Top-Down Cognitive Processing-Inspired Problem Decomposition for Agents

Sangbin Yun
Department of Brain and Cognitive Engineering
Korea University
Seoul, Republic of Korea, 02841
bin0113@korea.ac.kr

December 17, 2024

Abstract

This study explores a cognition-inspired approach to enhancing large language models (LLMs) by integrating top-down problem decomposition strategies. Inspired by human goal-directed behavior, the proposed framework breaks down complex tasks into structured sub-problems, solves them individually, and integrates the results into a final solution. Evaluated in a multi-step reasoning context, this method demonstrates improved accuracy and robustness compared to conventional approaches. The findings highlight the value of incorporating cognitive strategies into AI systems to enhance reasoning capabilities. Future work will focus on refining the framework, scaling it to diverse tasks, and integrating it into multi-agent systems for improved efficiency and performance.

1 Introduction

Top-down processing is a critical cognitive mechanism that underpins the goal-directed behavior of information-processing systems. It enables efficient decision-making by prioritizing task-relevant information while suppressing irrelevant details. In living organisms, this ability is essential for navigating complex environments, as it allows them to focus on significant stimuli and ignore distractions. The goal-directed behavior of living beings has long been a source of fascination and inspiration for cognitive scientists. This behavior highlights the brain's remarkable ability to modulate information processing to achieve specific objectives. At the core of this capability lies top-down processing, which facilitates selective attention, suppresses interference, and ensures efficient information management.

Motivated by these insights, this study explores whether the principles of top-down processing can be emulated in large language models (LLMs) to improve their reasoning and task execution capabilities. While LLMs have shown remarkable success in natural language tasks, they often struggle with multi-step reasoning and efficiently managing complex problems, particularly in environments with distracting or irrelevant contexts. To address this challenge, I hypothesized that incorporating top-down cognitive strategies into LLMs can facilitate goal-directedness by enhancing their ability to process information hierarchically. To test this hypothesis, I developed a LLM agent system specifically tailored for question-answering tasks. The system leverages LangGraph, a powerful platform for building and managing agent-based applications, to optimize multi-step reasoning processes.

This work represents a step toward integrating cognitive science principles into AI systems to improve their reasoning, robustness, and goal-directed behavior. By mimicking human-like cognitive strategies, this study aims to bridge the gap between biological cognition and artificial intelligence, paving the way for more efficient and trustworthy AI systems.

2 Related Work

Problem decomposition, also referred to as task decomposition, has recently garnered significant attention in the development of agent systems, particularly with the rapid advancements in LLMs. These models have demonstrated improved capabilities in zero-shot learning and in-context learning, enabling them to efficiently leverage problem

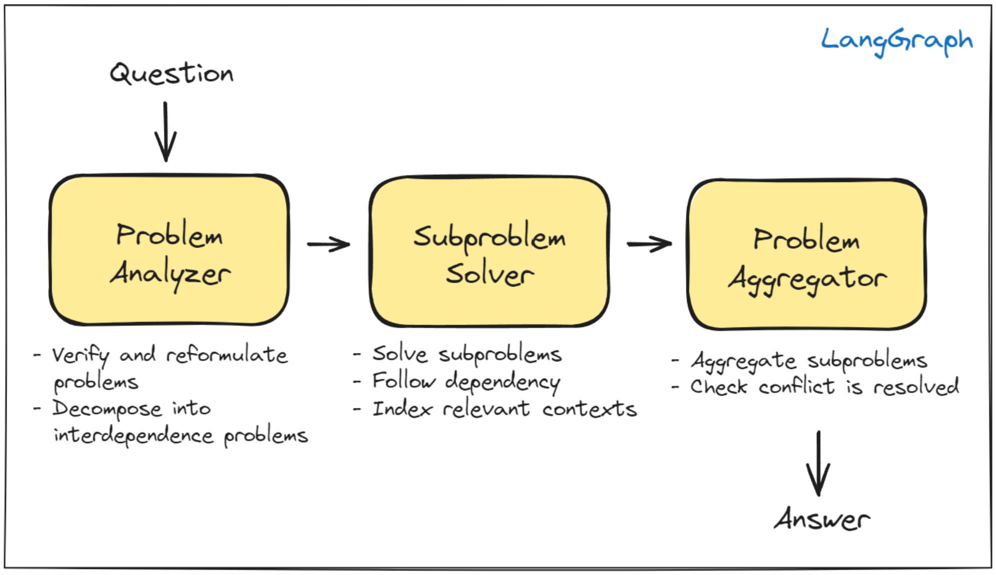

Figure 1: A visual representation of the problem decomposition process, which breaks down complex tasks into smaller, manageable sub-problems for efficient resolution and integration.

decomposition to enhance various agent heuristics. For instance, Huang et al. 2024 provide a comprehensive survey on the planning capabilities of LLM-based agents, highlighting how decomposition strategies benefit from the inherent learning mechanisms of LLMs [1]. Further advancements in prompt-based techniques have made problem decomposition more approachable and effective. Methods such as Chain-of-Thought (CoT) prompting have shown that structured reasoning steps can significantly improve LLM performance in complex reasoning tasks [2]. Similarly, zero-shot reasoning frameworks have emerged as powerful techniques for enabling reasoning without the need for task-specific examples [3]. Moreover, recent developments such as ReAct, which synergizes reasoning and acting, further demonstrate the growing importance of decomposition methods in enhancing LLM agent behavior [4].

Despite these advancements, current approaches often lack strong grounding in theories, which limits the systematic progression of problem decomposition frameworks. While techniques like CoT prompting and heuristic-driven decomposition yield practical improvements, a systematically established hypothesis doesn't exist. This gap highlights the need for a novel interpretation to advance a theoretically grounded approach, particularly problem decomposition by leveraging insights from cognitive science to enable more robust and goal-directed reasoning in LLM-based agent systems.

3 Method

To emulate the goal-directed problem-solving functionality inspired by top-down cognitive processing, I developed a problem decomposition process within a LangGraph-based LLM agent system. The system is designed to tackle complex problems by breaking them into smaller, manageable steps, thereby improving the LLM's reasoning capabilities and reducing cognitive load. The process consists of the following three key phases:

Problem Analyzer In this phase, the system accepts a complex question or task as input and decomposes it into smaller, structured sub-problems. This step ensures clarity and focus by isolating components of the original problem, thereby reducing ambiguity and providing a more manageable framework for solving. The reformulation phase mimics the hierarchical organization observed in human cognition, where individuals break tasks into smaller, goal-relevant steps.

Subproblem Solver Once the problem is reformulated, each sub-problem is addressed independently using the language model. This phase involves solving each sub-problem in isolation to reduce the cognitive load associated with solving the entire problem simultaneously. By resolving smaller tasks individually, the system can focus more effectively on the relevant details of each step without being overwhelmed by the complexity of the overall problem.

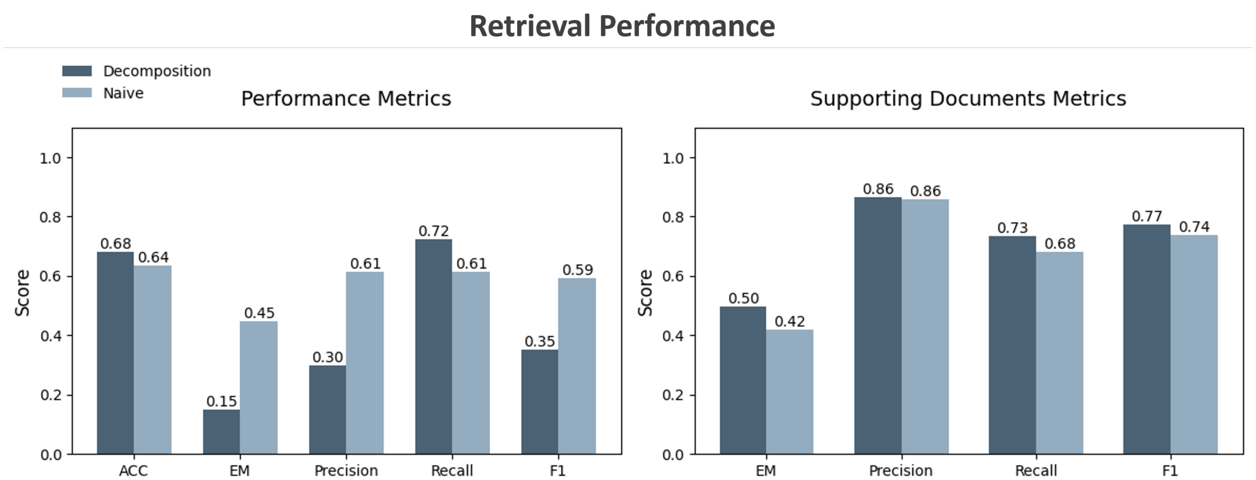

Figure 2: Evaluation metrics for answers and supporting documents, comparing the agent system and naive approach in terms of accuracy, exact match (EM), F1 score, precision, and recall.

Problem Aggregator In the final phase, the results from all sub-problems are aggregated and synthesized into a coherent answer that addresses the original query. This step combines the outputs from the individual resolutions, ensuring that the final solution is comprehensive and well-structured. By integrating sub-solutions, the system reflects a goal-oriented problem-solving process similar to human cognitive strategies, where multiple steps are consolidated to achieve a single, meaningful outcome.

Modular Approach and Cognitive Alignment The modularity of this approach mirrors the hierarchical and goal-directed strategies observed in human cognition. Each phase—reformulation, resolution, and integration—plays a distinct role in managing complexity while enhancing the system's ability to navigate multi-step reasoning challenges. This design allows the agent to focus on task-relevant information at each step, improving its efficiency and accuracy compared to single-step reasoning systems.

4 Experiment

4.1 Experimental Setup

To evaluate the proposed problem decomposition system, I utilized GPT-3.5 Turbo as the closed-source API and implemented the architecture using the LangChain libraries. The evaluation was conducted on the HotPotQA dataset [5], which is specifically designed to test multi-hop reasoning and challenges models with distracting contexts, requiring selective attention to derive the correct answers.

The ground-truth answers and supporting documents from the dataset were compared against the responses generated by the agent system. The evaluation focused on several key performance metrics, including accuracy (correct or incorrect answers), exact match (EM), and F1 scores. Additionally, I used methods from TrustLLM [6] for truthfulness assessment to ensure GPT-based evaluations were consistent and reliable. To benchmark the agent system's performance, I compared it against a single-step LLM baseline (a naive approach) that utilized prompts similar to those used for benchmarking HotPotQA within the TrustLLM framework.

4.2 Results

The overall performance of the agent system, evaluated in terms of answers and supporting documents, is summarized in Figure 2. For answer generation, the agent system achieved an accuracy of 68%, marking a 4% improvement over the nave single-step approach, which achieved 64%. However, while the agent system showed higher accuracy, it performed lower on other evaluation metrics. Specifically, the EM score for the agent system was 15%, representing a 30% decrease compared to the naïve approach, which achieved 45%. Similarly, the F1 score for the agent system dropped to 35%, a 24% decrease compared to the naïve approach's 59%.

In contrast, the agent system demonstrated superior performance in evaluating the supporting documents, outperforming the naïve approach across all key metrics. The agent system achieved an EM score of 50%, which was 8% higher than the naïve approach's 42%. Furthermore, the F1 score for the agent system reached 77%, a 3% improvement over the naïve approach, which achieved 74%. The recall score was also higher in the agent system, reaching 73% compared to 68% for the naïve approach, marking a 5% increase. While both approaches achieved a precision score of 86%, the agent system exhibited a slight advantage in decimal precision.

4.3 Summary

Overall, the results demonstrate that the problem decomposition-based agent system improves accuracy and supporting document performance compared to the nave single-step approach. However, the system exhibits notable weaknesses in EM and F1 scores when evaluating direct answers. These findings suggest that while the agent system excels at focusing on and integrating relevant information, further refinements are necessary to enhance the precision and alignment of individual sub-problem solutions.

5 Discussion

In this study, I demonstrated the effectiveness of leveraging problem decomposition, a strategy inspired by the cognitive top-down processing mechanism. By breaking down complex problems hierarchically, the agent system was able to focus more effectively on relevant information while suppressing distractions. This hierarchical approach allowed the LLM agent to navigate multi-step reasoning tasks more efficiently, as it processed smaller, manageable sub-problems before integrating them into a coherent solution. The ability to prioritize task-relevant details, while ignoring distracting contexts, aligns with the goal-oriented strategies observed in human cognition and highlights the benefits of structurally decomposing problems to improve reasoning performance.

While the system achieved notable improvements in accuracy, its lower performance on EM and F1 scores for answer generation requires further attention. Through manual inspection, I identified that this decline in performance largely stems from the absence of strict instructions for formatting final answers. My decision to avoid oversimplifying answers was intentional, as overly rigid answer structures could result in a loss of critical information. Consequently, the agent's outputs were often more flexible and nuanced, which did not strictly align with the fixed gold labels used in EM and F1 evaluation.

This observation underscores a significant limitation of EM and F1 scores as evaluation methods, particularly for LLMs. LLMs naturally generate diverse and contextually rich responses, which can deviate in format while still preserving the correct meaning. Relying on rigid assessment metrics that use pre-defined answers may overlook the true semantic accuracy of LLM-generated outputs. From this perspective, I incorporated LLM-based evaluations, such as those from TrustLLM, which offer a more flexible assessment framework. These methods better accommodate the variability in valid responses, capturing the true underlying meaning without distorting the evaluation process.

Despite these advancements, there remains a need for more elaborative indices to measure the level of goal-directedness in agent systems. Existing metrics, such as accuracy and EM, do not fully capture the relationship between structurally induced reasoning processes and the representation of goal-directed behavior. Developing refined evaluation methods will allow us to better understand and quantify how cognitive-inspired strategies, such as problem decomposition, contribute to achieving goal-oriented outcomes.

In conclusion, while problem decomposition demonstrates clear advantages in enabling LLMs to focus on relevant information and navigate complex tasks, further research is needed to address limitations in evaluation metrics and develop more robust frameworks for assessing goal-directedness. This will be crucial for advancing AI systems that are both cognitively aligned and capable of efficient, goal-oriented reasoning.

6 Conclusion

This study takes a significant step toward integrating cognitive science insights into AI, enhancing goal-directedness through a problem decomposition approach inspired by top-down processing. The developed LLM agent system demonstrated improved accuracy and robustness in complex tasks, showcasing the value of cognitive strategies in AI reasoning. Future work will focus on refining the module for more complex tasks, scaling to larger datasets, developing novel evaluation indices for goal-directed behavior, and exploring agent system dynamics to implement cognitive modularity. These advancements will contribute to building more efficient, accurate, and human-like goal-oriented AI systems.

References

[1] Liu W. Chen X. Wang X. Wang H. Lian D. ... & Chen E. Huang, X. Understanding the planning of llm agents: A survey. *arXiv preprint arXiv:2402.02716*, 2024.

[2] Wang X. Schuurmans D. Bosma M. Xia F. Chi E. ... & Zhou D. Wei, J. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems, 35, 24824-24837*, 2022.

[3] Gu S. S. Reid M. Matsuo Y. & Iwasawa Y. Kojima, T. Large language models are zero-shot reasoners. *Advances in neural information processing systems, 35, 22199-22213*, 2022.

[4] Zhao J. Yu D. Du N. Shafran I. Narasimhan K. & Cao Y. Yao, S. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

[5] Qi P. Zhang S. Bengio Y. Cohen W. W. Salakhutdinov R. & Manning C. D. Yang, Z. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *arXiv preprint arXiv:1809.09600*, 2018.

[6] Sun L. Wang H. Wu S. Zhang Q. Li Y. ... & Zhao Y. Huang, Y. Trustllm: Trustworthiness in large language models. *arXiv preprint arXiv:2401.05561*, 2024.