ISSN No:-2456-2165

Summarization and Visualization of Files based on Genai

S R Abhiram[1]; Suhas L[2]; Tejas S[3]; Tejaswini K. P.[4]
Department of Information Science
RNS Institute of Technology, Bengaluru

Abstract:- This survey examines advancements in augmenting language models (LMs) with enhanced reasoning abilities and tool-usage capabilities. Reasoning in this context involves breaking down complex tasks into simpler subtasks, while tool use refers to engaging with external modules, such as a code interpreter. LMs can apply these capabilities independently or together through heuristics or through learning from example demonstrations. By utilizing various, often non-parametric external modules, these enhanced LMs expand their ability to process context, shifting beyond traditional language modeling. This type of model is referred to as an Augmented Language Model (ALM). The standard missing token objective enables ALMs to develop reasoning skills, utilize tools, and even perform actions, while still handling typical language tasks—and in some cases, outperforming standard LMs in benchmark tests. This survey concludes that ALMs could potentially overcome significant limitations found in traditional LMs, including issues with interpretability, consistency, and scalability.

Keywords:- *Reasoning, Tool Use, Non-Parametric Module, Missing Token Prediction), Heuristics, Demonstrations, Interpretability, Consistency, Scalability.*

I. INTRODUCTION

The survey investigates recent developments in enhancing language models (LMs) by adding reasoning skills and the ability to use external tools. Reasoning refers to breaking down complex tasks into simpler parts, while tool usage involves integrating with modules like code interpreters to extend functionality. These enhancements allow LMs to apply reasoning and tool-usage abilities independently or jointly, often learned through heuristics or demonstrations. Referred to as Augmented Language Models (ALMs), these models can utilize various external, non-parametric modules to broaden their context capabilities. The ALMs retain the core missing token prediction objective, enabling them to perform typical language tasks while also outperforming many conventional LMs in benchmarks. The survey concludes that ALMs offer a promising approach to address key challenges in traditional LMs, including limitations in interpretability, consistency, and scalability.

A growing trend in research has emerged aimed at addressing the challenges associated with large language models (LLMs), moving slightly away from traditional statistical language modeling approaches. For instance, one line of research enhances the relevance of LLMs by incorporating information from pertinent external documents, effectively mitigating the limitations posed by their constrained context size. By integrating a retrieval module that extracts relevant documents from a database based on the given context, it becomes feasible to achieve comparable capabilities to some of the largest LLMs while utilizing fewer parameters (Borgeaud et al., 2022; Izacard et al., 2022). This results in a non-parametric model capable of querying external data sources. Furthermore, LLMs can enhance their context through reasoning strategies (Wei et al., 2022c; Taylor et al., 2022; Yang et al., 2022c), producing a more relevant context by investing additional computational resources prior to generating responses.

Another approach involves enabling LLMs to utilize external tools (Press et al., 2022; Gao et al., 2022; Liu et al., 2022b) to fill in critical gaps in information not captured within the model's weights. While many of these studies target specific shortcomings of LLMs, it is clear that a systematic integration of both reasoning and tools could yield significantly more powerful models. We will refer to these models as Augmented Language Models (ALMs). As this trend continues to grow, it becomes increasingly challenging to monitor and comprehend the breadth of results, highlighting the need for a taxonomy of ALM research and clear definitions of the technical terms that are sometimes used interchangeably.

II. REASONING

- Previous studies have indicated that while LLMs can tackle simple reasoning tasks, they struggle with more complex ones (Creswell et al., 2022). Consequently, this section will explore various strategies aimed at enhancing the reasoning capabilities of LMs.
- A significant challenge for LMs when faced with complex reasoning problems is accurately deriving solutions by combining the correct answers predicted for sub-problems. For instance, a language model might accurately predict a celebrity's birth and death dates but fail to calculate their age correctly. This issue has been identified by Press et al. (2022) as the compositionality gap in LMs. In the remainder of this section, we will examine three prominent approaches to eliciting reasoning in LMs. It is worth noting that Huang and Chang (2022) have conducted a survey on reasoning within language models, while Qiao et al. (2022) have focused on reasoning through prompting.

- Several studies aim to elicit intermediate reasoning steps by explicitly breaking down problems into sub-problems, facilitating a divide-and-conquer approach. This recursive strategy is particularly beneficial for complex tasks, as compositional generalization can pose significant challenges for language models (Lake and Baroni, 2018; Keysers et al., 2019; Li et al., 2022a). Approaches that utilize problem decomposition can either address the sub-problems independently.
- Despite their impressive outcomes, prompting methods have notable drawbacks, particularly their reliance on model scale. Specifically, they necessitate the identification of effective prompts that can elicit step-by-step reasoning, as well as the manual provision of examples for few-shot learning on new tasks. Additionally, using long prompts can be computationally intensive, and the limited context size of models restricts the ability to take advantage of a large number of examples. Recent research proposes addressing these challenges by training language models (LMs) to utilize a form of working memory.
- Reasoning can generally be understood as the process of breaking down a problem into a series of sub-problems, approached either iteratively or recursively. However, exploring numerous reasoning pathways can be challenging, and there is no assurance that the intermediate steps are valid. One method to create reliable reasoning traces involves generating pairs of questions and their corresponding answers for each reasoning step (Creswell and Shanahan, 2022), but this still does not guarantee the accuracy of those intermediate steps.
- Ultimately, a reasoning language model aims to enhance its context independently to improve its likelihood of producing the correct answer. The extent to which language models actually utilize the identified reasoning steps to inform their final predictions is still not well understood (Yu et al., 2022).
- Often, certain reasoning steps can contain errors that negatively impact the correctness of the final output. For instance, errors in complex mathematical calculations during a reasoning step can result in an incorrect conclusion. Similarly, mistakes regarding well-known facts, such as identifying a president during a specific year, can lead to inaccuracies. Some of the studies mentioned earlier (Yao et al., 2022b; Press et al., 2022) have begun to explore the use of simple external tools.
- Tool such as search engines or calculators, to verify these intermediate steps. The following section of this survey will delve into the various tools that language models can query to enhance their chances of generating correct answers, which could be particularly relevant for your interests in machine learning and language models..
- Similarly, Khot et al. (2022) uses prompts to break down tasks into specific operations but permits each sub-problem to be addressed by a library of specialized handlers, each designed for a particular sub-task (e.g., retrieval).

III. USING TOOLS AND ACTING

A. Iterative LM calling

A growing body of research explores how LMs can access knowledge beyond their internal parameters by interacting with external tools for tasks like precise computations or retrieving information. These tools allow models to "act" when their outputs affect external environments. For example, LMs can be configured to call another model or external tool to refine a generated response iteratively or to connect with modules trained on diverse data types. This multimodal approach expands the model's ability to perform actions or use other resources, such as search engines, web browsers, and virtual or physical agent control, allowing ALMs to perform a broader range of tasks with increased reliability.

Incorporating diverse modalities can enhance the effectiveness of language models (LMs), especially in tasks where context is crucial. For instance, the tone of a question—whether serious or ironic—can significantly affect the type of response required. Recent studies by Hao et al. (2022) and Alayrac et al. (2022) highlight the potential of using LMs as universal interfaces for models that have been pre-trained on various modalities. Hao et al. (2022) integrate several pre-trained encoders that can process different forms of data, such as text and images, into an LM that acts as a universal task layer. This integration, achieved through semi-causal language modeling, combines the advantages of both causal and non-causal approaches, facilitating in-context learning and open-ended generation while also allowing for easy fine-tuning of the encoders.

Language models can be improved through memory units, such as neural caches that store recent inputs (Grave et al., 2017; Merity et al., 2017), which bolster their reasoning capabilities. Alternatively, knowledge can be retrieved from external sources, offloading it from the LM. These memory augmentation strategies help the LM avoid generating outdated information.

Isolation, focusing solely on digital artifacts and struggling to integrate findings across other forensic domains like DNA or physical evidence. Their "black box" nature makes it challenging to present transparent, legally acceptable outputs.

Two types of retrievers can enhance LMs: dense and sparse. Sparse retrievers rely on bag-of-words representations for documents and queries (Robertson and Zaragoza, 2009), whereas dense neural retrievers utilize dense vectors generated from neural networks (Asai et al., 2021). Both types evaluate the relevance of documents to information-seeking queries through either (i) term overlap or (ii) semantic similarity. Sparse retrievers excel at the former, while dense retrievers perform better at the latter (Luan et al., 2021).

When augmenting LMs with dense retrievers, various studies have found success by appending retrieved documents to the existing context (Chen et al., 2017; Clark and Gardner, 2017; Lee et al., 2019; Guu et al., 2020; Khandelwal et al., 2020; Lewis et al., 2020; Izacard and Grave, 2020; Zhong et al., 2022; Borgeaud et al., 2022; Izacard et al., 2022). While retrieving documents for question answering is not a novel concept, retrieval-augmented LMs have recently shown strong performance in other knowledge-intensive tasks beyond Q&A, effectively narrowing the performance gap with larger LMs that require significantly more parameters. REALM (Guu et al., 2020) was the first approach to jointly train a retrieval system with an encoder LM end-to-end. RAG (Lewis et al., 2020) fine-tunes both the retriever and a sequence-to-sequence model together. Izacard and Grave (2020) introduced a modified seq2seq architecture designed to efficiently handle multiple retrieved documents. Borgeaud et al. (2022) developed an auto-regressive LM named RETRO, demonstrating that combining a large corpus with pre-trained, frozen BERT embeddings for the retriever can yield performance comparable to GPT-3 on various downstream tasks without requiring additional training for the retriever.

Overall Limitations Across Solutions: Fragmented Analysis: Most existing AI tools focus on specific evidence types (e.g., digital, genetic, or medicolegal), resulting in fragmented forensic investigations. Lack of Transparency: Many AI models are “black-boxes,” making it difficult to interpret and validate their results, which affects legal acceptability. Slow, Sequential Processing: AI models often analyse evidence sequentially rather than simultaneously, resulting in longer investigation times and delayed insights.

B. Acting on the Virtual and Physical World

Integrated Comprehensive Recent research has shown that language models (LMs) can effectively control virtual agents in both 2D and 3D simulated environments by generating executable functions. For instance, Li et al. (2022b) fine-tuned control virtual agents in both 2D and 3D simulated environments by generating executable functions. For instance, Li et al. (2022b) fine-tuned a pre-trained GPT-2 (Radford et al., 2019) to handle sequential decision-making tasks by encoding goals and observations as a sequence of embeddings and predicting subsequent actions. This framework demonstrated strong combinatorial generalization across various domains, including a simulated household environment. It indicates that LMs can generate representations beneficial for modeling not only language but also sequential objectives and plans, enhancing their learning and generalization capabilities in tasks beyond mere language processing.

In a similar vein, Huang et al. (2022a) explored whether LMs can leverage world knowledge to execute specific actions in response to high-level tasks articulated in natural language, such as "make breakfast." This study was pioneering in demonstrating that, if the LM is sufficiently large and appropriately prompted, it can decompose high-level tasks into a series of simple commands without needing additional training. However, the agent is limited to a predefined set of actions, meaning not all natural language instructions can be executed within the environment. To overcome this limitation, the authors proposed using the cosine similarity function to map the LM-generated commands to feasible actions for the agent. This approach was evaluated in a virtual household setting, where it showed enhanced task execution capabilities compared to relying solely on the LM-generated plans.

While these studies highlight the potential of LMs in controlling virtual robots, other research has focused on physical robots. Zeng et al. (2022) integrated a LM with a visual-language model (VLM) and a pre-trained language-conditioned policy to control a simulated robotic arm. Here, the LM functions as a multi-step planner that decomposes high-level tasks into subgoals, while the VLM describes the objects in the environment. The results are passed to the policy, which executes actions based on the specified goals and the observed state of the world. Dasgupta et al. (2023) employed the 7B and 70B Chinchilla models as planners for an agent that acts and observes results in a PycoLab environment. A reporter module was also utilized to translate actions and observations from pixel data to text format. Lastly, in Carta et al. (2023), an agent employs a LM to generate action policies for text-based tasks, with interactive learning through online reinforcement learning helping to ground the LM's internal representations in the environment, moving away from solely relying on the statistical surface structure of pre-training.

Liang et al. (2022) utilized a LM to create robot policy code based on natural language commands by prompting the model with several demonstrations. By integrating traditional logic structures and referencing external libraries for tasks such as arithmetic operations, LMs can develop policies that exhibit spatial and geometric reasoning, generalize to novel instructions, and provide precise values for ambiguous descriptions. This method proved effective across various real robot platforms. LMs possess common sense knowledge about the world, which can facilitate robots in complex ways.

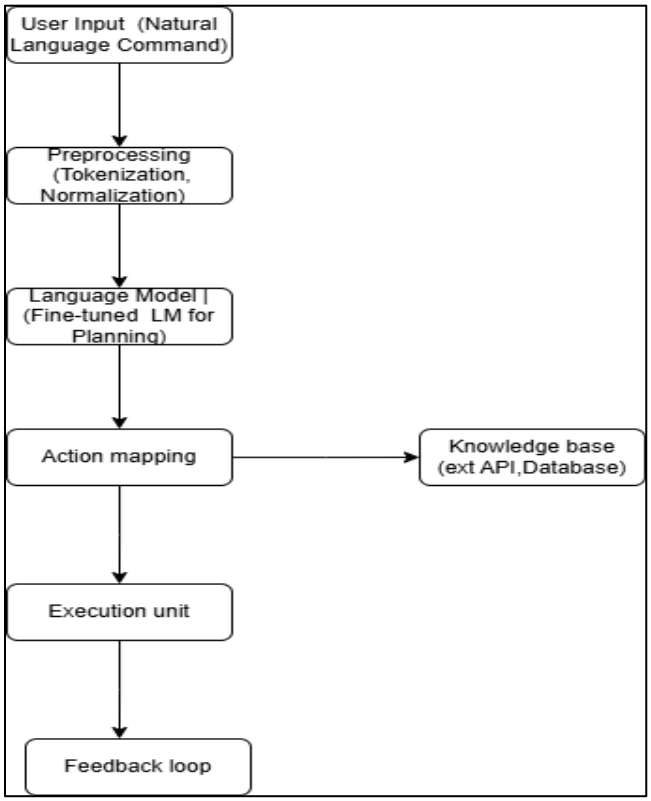

Fig 1: Block Diagram of Proposed Analysis System

Table 1: Traditional method vs Proposed method

Traditional Method	Proposed Method
Direct rule-based programming	Fine-tuning of language models (LMs) for specific tasks.
Sequential command execution.	Multi-step planning and contextual understanding
Static algorithms without learning.	Interactive learning through online reinforcement learning
Limited flexibility and context-awareness	Decomposes high-level tasks into simpler subgoals

IV. CONCLUSION

The development of Augmented Language Models (ALMs) addresses several inherent limitations in traditional language models (LMs). By equipping LMs with reasoning abilities and the capacity to interact with external tools, ALMs demonstrate a notable improvement in handling complex, multi-step tasks. This enhancement allows them to apply structured reasoning and retrieve or compute missing information in ways that traditional LMs cannot achieve within a limited context. The integration of tools—whether for retrieving information, performing computations, or controlling virtual and physical agents—provides ALMs with a more adaptable, context-sensitive approach. However, the field still faces challenges, such as achieving greater interpretability and optimizing models for scalability. The interplay between reasoning and tool use also raises questions about finding the right balance for improved generalization and task efficiency.

This survey concludes that ALMs represent a significant advance in natural language processing, enhancing model versatility and task performance across benchmarks. By moving away from the limitations of purely parametric models, ALMs leverage external modules to expand contextual understanding and real-time adaptability. This shift provides a foundation for LMs to become more autonomous and capable agents, potentially applicable to a wide range of real-world scenarios. Continued research is needed to further refine these capabilities, address current scalability challenges, and explore the broader applications of ALMs in fields that demand robust reasoning and decision-making. In summary, the GenAI-Based Forensic Simulation System sets a new benchmark for AI-driven forensics, offering a holistic approach that enhances the speed, accuracy, and integrity of forensic investigations. It not only bridges the gaps present in existing solutions but also redefines the potential of AI in aiding the criminal justice system.

REFERENCES

[1]. Deborah Cohen, Moonkyung Ryu, Yinlam Chow, Orgad Keller, Ido Greenberg, Avinatan Hassidim, Michael Fink, Yossi Matias, Idan Szpektor, Craig Boutilier, et al. Dynamic planning in open-ended dialogue using reinforcement learning. arXiv preprint
[2]. Antonia Creswell and Murray Shanahan. Faithful reasoning using large language models.
[3]. Satanjeev Banerjee and Alon Lavie. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, pages 65–72.
[4]. Antonia Creswell, Murray Shanahan, and Irina Higgins. Selection-inference: Exploiting large language models for interpretable logical reasoning
[5]. Ishita Dasgupta, Andrew K Lampinen, Stephanie CY Chan, Antonia Creswell, Dharshan Kumaran, James L McClelland, and Felix Hill. Language models show human-like content effects on reasoning.
[6]. Ishita Dasgupta, Christine Kaeser-Chen, Kenneth Marino, Arun Ahuja, Sheila Babayan, Felix Hill, and Rob Fergus. Collaborating with language models for embodied reasoning, 2023.
[7]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.
[8]. Pierre L Dognin, Inkit Padhi, Igor Melnyk, and Payel Das. Regen: Reinforcement learning for text and knowledge base generation using pretrained language models. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
[9]. Chris Donahue, Mina Lee, and Percy Liang. Enabling language models to fill in the blanks. In Proceedings of the Annual Meeting of the Association for Computational Linguistics(ACL), 2020.

[10]. Iddo Drori, Sarah Zhang, Reece Shuttleworth, Leonard Tang, Albert Lu, Elizabeth Ke, Kevin Liu, Linda Chen, Sunny Tran, Newman Cheng, et al. A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. Proceedings of the National Academy of Sciences, 119(32), 2022.
[11]. Andrew Drozdov, Nathanael Schärli, Ekin Akyürek, Nathan Scales, Xinying Song, Xinyun Chen, Olivier Bousquet, and Denny Zhou. Compositional semantic parsing with large language models.
[12]. Dheeru Dua, Shivanshu Gupta, Sameer Singh, and Matt Gardner. Successive prompting for decomposing complex questions. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022.
[13]. Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Jamie Callan, and Graham Neubig. Pal: Program-aided language models, 2022.
[14]. lge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas, et al. Solving rubik's cube with a robot hand.
[15]. Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katie Millican, Malcolm Reynolds, et al. Flamingo: a visual language model for few-shot learning. Advances in Neural Information Processing Systems (NeurIPS), 2022.
[16]. Daniel Andor, Luheng He, Kenton Lee, and Emily Pitler. Giving BERT a calculator: Finding operations and arguments with reading comprehension. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.
[17]. Akari Asai, Xinyan Yu, Jungo Kasai, and Hannaneh Hajishirzi. One question answering model for many languages with cross-lingual dense passage retrieval. Advances in Neural Information Processing Systems (NeurIPS), 2021.
[18]. Akari Asai, Timo Schick, Patrick Lewis, Xilun Chen, Gautier Izacard, Sebastian Riedel, Hannaneh Hajishirzi, and Wen-tau Yih. Task-aware retrieval with instructions
[19]. Lalit R. Bahl, Frederick Jelinek, and Robert L. Mercer. A maximum likelihood approach to continuous speech recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5(2):179–190, 1983.
[20]. Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. arXiv preprint arXiv:2212.08073, 2022.