



**HAL**  
open science

# Chain of thoughts, Tree of thoughts and Retrieval Augmented thoughts

Nabil Moncef Boukhatem, Davide Buscaldi, Leo Liberti

► **To cite this version:**

Nabil Moncef Boukhatem, Davide Buscaldi, Leo Liberti. Chain of thoughts, Tree of thoughts and Retrieval Augmented thoughts. LIX, École Polytechnique. 2024. hal-04748901

**HAL Id: hal-04748901**

**<https://hal.science/hal-04748901v1>**

Submitted on 22 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# Chain of thoughts, Tree of thoughts and Retrieval Augmented thoughts

Nabil Moncef Boukhatem<sup>1,3</sup>, Davide Buscaldi<sup>2</sup>, and Leo Liberti<sup>3</sup>

<sup>1</sup> LundiMatin, Paris, France [moncef-nabil.boukhatem@lundimatin.fr](mailto:moncef-nabil.boukhatem@lundimatin.fr)

<sup>2</sup> LIPN CNRS, Université de Paris-Nord, Villetaneuse, France,  
[buscaldi@lipn.univ-paris13.fr](mailto:buscaldi@lipn.univ-paris13.fr)

<sup>3</sup> LIX CNRS Ecole Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France, [liberti@lix.polytechnique.fr](mailto:liberti@lix.polytechnique.fr)

**Abstract.** This paper presents novel contributions to improving large language models (LLMs) through the integration and enhancement of three reasoning frameworks: Chain-of-Thoughts (CoT), Tree-of-Thoughts (ToT), and Retrieval-Augmented Thoughts (RAT). CoT generates linear intermediate reasoning steps, ToT introduces a branching structure for exploring multiple reasoning pathways, and RAT incorporates real-time external knowledge retrieval. A key innovation discussed is the use of decision trees, both handcrafted and automatically generated, to guide and validate the reasoning processes of CoT and ToT. This approach enhances the interpretability, accuracy, and adaptability of LLMs, particularly in complex problem-solving tasks like medical diagnostics and financial forecasting.

## 1 Introduction

While Chain-of-Thoughts (CoT) [3], Tree-of-Thoughts (ToT) [4], and Retrieval-Augmented-Thoughts (RAT) [2] share the common goal of enhancing the reasoning capabilities of LLMs, they each employ distinct strategies to achieve this, addressing different challenges in complex problem-solving.

## 2 Methodological Differences

Chain-of-Thoughts (CoT) [3] serves as the foundation among the three techniques. By generating intermediate reasoning steps in a linear sequence, CoT allows for a step-by-step elucidation of the model’s thought process. This method is not only instrumental in improving the transparency of the model’s decisions but also enhances accuracy, as each step can be scrutinized and adjusted independently before arriving at the final solution. The linear nature of CoT makes it particularly well-suited for tasks where reasoning is sequential and straightforward, such as arithmetic calculations or logical deductions.

Tree-of-Thoughts (ToT) [4], on the other hand, builds upon CoT by introducing a hierarchical and branching structure to the reasoning process. Unlike

CoT, which follows a single path from premise to conclusion, ToT allows the model to explore multiple potential reasoning pathways simultaneously. This tree-like structure enables the model to consider various alternative hypotheses or solutions before converging on the most optimal outcome. ToT is particularly advantageous in complex decision-making scenarios where multiple factors must be weighed, such as strategic planning or multi-step problem-solving tasks. The ability to explore different branches of reasoning enhances the robustness of the final decision by ensuring that all plausible alternatives have been considered.

Retrieval-Augmented-Thoughts (RAT) [2] introduces a different dimension to the reasoning process by incorporating external knowledge retrieval directly into the model’s cognitive workflow. While CoT and ToT rely primarily on the internal reasoning capabilities of the model, RAT extends this by allowing the model to access and integrate relevant external information in real time. This approach is particularly powerful in domains where up-to-date or highly specialized knowledge is crucial for accurate reasoning, such as in medical diagnostics or financial analysis. By augmenting the model’s internal reasoning with external data sources, RAT enables the generation of more contextually aware and informed outputs, significantly enhancing the model’s overall performance in dynamic and information-rich environments.

### 3 Applications and Implications

Each of these techniques has unique strengths that make them suitable for different types of tasks. CoT is particularly effective in scenarios that require clear, transparent reasoning with a focus on interpretability. Its linear approach ensures that each step in the reasoning process can be easily understood and validated, making it ideal for educational tools, tutoring systems, and any application where the clarity of the reasoning process is paramount, for example :

- **Mathematical Problem Solving:** Breaking down complex equations into sequential steps for better understanding and accuracy.
- **Logical Reasoning:** Solving logic puzzles or syllogisms by outlining each inferential step.
- **Instruction Generation:** Providing detailed procedural instructions for tasks like assembling equipment or following a recipe.
- **Education:** Assisting in tutoring systems where explaining each step enhances learning outcomes.

ToT’s ability to explore multiple reasoning paths simultaneously makes it highly valuable in more complex problem-solving tasks where multiple potential outcomes need to be evaluated. This includes applications in areas such as game theory, strategic decision-making, and any situation where exploring a broad range of possibilities is beneficial. The hierarchical structure of ToT allows for a more comprehensive exploration of the problem space, leading to more robust and reliable solutions in scenarios that benefit from exploring multiple possibilities, as in:

- **Strategic Planning:** Evaluating various strategies in business, military, or game scenarios to determine the optimal course of action.
- **Creative Writing:** Generating diverse story plots or brainstorming ideas by considering different narrative paths.
- **Problem Solving:** Tackling complex problems with multiple solutions by exploring alternative methods and comparing their outcomes.
- **Decision Support Systems:** Assisting decision-makers by presenting various options and their potential impacts.

RAT, with its focus on integrating external information, is particularly well-suited for tasks that require access to the most current or specialized knowledge. This makes RAT a powerful tool in fields such as medical diagnostics, where the latest research and data can significantly impact decision-making, or in financial forecasting, where real-time data integration is essential. By dynamically incorporating relevant external knowledge, RAT enables LLMs to provide more accurate and contextually relevant outputs, thereby extending the utility of LLMs in knowledge-intensive domains, as in:

- **Medical Diagnostics:** Incorporating the latest research findings into patient diagnosis and treatment recommendations.
- **Legal Analysis:** Accessing up-to-date legal precedents and statutes for case analysis.
- **Financial Forecasting:** Integrating real-time market data into economic models and investment strategies.
- **News Summarization:** Providing summaries that include the most recent developments on a topic.

## 4 Decision Trees as a Foundation for Chain of Thoughts and Tree of Thoughts

Decision trees are useful in structured decision-making tasks. The integration of decision trees, both handcrafted and automatically extracted, into the reasoning frameworks of ToT and ToT presents a promising avenue for further enhancing the interpretability, accuracy, and robustness of LLMs by enhancing reasoning through providing structured pathways. Decision trees [1], well-known for their clear, rule-based structure, can serve as a scaffold for generating and validating reasoning paths in CoT and ToT methodologies.

In the context of CoT reasoning, decision trees can structure the sequence of reasoning steps, providing predefined pathways that models follow to reach conclusions. Handcrafted decision trees, designed by domain experts, encapsulate best practices and critical decision points, ensuring that the model’s reasoning aligns with expert knowledge. For instance, in medical diagnostics, a decision tree could guide differential diagnosis by systematically mapping symptoms, test results, and patient history, enhancing both the transparency and trustworthiness of the model’s reasoning process. This approach not only improves

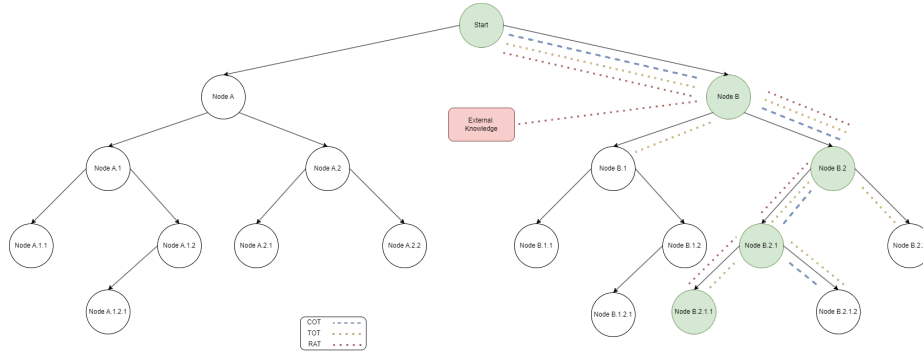


Fig. 1: An example of decision tree integration in CoT, ToT and RAT reasoning frameworks

interpretability but also increases accuracy, making the decision-making process more comprehensible.

Automatically extracted decision trees, on the other hand, offer scalability and adaptability by deriving reasoning pathways from large datasets. These trees capture patterns from historical data, allowing models to dynamically adjust to new information. Such flexibility is especially valuable in fast-evolving fields like financial forecasting, where decision trees can reflect the latest trends and evolving market conditions. By combining expert-driven and data-driven decision trees, CoT reasoning becomes both structured and adaptable, reducing ambiguity, enhancing transparency, and improving the model’s ability to learn from emerging data. This dual approach ensures that the reasoning remains aligned with best practices while remaining flexible enough to evolve with new information.

For ToT, decision trees can extend their utility by structuring the exploration of multiple reasoning paths. In this case, a decision tree could represent a hierarchical map of potential solutions, with each branch corresponding to a different reasoning path. The model can then explore these branches simultaneously, leveraging the tree’s structure to ensure that all relevant possibilities are considered. By guiding the exploration process, decision trees can help prevent the model from pursuing dead-end paths or missing critical alternative solutions, thus enhancing the robustness and completeness of the final decision.

## 5 Validation of Reasoning Paths through Decision Tree Metrics

Beyond their role in generating reasoning pathways, decision trees can also be employed to validate the thought processes generated by CoT, ToT, and RAT. One promising approach is to use a completion metric based on the decision tree structure, where the model’s reasoning is evaluated by how well it navigates through the tree following the indications provided in the prompt.

In practice, this validation process involves comparing the model’s reasoning path with the optimal path defined by the decision tree. For CoT, this could mean checking whether each step in the chain aligns with a corresponding node in the tree, ensuring that the model adheres to the established decision-making process. For ToT, the validation could involve evaluating how thoroughly the model explores the branches of the tree, ensuring that it does not prematurely converge on a single solution without considering viable alternatives.

The completion metric can be quantified in several ways. A basic approach is to measure the percentage of the decision tree that the model correctly navigates, with higher percentages indicating a more complete and accurate reasoning process:

$$CM = \frac{N_C}{N_T} \times 100 \quad (1)$$

Where  $CM$  is the Completion Metric (percentage),  $N_C$  is the number of correct nodes traversed by the model, and  $N_T$  is the total number of nodes in the optimal path.

A more nuanced approach involves weighting different parts of the tree based on their importance or the likelihood of leading to a correct solution:

$$WCM = \frac{\sum_{i=1}^n (w_i \times c_i)}{\sum_{i=1}^n w_i} \times 100 \quad (2)$$

Where  $WCM$  is the Weighted Completion Metric (percentage),  $w_i$  is the weight of node  $i$ , and  $c_i$  is 1 if node  $i$  is correctly traversed, 0 otherwise.

In addition to the completion metric, other metrics can be valuable when evaluating decision tree-based reasoning models. For instance, accuracy evaluation involves comparing the model’s reasoning path to the optimal path in the decision tree, helping identify deviations and errors in the model’s reasoning process:

$$AE = 1 - \frac{ED}{N_D + N_M} \quad (3)$$

Where  $AE$  is the Accuracy Evaluation (0 to 1),  $ED$  is the edit distance between the model’s path and the optimal path,  $N_D$  is the number of nodes in the optimal path, and  $N_M$  is the number of nodes in the model’s path.

Furthermore, performance benchmarks can be established by using decision trees to create standardized tests, allowing the evaluation of reasoning capabilities across different models:

$$PB = \frac{\sum_{i=1}^n (CM_i \times w_i)}{\sum_{i=1}^n w_i} \quad (4)$$

Where  $PB$  is the Performance Benchmark,  $CM_i$  is the Completion Metric for task  $i$ , and  $w_i$  is the weight of task  $i$  in the benchmark.

Feedback mechanisms can play a crucial role by offering corrective guidance when the model’s reasoning diverges from the expected path. The amount of feedback needed can be quantified as:

$$FM = \frac{N_D - N_C}{N_D} \quad (5)$$

Where  $FM$  is the Feedback Metric (0 to 1),  $N_D$  is the number of nodes in the optimal path, and  $N_C$  is the number of correct nodes traversed by the model.

In the context of RAT, decision trees can also play a crucial role in ensuring that the external knowledge retrieved by the model is appropriately integrated into the reasoning process. We can measure this integration using the RAT Integration Metric:

$$RIM = \frac{N_R}{N_E} \times \frac{N_C}{N_T} \quad (6)$$

Where  $RIM$  is the RAT Integration Metric (0 to 1),  $N_R$  is the number of retrieved knowledge points correctly applied,  $N_E$  is the total number of external knowledge points retrieved,  $N_C$  is the number of correct nodes traversed by the model, and  $N_T$  is the total number of nodes in the optimal path.

For ToT, we can evaluate how thoroughly the model explores the branches using the ToT Exploration Metric:

$$TEM = \frac{N_E}{N_P} \times \left( 1 - \frac{|D_O - D_A|}{D_O} \right) \quad (7)$$

Where  $TEM$  is the ToT Exploration Metric (0 to 1),  $N_E$  is the number of branches explored by the model,  $N_P$  is the total number of possible branches,  $D_O$  is the optimal depth of exploration, and  $D_A$  is the actual depth of exploration by the model.

These metrics, when combined with decision tree frameworks, offer comprehensive tools for enhancing and evaluating model performance. They provide a consistent framework for assessing model effectiveness in various domains, ensuring that reasoning quality is measured accurately and that the external data not only supports the reasoning process but also aligns with the established decision-making framework, thereby enhancing the coherence and reliability of the model’s outputs.

## 6 Hybrid Approaches: Combining Handcrafted and Automatically Extracted Decision Trees

The combination of handcrafted and automatically extracted decision trees represents a hybrid approach that could further enhance the effectiveness of CoT, ToT, and RAT. Handcrafted trees, with their expert-driven structure, provide a solid foundation for reasoning, ensuring that the model’s outputs are grounded in well-established knowledge and best practices. Automatically extracted trees,

with their data-driven flexibility, allow the model to adapt to new and emerging information, ensuring that the reasoning process remains relevant in dynamic environments.

In practice, a hybrid system could begin with a handcrafted decision tree as a base, guiding the model through the initial stages of reasoning. As the model encounters new data or complex scenarios, it could augment this base with automatically extracted trees, dynamically integrating new pathways that reflect the latest information. This approach would allow the model to benefit from the strengths of both handcrafted and automated reasoning frameworks, providing a balance between stability and adaptability.

Such a hybrid system could facilitate continuous learning, where the decision trees themselves are updated over time based on the model’s performance and feedback from human experts. This iterative refinement process could lead to increasingly sophisticated reasoning frameworks, enabling the model to handle even more complex and nuanced tasks with greater accuracy and confidence.

## 7 Challenges and Future Directions

Despite the advancements brought by CoT, ToT, and RAT, each approach also presents unique challenges. CoT, while effective in enhancing interpretability, may struggle with tasks that require considering multiple simultaneous hypotheses, as its linear nature restricts exploration to a single pathway. ToT, although more versatile in this regard, introduces complexity in managing and evaluating multiple reasoning paths, which can be computationally intensive and challenging to scale. RAT, while powerful in integrating external knowledge, relies heavily on the quality and relevance of the retrieved information, and there remains the challenge of ensuring that this integration is seamless and does not introduce noise or irrelevant data into the reasoning process.

Future research in this area could focus on hybrid approaches that combine the strengths of CoT, ToT, and RAT, potentially leading to even more powerful reasoning frameworks. For example, a system that starts with a CoT approach for initial reasoning, then expands into a ToT structure for exploring alternative solutions, and finally employs RAT to verify and augment the reasoning with external data, could offer a comprehensive solution that maximizes the strengths of all three techniques.

Additionally, improving the efficiency of these methods, particularly in the context of ToT and RAT, will be crucial for their broader application in real-time and resource-constrained environments. This may involve developing more sophisticated algorithms for path pruning in ToT or more efficient retrieval mechanisms in RAT, ensuring that these advanced reasoning techniques can be applied at scale without prohibitive computational costs.

Moreover, the integration of decision trees into CoT, ToT, and RAT opens up new avenues for research and application, particularly in fields where complex decision-making and dynamic information integration are critical. For instance, in healthcare, the ability to generate and validate reasoning paths based on



decision trees could lead to more accurate and reliable diagnostic tools, reducing the risk of misdiagnosis and improving patient outcomes. In finance, the use of decision trees could enhance the robustness of forecasting models, allowing for more informed investment decisions and risk management strategies.

Moreover, the development of completion metrics based on decision trees could lead to new benchmarks for evaluating the performance of LLMs in reasoning tasks. These metrics could be used to assess not only the accuracy of the model’s outputs but also the transparency and completeness of its reasoning process, providing a more comprehensive evaluation of the model’s capabilities.

In conclusion, CoT, ToT, and RAT represent significant advancements in the reasoning capabilities of LLMs, each offering unique benefits that address different aspects of complex problem-solving. By continuing to refine these techniques and explore their potential synergies, future AI systems could achieve even greater levels of performance, reliability, and applicability across a wide range of domains. In addition, the use of handcrafted and automatically extracted decision trees to generate and validate reasoning paths represents a significant advancement in the field of LLMs. By leveraging the structured, rule-based nature of decision trees, these approaches can enhance the interpretability, accuracy, and robustness of CoT, ToT, and RAT, pushing the boundaries of what AI systems can achieve in complex, real-world applications. As research in this area continues to evolve, the integration of decision trees into LLMs holds the potential to revolutionize how AI systems approach and solve complex reasoning tasks, leading to more effective and reliable outcomes across a wide range of domains.

## References

1. C. KINGSFORD AND S. SALZBERG, *What are decision trees?*, Nature biotechnology, 26 (2008), pp. 1011–1013.
2. Z. WANG, A. LIU, H. LIN, J. LI, X. MA, AND Y. LIANG, *Rat: Retrieval augmented thoughts elicit context-aware reasoning in long-horizon generation*, 2024.
3. J. WEI, X. WANG, D. SCHUURMANS, M. BOSMA, B. ICHTER, F. XIA, E. CHI, Q. LE, AND D. ZHOU, *Chain-of-thought prompting elicits reasoning in large language models*, 2023.
4. S. YAO, D. YU, J. ZHAO, I. SHAFRAN, T. GRIFFITHS, Y. CAO, AND K. NARASIMHAN, *Tree of thoughts: Deliberate problem solving with large language models*, Advances in Neural Information Processing Systems, 36 (2024).