for the design of next-generation reasoning-specialized models. The code is available at https://github.com/diagram-of-thought/diagram-of-thought.

1 Introduction

Large language models (LLMs) have achieved remarkable success across a wide array of tasks. Nonetheless, their capacity for intricate reasoning remains an open challenge. Conventional methods, such as Chain-of-Thought (CoT) (Wei et al., 2022), enable LLMs to generate intermediate reasoning steps in a linear fashion—effectively allowing the model to "think aloud." However, the linearity inherent in CoT can be insufficient to capture the dynamic, non-linear, and self-correcting nature of human reasoning, which often involves revisiting, critiquing, and refining initial propositions.

Recent extensions have sought to overcome these limitations. Tree-of-Thought (ToT) (Yao et al., 2023) introduces branching structures to explore alternative reasoning pathways, while Graph-of-Thought (GoT) (Besta et al., 2024) generalizes this concept by modeling reasoning as a graph. Cumulative Reasoning (CR) (Zhang et al., 2023) orchestrates an iterative process in which specialized LLMs—playing roles such as proposer, verifier, and reporter—collaborate to incrementally refine reasoning. Despite their merits, these approaches typically depend on external control mechanisms or the coordination of multiple models, thereby complicating both training and deployment.

In this work, we propose the *Diagram of Thought* (DoT) framework, which re-conceptualizes logical deduction as the construction of a directed acyclic graph (DAG) within a single LLM. In

DoT, each node of the DAG corresponds to a proposition that has been proposed, critiqued, refined, or verified. By incorporating natural language critiques rather than mere binary signals, DoT furnishes richer feedback that enables the model to pinpoint errors and effect more nuanced refinements.

Crucially, DoT leverages **auto-regressive** next-token prediction augmented with role-specific tokens (e.g., `<proposer>`, `<critic>`, `<summarizer>`). This design allows the model to seamlessly transition between generating novel ideas and engaging in critical evaluation—all within a unified framework. The resultant integration not only streamlines the overall architecture but also ensures compatibility with standard training paradigms, thereby facilitating both efficient deployment and enhanced interpretability.

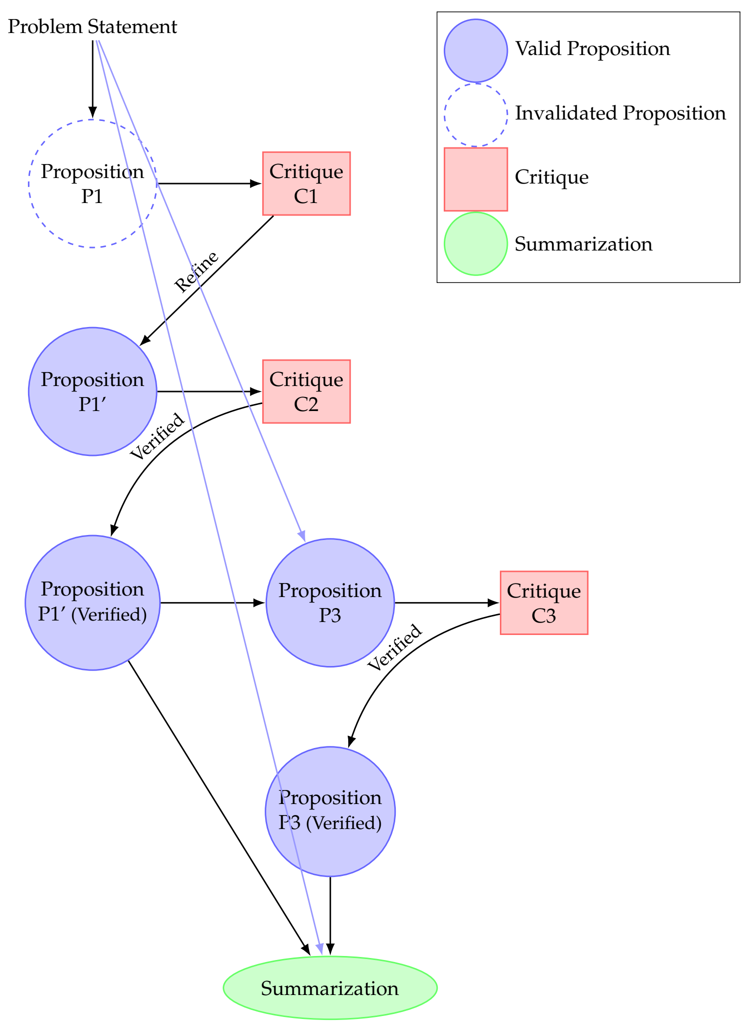

Figure 1: Diagram of Thought: Directed Acyclic Graph of Iterative Reasoning

Our main contributions are summarized as follows:

1. We introduce the Diagram of Thought (DoT) framework, which models iterative reasoning as the construction of a DAG within a single LLM. This approach systematically integrates propositions, critiques, refinements, and verifications into a cohesive structure.

2. We demonstrate that employing auto-regressive next-token prediction in conjunction with role-specific tokens allows the model to fluidly transition between generating ideas and critically evaluating them, thereby enriching the feedback mechanism without recourse to external control.
3. We formalize the DoT framework within Topos theory, thereby elucidating the connections between DoT and categorical logic while ensuring logical consistency and soundness throughout the reasoning process.

By embedding the entire reasoning process within a unified model, the DoT framework simplifies system design and streamlines deployment, while simultaneously enhancing interpretability and robustness. This integration paves the way for the development of next-generation reasoning systems that are both computationally efficient and theoretically well-grounded.

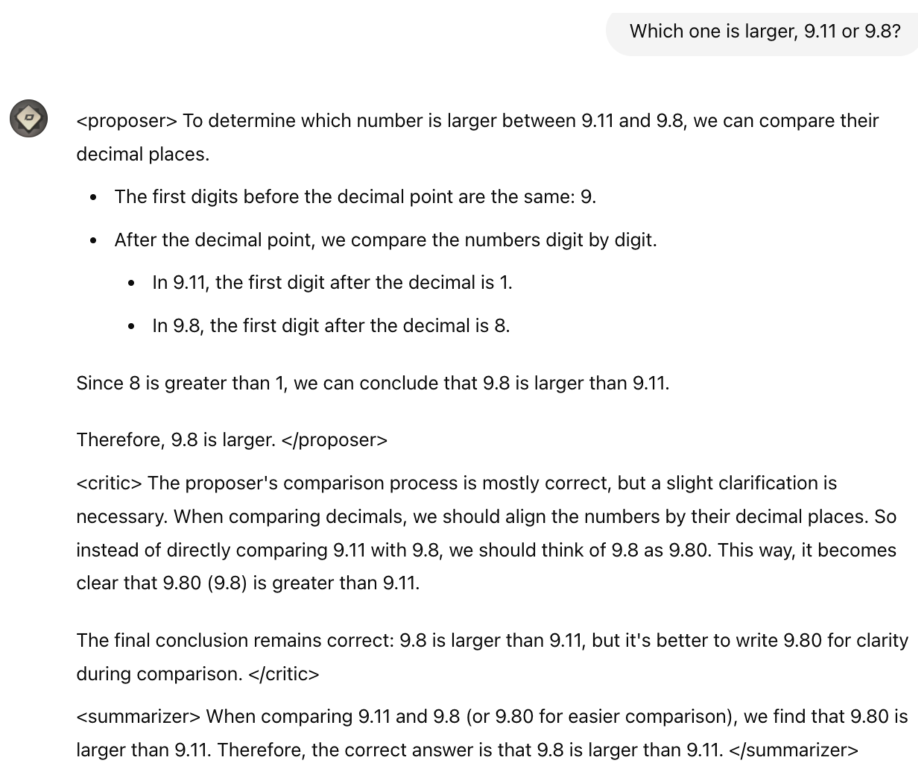

Figure 2: Asking LLM to answer the question: "Which one is larger, 9.11 or 9.8?".

2 Related Work

The evolution of intermediate reasoning paradigms in large language models (LLMs) has been marked by several influential approaches that explicitly articulate the reasoning process. Chain-of-Thought (CoT) (Wei et al., 2022) pioneered this effort by prompting LLMs to generate intermediate reasoning steps in a linear sequence. This "think-aloud" strategy has been empirically

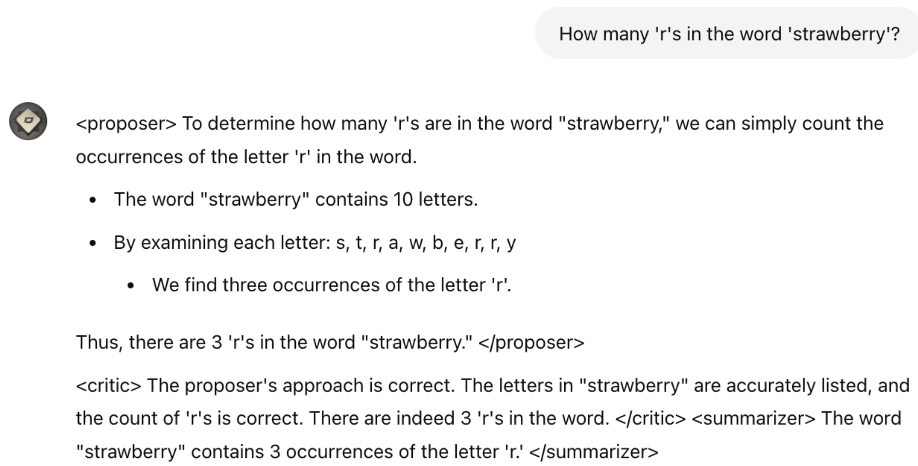

Figure 3: Asking LLM to answer the question: "How many 'r's in the word 'strawberry'?".

shown to improve performance on complex problem-solving tasks by making the hidden reasoning process explicit. Nevertheless, the linear structure of CoT is inherently limited in its capacity to capture the non-linear, iterative dynamics that characterize human reasoning.

To address these limitations, Tree-of-Thought (ToT) (Yao et al., 2023) extends the CoT framework by organizing reasoning processes into a tree structure. This branching architecture enables the exploration of multiple reasoning trajectories concurrently and facilitates backtracking, thereby accommodating alternative solutions and corrections. In a similar vein, Graph-of-Thought (GoT) (Besta et al., 2024) generalizes this idea further by representing the reasoning process as a graph. This representation is capable of capturing intricate interdependencies among reasoning steps, going beyond strictly hierarchical relationships. However, both ToT and GoT typically require the management of multiple reasoning pathways, which can introduce substantial computational overhead and complexity, particularly when integrating these approaches within a single LLM framework.

In contrast, Cumulative Reasoning (CR) (Zhang et al., 2023) adopts a multi-agent paradigm in which specialized LLMs fulfill distinct roles—namely, proposer, verifier, and reporter—to collaboratively refine reasoning chains. While CR effectively mirrors human collaborative problem-solving by iteratively building upon previous propositions and validations, the involvement of multiple models and external orchestration mechanisms further complicates both the training and deployment processes.

Diagram of Thought (DoT), as introduced in this paper, distinguishes itself by unifying the strengths of these prior approaches within a single, self-contained framework. By modeling reasoning as the construction of a directed acyclic graph (DAG), DoT is able to capture the non-linear and iterative facets of logical deduction while maintaining computational efficiency. The framework leverages auto-regressive next-token prediction to internally manage role transitions, thereby seamlessly alternating between the generation of propositions, the provision of critiques, and the synthesis of final outputs. This internal role management not only simplifies the reasoning process but also eliminates the need for external control or multi-model collaboration.

Furthermore, the DoT framework integrates iterative refinement and verification directly into the reasoning pipeline. By preserving all intermediate reasoning steps within a coherent DAG structure, DoT allows for continuous error correction and systematic validation. This approach enhances both the interpretability and robustness of the generated reasoning process, aligning well with established LLM training paradigms and offering a theoretically grounded alternative to linear or tree-based methods.

In summary, while CoT, ToT, GoT, and CR each contribute valuable insights into the modeling of intermediate reasoning, DoT advances the state-of-the-art by consolidating these insights into a unified, internally coherent framework. This consolidation not only streamlines the implementation and training of LLMs but also fosters enhanced logical consistency and robustness in iterative reasoning.

3 Diagram of Thought

In this section, we introduce the *Diagram of Thought* (DoT) framework, which models logical reasoning as the cumulative construction of a **directed acyclic graph (DAG)** within a single large language model (LLM). In DoT, each node of the DAG represents a distinct element of the reasoning process—namely, propositions, critiques, refinements, and verifications—while edges encode the logical dependencies and inferential relationships between these elements. The inherent acyclic structure guarantees a well-founded progression, thereby precluding circular dependencies and ensuring rigorous logical deduction.

3.1 Roles and Auto-Regressive Next-Token Prediction

The DoT framework exploits the LLM's auto-regressive next-token prediction to internally manage distinct reasoning roles through the use of role-specific tokens. These roles are defined as follows:

- **Proposer (`<proposer>`)**: Generates new propositions or intermediate reasoning steps, thereby contributing nodes to the DAG.
- **Critic (`<critic>`)**: Evaluates the propositions by identifying errors, inconsistencies, or logical fallacies, and appends corresponding critique nodes to the graph.
- **Summarizer (`<summarizer>`)**: Synthesizes the validated propositions by performing a topological sort on the DAG, ultimately yielding a coherent chain-of-thought.

Transitions between these roles are seamlessly integrated into the generation process, as the LLM leverages contextual cues provided by the role-specific tokens to predict subsequent tokens in a consistent and structured manner.

3.2 Iterative Reasoning Process

The iterative reasoning process in DoT unfolds through a systematic cycle that mirrors human problem-solving:

1. **Proposition Initiation**: The **Proposer** introduces an initial proposition, thereby creating a new node in the DAG.

2. **Critical Evaluation**: The **Critic** assesses the proposition. If flaws or inconsistencies are detected, it generates a critique node and establishes an edge from the original proposition to this critique.
3. **Refinement**: In response to the critique, the **Proposer** refines the initial proposition, producing a revised node that incorporates the corrective feedback.
4. **Iteration**: This cycle of proposing, critiquing, and refining continues until the proposition meets the criteria for logical validity and is formally verified.
5. **Synthesis**: Once a sufficient set of valid propositions is accumulated, the **Summarizer** performs a topological sort of the DAG, thereby synthesizing the overall reasoning into a coherent final output.

By engaging with both valid and invalid reasoning pathways, the DoT framework enables the model to learn iteratively from its own errors, akin to the hypothesis-testing cycle observed in human cognition.

Figure 1 provides a schematic illustration of the DoT framework. Each proposition, along with its refinements, is depicted as a node; critiques generate additional nodes without forming cycles, thus preserving the acyclic structure essential for coherent topological ordering.

3.3 Training and Inference Protocols

Training within the DoT paradigm involves curating examples that conform to the structured DoT format, incorporating explicit role-specific tokens and DAG representations. This exposure enables the LLM to learn the distinct functions associated with each role and to generate content that adheres to the underlying logical framework.

During inference, the LLM employs its auto-regressive next-token prediction to generate sequences that reflect the full spectrum of roles—proposing, critiquing, and summarizing—thereby constructing the reasoning DAG in a self-contained manner. This integrated approach not only streamlines the deployment process but also obviates the need for multi-model coordination or external control mechanisms, ensuring that the entire reasoning process remains both efficient and internally consistent.

4 Topos-Theoretic Formalization of DoT

In this section, we present a rigorous mathematical formalization of the Diagram of Thought (DoT) framework using Topos Theory, categorical logic, and PreNet categories. This formalism establishes a solid foundation for iterative reasoning by large language models, ensuring both logical consistency and the robust integration of inference, critique, and refinement processes.

4.1 Topos-Theoretic Foundations

A *topos* $\mathcal{E}$ is a category that, much like the category of sets, satisfies several key properties:

- $\mathcal{E}$ has all finite limits and colimits,
- $\mathcal{E}$ is *cartesian closed* (i.e., for every pair of objects $A, B \in \mathcal{E}$ there exists an exponential object B^A),

- $\mathcal{E}$ contains a *subobject classifier* Ω, which internalizes the notion of truth values (MacLane & Moerdijk, 2012; Lambek & Scott, 1988).

Within the internal language of $\mathcal{E}$, we model the elementary constituents of DoT as follows:

1. A *proposition* is represented as a subobject
$$P \hookrightarrow 1_{\mathcal{E}},$$
where $1_{\mathcal{E}}$ denotes the terminal object. This identifies P with the collection of conditions or states in which the proposition holds.
2. Logical inference between propositions is encoded via a morphism
$$f : P \to Q,$$
which, in the internal logic, asserts that P entails Q.
3. A *critique* is modeled as an evaluation morphism
$$c_P : P \to \Omega,$$
assigning to each element of P a truth value (or degree of validity) within Ω.
4. A *refinement* operation is formalized by a morphism
$$r : P \to P',$$
transforming an initial (possibly flawed) proposition P into a refined proposition P'.

This categorical representation embeds the DoT operations into the rich internal logic of the topos, allowing us to leverage additional structures—such as exponential objects and internal homs—for meta-level reasoning.

4.2 Iterative Reasoning as Diagrams and Colimits

The iterative nature of reasoning in DoT is naturally captured by a *diagram* in $\mathcal{E}$. Formally, we consider a functor
$$D : \mathcal{J} \to \mathcal{E},$$
where $\mathcal{J}$ is a small index category whose objects correspond to the nodes (propositions, critiques, refinements, etc.) and whose arrows represent the inferential or corrective relationships. The acyclicity of $\mathcal{J}$ guarantees that the reasoning process is well-founded and free of circular dependencies.

The *colimit* of the diagram,
$$\varinjlim D,$$
serves as a canonical aggregation of the entire reasoning process. In categorical terms, the colimit is characterized by a universal property: any cocone from D factors uniquely through $\varinjlim D$. This construction is analogous to the role of the **Summarizer** (`<summarizer>`) in DoT, which synthesizes the verified propositions into a unified chain-of-thought. The uniqueness (up to isomorphism) of the colimit ensures that the overall reasoning remains invariant under different presentations or orderings of the refinement process.

4.3 Dynamic Reasoning in PreNet Categories

To model concurrent and dynamic aspects of reasoning, we extend our framework to include *PreNet categories* (Baez et al., 2021). A PreNet category $\mathcal{C}$ generalizes the concept of Petri nets in a categorical setting:

- Objects in $\mathcal{C}$ represent *states* or propositions,
- Morphisms in $\mathcal{C}$ correspond to transitions (i.e., inference steps) between these states.

In this context, the DAG structure underlying DoT is modeled as a functor

$$D : \mathcal{J} \to \mathcal{C},$$

with the acyclic index category $\mathcal{J}$ reflecting both sequential and parallel reasoning processes. The colimit $\varinjlim D$ in $\mathcal{C}$ aggregates all the inference steps into a single, coherent state representing the final chain-of-thought. Importantly, the equivalence of the category of PreNets to a presheaf category ensures that $\mathcal{C}$ inherits the structure of a topos, thereby unifying the logical and dynamic aspects of DoT.

Proposition 4.1. The category of PreNets is equivalent to a presheaf category and thus inherits the structure of a topos (Baez et al., 2021). Consequently, the full power of topos-theoretic logic is available for modeling and verifying the DoT framework.

4.4 Soundness, Consistency, and Categorical Semantics

The topos-theoretic and PreNet formulations together ensure that the iterative reasoning process is both *sound* and *consistent*:

- The internal logic of a topos $\mathcal{E}$ provides a precise semantic interpretation of truth values, logical entailment, and inference.
- The acyclicity of the index category $\mathcal{J}$ ensures that the diagram D is well-founded, precluding any circular reasoning.
- The colimit $\varinjlim D$ aggregates all valid inferences, critiques, and refinements into a single object, guaranteeing that every logical contribution is coherently integrated and free from contradiction.

Thus, by identifying the final chain-of-thought with $\varinjlim D$, we achieve a canonical and invariant representation of the reasoning process that is robust under various decompositions and refinements.

4.5 Implications and an Illustrative Example

To illustrate the formalism, consider two propositions P and Q, represented as subobjects $P, Q \hookrightarrow 1_{\mathcal{E}}$, and assume an inference morphism

$$f : P \to Q.$$

Suppose a critique of P is expressed via a morphism

$$c_P : P \to \Omega,$$

signaling a deficiency in P. In response, a refinement morphism

$$r : P \to P'$$

produces an improved proposition P' and a corresponding inference

$$f' : P' \to Q.$$

The composite diagram

$$D : \quad P \xrightarrow{f} Q, \quad P \xrightarrow{c_P} \Omega, \quad P \xrightarrow{r} P' \xrightarrow{f'} Q,$$

captures both the original and refined reasoning processes. The colimit $\varinjlim D$ then unifies these steps into a single, coherent object in $\mathcal{E}$ that represents the final, verified chain-of-thought.

This rigorous categorical treatment of DoT not only validates the logical coherence of the reasoning process but also bridges abstract categorical theory with practical inference mechanisms in large language models. The framework thus provides a principled foundation for developing next-generation reasoning systems that are both mathematically sound and computationally effective.

5 Conclusion

In this paper, we have introduced the *Diagram of Thought* (DoT) framework, which re-conceptualizes iterative reasoning in large language models (LLMs) as the construction of a directed acyclic graph (DAG) within a single model. By integrating propositions, critiques, refinements, and verifications into a unified DAG, DoT captures the nuanced, non-linear aspects of logical deduction that extend beyond conventional linear or tree-based models. The framework exploits auto-regressive next-token prediction with role-specific tokens, thereby enabling seamless transitions between idea generation, critical evaluation, and synthesis—all without relying on external control mechanisms.

Moreover, we have provided a topos-theoretic formalization of the DoT framework, establishing a solid mathematical foundation by mapping reasoning elements to categorical structures. By representing propositions as subobjects, inferences as morphisms, and critiques as evaluative mappings within a topos, our approach guarantees logical consistency and soundness throughout the reasoning process. This formal underpinning not only substantiates the practical efficacy of DoT but also bridges the gap between abstract categorical logic and empirical performance in LLM reasoning tasks.

In summary, the DoT framework represents a substantial advancement in the design of reasoning-specialized models. By harmonizing auto-regressive prediction with a rigorous categorical foundation, our approach paves the way for the development of more capable and transparent reasoning systems, thereby contributing both to the theoretical understanding and practical realization of advanced language models.

References

John C Baez, Fabrizio Genovese, Jade Master, and Michael Shulman. Categories of nets. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pp. 1–13. IEEE, 2021.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17682–17690, 2024.

Joachim Lambek and Philip J Scott. *Introduction to higher-order categorical logic*, volume 7. Cambridge University Press, 1988.

Saunders MacLane and Ieke Moerdijk. *Sheaves in geometry and logic: A first introduction to topos theory*. Springer Science & Business Media, 2012.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*, 2022.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601*, 2023.

Yifan Zhang, Jingqin Yang, Yang Yuan, and Andrew Chi-Chih Yao. Cumulative reasoning with large language models. *arXiv preprint arXiv:2308.04371*, 2023.