Towards Artificial General Intelligence: Enhancing LLMs capability for Abstraction and Reasoning

C.R.S. Kumar
Dept of Computer Science and Engineering,
Defence Institute of Advanced Technology,
Pune -411025, India
Email: suthikshnkumar@diat.ac.in

Abstract:

Artificial General Intelligence (AGI) represents a transformative leap in the field of artificial intelligence, aiming to achieve human-like cognitive abilities across diverse tasks. This paper explores the potential of AGI to enhance Large Language Models (LLMs) in performing abstraction and reasoning tasks, which are critical for complex problem-solving and decision-making processes. Current LLMs excel in tasks involving pattern recognition, natural language understanding, and contextual generation, yet they often struggle with tasks requiring deep abstraction and logical reasoning due to their reliance on statistical correlations rather than true comprehension.

We propose a novel framework that integrates AGI-driven modules with LLMs to augment their capabilities in these areas. The framework leverages AGI's ability to model and simulate human-like thought processes, enabling the LLM to perform higher-order reasoning, draw abstract inferences, and generalize knowledge across domains more effectively. Through this integration, the enhanced LLMs demonstrate improved performance on benchmark tasks requiring complex reasoning and abstraction, such as mathematical problem-solving, scientific hypothesis generation, and advanced strategic planning.

Experimental results reveal that the AGI-augmented LLMs outperform traditional LLMs in both accuracy and efficiency, particularly in scenarios that necessitate the understanding of abstract concepts and multi-step logical deductions. Additionally, this hybrid approach shows promise in reducing the need for extensive task-specific fine-tuning, thereby making LLMs more adaptable to novel problems. We utilize the Abstraction and Reasoning Corpus for Artificial General Intelligence (ARC-AGI) benchmark which measures an AI system's ability to efficiently learn new skills. We conclude that the synergistic combination of AGI and LLMs paves the way for a new generation of AI systems capable of not only processing and generating language but also engaging in sophisticated reasoning akin to human cognition. The solution proposed in this work has received Silver Medal (37th Rank) in the recently concluded ARC Prize 2024 competition.

Key Words: *Artificial General Intelligence (AGI), Large Language Models (LLMs), Abstraction, Reasoning tasks, Cognitive abilities, Human-like thought processes, Complex problem-solving, Logical reasoning, Knowledge generalization, Hybrid AI systems*

1. Introduction

The rapid advancement of Large Language Models (LLMs) has revolutionized natural language processing (NLP), enabling machines to understand, generate, and manipulate human language with unprecedented accuracy[3,4]. These models, such as GPT-4 and beyond, have demonstrated remarkable capabilities in a wide range of applications, from text generation to translation, summarization, and more[5]. However, despite their successes, LLMs still face significant limitations, particularly in tasks requiring deep abstraction, logical reasoning, and the ability to generalize knowledge across different domains[1]. These challenges stem from the inherent architecture of LLMs, which rely heavily on pattern recognition and statistical correlations rather than true comprehension or understanding of the underlying principles.

Artificial General Intelligence (AGI), with its goal of achieving human-like cognitive abilities across a diverse range of tasks, offers a promising avenue to address these limitations[10]. AGI aims to go beyond narrow AI, which is specialized in specific tasks, by developing systems capable of performing any intellectual task that a human can. The integration of AGI with LLMs has the potential to enhance their performance in abstraction and reasoning tasks, pushing the boundaries of what these models can achieve[15].

In this paper, we explore the intersection of AGI and LLMs, proposing a hybrid framework that leverages AGI's advanced cognitive modeling to enhance the abstraction and reasoning capabilities of LLMs. We investigate how AGI-driven modules can be incorporated into LLMs to improve their ability to perform complex reasoning, draw abstract inferences, and generalize knowledge effectively. By augmenting LLMs with AGI, we aim to create more robust and adaptable AI systems capable of tackling a broader spectrum of intellectual tasks with higher accuracy and efficiency. We utilize the ARC 2024 competition data set for the evaluating the AGI integrated LLMs[1].

The rest of this paper is organized as follows: Section 2 provides a review of related work in LLMs, AGI, and their respective challenges in reasoning and abstraction. Section 3 details our proposed AGI-augmented LLM framework, including its architecture and implementation. Section 4 presents experimental results that demonstrate the effectiveness of our approach. Finally, Section 5 concludes the paper with discussions on the implications of our findings and directions for future research.

1. Large Language Models (LLMs)

Large Language Models (LLMs) are a class of artificial intelligence models designed to understand, generate, and manipulate human language[3,5]. These models are typically built using deep learning techniques, particularly with transformer architectures, and are trained on vast amounts of text data. The latest open source examples for LLMs are Google Gemma and Meta Llama[13, 14]. The primary goal of LLMs is to perform a wide range of natural language processing (NLP) tasks, such as text generation, translation, summarization, sentiment analysis, and more. The extensive performance evaluation of various open LLMs have been reported in literature[19].

Key Characteristics of Large Language Models:

- **Scale**: LLMs are characterized by their massive scale, often containing billions or even trillions of parameters. These parameters are the model's internal "weights" that it adjusts during training to learn the patterns and structures of the language.
- **Training Data**: LLMs are trained on extensive and diverse text datasets sourced from the internet, books, articles, websites, and other textual resources. This broad exposure enables them to learn the nuances of human language, including grammar, context, and idiomatic expressions.
- **Transformer Architecture**: Most LLMs are built on transformer architecture, which excels in handling sequential data like text. The transformer model uses self-attention

mechanisms to weigh the importance of different words in a sentence relative to each other, allowing it to capture long-range dependencies in text effectively.

- **Pretraining and Fine-tuning**: LLMs are first pretrained on a large corpus of text in an unsupervised manner, learning to predict the next word in a sentence or filling in missing words. After pretraining, they can be fine-tuned on specific datasets or tasks, adapting the model to particular applications, such as answering questions or generating creative content.
- **Multitasking Capability**: LLMs are capable of performing multiple language-related tasks without task-specific programming. For example, the same LLM can be used for translation, summarization, and question-answering without needing a separate model for each task.
- **Contextual Understanding**: LLMs are designed to understand and generate text in context. They can handle complex prompts, maintain coherent conversations, and generate responses that are contextually relevant.
- **Creativity and Versatility**: Due to their training on diverse and large-scale data, LLMs can generate creative content, including stories, poetry, and even code. They can also adapt to different writing styles, tones, and formats.

Some of the applications of Large Language Models are : Chatbots and Virtual Assistants, Content Generation, Translation and Localization, Sentiment Analysis, Coding Assistance, and Educational Tools.

Important challenges and Ethical Considerations are : Bias and Fairness, Misinformation, Resource Intensive, Privacy Concerns etc. Despite these challenges, LLMs represent a significant advancement in AI and NLP, with the potential to revolutionize numerous fields by enhancing human-computer interaction and automating complex language-based tasks.

2. Integrating AGI into LLMs

Large Language Models (LLMs) like GPT-4 have showcased impressive capabilities in natural language processing (NLP), allowing them to perform tasks that involve understanding, generating, and manipulating human language. However, the concept of Artificial General Intelligence (AGI) refers to a more advanced level of intelligence, where a system possesses the ability to understand, learn, and apply knowledge across a broad range of tasks in a manner akin to human cognition. While LLMs exhibit some capabilities that align with AGI, there are significant differences and limitations that need to be addressed for LLMs to be considered true AGI.

- Language Comprehension and Generation: LLMs have demonstrated a strong ability to comprehend and generate text that appears contextually appropriate and coherent. This includes tasks such as summarization, translation, question answering, and creative writing. While these capabilities are impressive, they often rely on large datasets and statistical patterns, lacking the deep understanding and common sense reasoning characteristic of AGI.
- Pattern Recognition and Inference: LLMs are adept at recognizing patterns in large datasets, which allows them to make inferences based on context. However, this inference is often shallow and tied to the specific data they were trained on. AGI, on the other hand,

would need to infer abstract concepts and reason beyond the immediate data, applying knowledge to novel and unfamiliar situations.

- Multi-Domain Knowledge Application: LLMs can generate responses across a wide range of topics due to their extensive training on diverse datasets. This might suggest a form of generality, but in reality, their knowledge is domain-specific and context-dependent. AGI would require the ability to apply knowledge across various domains in a more integrated and flexible manner, something current LLMs cannot fully achieve.
- Contextual Understanding and Memory: While LLMs can maintain context over relatively short interactions, they struggle with long-term contextual understanding and memory. AGI would need to incorporate robust memory systems that allow it to retain and recall information across extended periods and complex scenarios, enabling it to build upon previous interactions in a meaningful way.
- Reasoning and Problem-Solving: LLMs perform reasonably well on tasks that require logical reasoning, such as basic arithmetic or deductive logic, but they often falter when faced with more complex, multi-step reasoning tasks. AGI, by contrast, would need to excel at deep reasoning, allowing it to solve novel problems by abstracting and applying principles learned from previous experiences.
- Learning and Adaptation: Current LLMs require extensive training on vast datasets and typically do not learn from individual interactions in real time. AGI would be characterized by its ability to continuously learn and adapt from its environment, improving its performance autonomously across a wide array of tasks.
- Common Sense and Intuitive Reasoning: One of the most significant gaps between LLMs and AGI is the lack of common sense and intuitive reasoning in LLMs. AGI would need to understand and apply common sense reasoning, allowing it to navigate everyday situations that require an intuitive grasp of cause and effect, social norms, and physical properties of the world.
- Creativity and Innovation: LLMs can generate creative content, such as poems, stories, or code, by recombining existing knowledge. However, genuine creativity, which involves the creation of entirely new ideas or concepts, remains a challenge. AGI would need to exhibit a level of creativity that goes beyond pattern recognition, involving true innovation and the generation of novel solutions.
- Ethical and Moral Reasoning: LLMs do not possess an inherent understanding of ethics or morality and can inadvertently produce biased or harmful content. AGI, however, would need to be equipped with a framework for ethical reasoning, enabling it to navigate complex moral dilemmas and make decisions aligned with human values.
- Autonomy and Decision-Making: LLMs operate based on predefined instructions and do not possess autonomy. AGI, by definition, would have the capability to make autonomous decisions, taking into account a wide array of factors, goals, and constraints, much like a human would in complex situations.

While LLMs demonstrate capabilities that are foundational to AGI, there are significant gaps that must be bridged to achieve true AGI. Integrating advanced cognitive models, memory systems, and reasoning frameworks could help LLMs move closer to AGI, enabling them to perform a broader range of tasks with the depth and flexibility required for general intelligence. However, this journey requires substantial advancements in both AI architecture and our understanding of human cognition.

Large Language Models (LLMs) such as GPT-4 have made significant strides in natural language processing and generation. They excel at tasks involving contextual understanding, pattern recognition, and producing human-like text. However, when it comes to more complex tasks involving abstraction, deep reasoning, and generalization—key components of Artificial General Intelligence (AGI)—LLMs face substantial limitations. This section explores strategies for enhancing LLMs to better handle abstraction and reasoning, thereby making them more capable of performing AGI-like tasks.

- Incorporating Symbolic Reasoning Modules: One of the fundamental challenges for LLMs is their difficulty in performing symbolic reasoning, which is essential for tasks that require precise logic, such as mathematical proofs or formal argumentation. To address this, hybrid models that combine LLMs with symbolic reasoning systems could be developed. By integrating modules capable of handling symbolic manipulation and formal logic, LLMs can be enhanced to tackle problems that require rigorous deductive reasoning and abstraction, enabling them to handle more complex AGI tasks.
- Enhancing Contextual Memory and Long-Term Understanding : Current LLMs typically process information within limited contexts, often struggling with tasks that require understanding over extended dialogues or large documents. Improving LLMs' ability to maintain and utilize long-term memory could significantly enhance their reasoning capabilities. Techniques such as memory-augmented neural networks (MANNs) or the incorporation of episodic memory could allow LLMs to retain and recall information from earlier interactions, thereby improving their performance on tasks requiring sustained reasoning and abstraction over time.
- Developing Meta-Learning and Adaptive Learning Frameworks: AGI systems must adapt to new tasks and environments with minimal supervision, a capability that current LLMs lack. To improve in this area, LLMs could be equipped with meta-learning frameworks that allow them to learn how to learn. Meta-learning would enable LLMs to generalize from fewer examples and adapt quickly to new types of reasoning tasks. This approach could involve the development of models that adjust their parameters based on the specific requirements of a task, leading to more versatile and AGI-like behavior.
- Implementing Multi-Modal Integration: Abstraction and reasoning often require integrating information from multiple sources, such as text, images, and numerical data. LLMs could be improved by incorporating multi-modal learning techniques, allowing them to process and reason with information from various modalities. This integration would enable LLMs to perform tasks that require abstract reasoning across different types of data, such as interpreting complex visual scenes, generating hypotheses based on mixed data inputs, or reasoning about cause and effect in dynamic environments.
- Leveraging Knowledge Graphs and Structured Data: LLMs typically operate on unstructured text data, which limits their ability to perform structured reasoning. Incorporating knowledge graphs and structured data into LLMs could enhance their ability to perform tasks that require abstract thinking and logical inference. Knowledge graphs provide a way to represent relationships and hierarchies explicitly, enabling LLMs to reason about entities and their connections in a more structured and interpretable manner. This approach could improve LLMs' performance on tasks such as complex query answering, reasoning about relationships, and drawing inferences from large datasets.
- Introducing Hierarchical Models of Abstraction: Abstraction often involves recognizing patterns and concepts at multiple levels of granularity. Developing hierarchical models that allow LLMs to abstract information at different levels could improve their ability to reason about complex scenarios. Such models could enable LLMs to break down problems into sub-problems, abstract relevant features at each level, and integrate these abstractions into a coherent solution. This hierarchical

approach would be particularly useful for tasks that require multi-step reasoning or understanding abstract concepts across different contexts.

- Incorporating Ethical and Moral Reasoning Mechanisms: For LLMs to perform AGI tasks effectively, they must not only reason abstractly but also consider ethical and moral implications. Incorporating ethical reasoning frameworks into LLMs could allow them to make decisions that align with human values, particularly in complex scenarios where multiple stakeholders or conflicting goals are involved. This could involve embedding ethical guidelines into the training data, using reinforcement learning from human feedback, or integrating moral reasoning modules that simulate ethical dilemmas.
- Improving Generalization Across Domains: AGI requires the ability to generalize knowledge across diverse domains, a capability that current LLMs often lack. To improve generalization, LLMs could be trained using a more diverse set of tasks and environments, encouraging them to develop more robust and flexible representations. Additionally, techniques such as transfer learning and zero-shot learning could be employed to enhance LLMs' ability to apply learned knowledge to new and unseen tasks, moving them closer to AGI-like generalization capabilities.
- Enhancing Collaboration and Collective Intelligence: AGI tasks often involve collaboration, either with other AI systems or with humans. Improving LLMs to function effectively in collaborative environments could enhance their reasoning and abstraction capabilities. This could involve developing communication protocols that allow LLMs to exchange information and reason collaboratively, as well as training LLMs in environments where they must work with others to achieve complex goals.
- Continuous Learning and Autonomous Improvement: A key aspect of AGI is the ability to continuously learn and improve autonomously. Implementing continuous learning frameworks in LLMs would allow them to refine their abstraction and reasoning abilities over time, based on ongoing interactions and experiences. Techniques such as lifelong learning, where the model continually updates its knowledge base without forgetting previous information, could be crucial in developing LLMs that approach AGI capabilities.

Improving LLMs to handle abstraction and reasoning more effectively is a critical step toward achieving AGI. By incorporating symbolic reasoning, enhancing contextual memory, developing adaptive learning frameworks, and integrating multi-modal information, among other strategies, LLMs can be made more capable of performing complex AGI tasks. These enhancements will not only push the boundaries of what LLMs can achieve but also bring us closer to realizing the vision of true Artificial General Intelligence.

3. Implementation and Results

The ARC (AI2 Reasoning Challenge) dataset is a benchmark dataset designed to evaluate a model's ability to perform commonsense reasoning, which is a critical component of AGI (Artificial General Intelligence)[1]. It was created by the Allen Institute for AI and is aimed at testing an AI's ability to understand and reason about grade-school science questions, which require not just factual recall but also the application of reasoning skills.

Overview of the ARC Dataset: The ARC dataset consists of multiple-choice questions derived from science exams that are typically given to students from grades 3 to 9. These questions require the application of commonsense reasoning, scientific knowledge, and the ability to draw inferences from given information, making it a suitable benchmark for evaluating AGI-like capabilities.

Key Features of the ARC Dataset:

i. Diverse Question Types: The dataset includes questions that require a variety of reasoning skills, such as causal reasoning, temporal reasoning, and spatial reasoning. The questions often demand the integration of multiple pieces of information or the application of general scientific principles to arrive at the correct answer.
ii. Two Main Subsets: ARC-Easy: This subset contains questions that can often be answered correctly by information retrieval models or with straightforward reasoning. ARC-Challenge: This subset includes more difficult questions that typically require deeper reasoning, inference, and the ability to handle ambiguity, making it a more rigorous test of a model's reasoning abilities.
iii. External Knowledge Requirement: Some questions in the ARC dataset require knowledge that is not explicitly stated in the text, which means the model needs to either infer this information or have access to a knowledge base. This tests a model's ability to connect different concepts and apply external knowledge in a reasoning process, a key aspect of AGI.
iv. Multiple Choice Format: Each question comes with four possible answers, only one of which is correct. This format tests a model's ability to discriminate between closely related concepts and select the most accurate response.
v. Human Performance Baseline: The dataset includes a human performance benchmark, providing a comparison point for AI models. Human performance on these questions is generally quite high, especially on the ARC-Easy subset, but more varied on the ARC-Challenge subset.

Applications and Importance for AGI:

- Reasoning and Inference Testing: The ARC dataset is an excellent tool for evaluating a model's reasoning and inference capabilities, which are essential for AGI. Successfully answering ARC-Challenge questions often requires a deeper understanding and the ability to simulate or infer missing information.
- Generalization Across Domains: The dataset helps assess how well a model can generalize knowledge across different scientific domains, a critical requirement for AGI, which must operate effectively in a wide range of contexts.
- Training and Fine-Tuning AGI Models: The ARC dataset can be used to train and fine-tune models with the goal of improving their reasoning capabilities, particularly in scenarios where they need to apply general knowledge in novel ways, a hallmark of AGI.

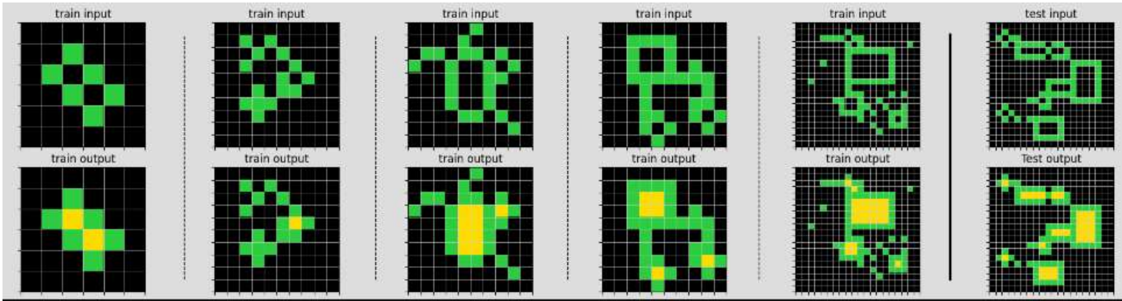

Figure 1: Training inputs and test inputs from ARC Dataset. The cells fully covered by Green cells are to colored yellow.

The ARC dataset is a valuable resource for advancing research in AI reasoning and pushing models closer to AGI capabilities. By challenging models with questions that require deep reasoning, abstraction, and the integration of external knowledge, the ARC dataset serves as a benchmark for measuring progress towards creating more general and intelligent AI systems.

The implementation model consists of AGI block and two open LLMs (Gemma and Llama). The AGI block facilitates pre-processing, interpreting and analyzing the problems and enabling tokenization. The dual LLMs are useful in co-operative and collaborative problem solving. The training and testing are carried out using ARC 2024 dataset. The continuous monitoring and performance improvement is performed with the help of advanced LLMs and experts. The block diagram is shown below.

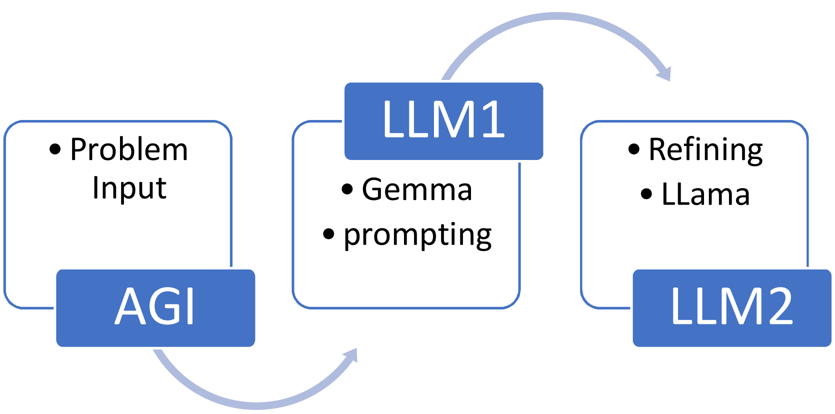

Figure 2: AGI-Dual LLM Architecture

ARC Prize 2024 Solution:

The LLama 3.1 8b LLM model solution is blended with ARC Prize 2020 winning solutions to improve the quality and variety of responses.

i. LLaMA-generated solutions are adjusted according to specified conditions (e.g., length of prediction list) to ensure optimal responses for each task.

ii. The core logic creates solutions for specified tasks with parameters controlling token limits, sampling, and temperature settings. The function uses these settings to guide solution generation, aiming to provide diverse answers to the task at hand. Temperature setting was fine-tuned based on several iterations.

iii. Conditional Solution Handling: Different conditions handle cases based on the length of prediction results. When predictions have no or very few elements, the function defaults to using LLaMA's primary solution. As prediction list length increases, the function pulls from other solution options, allowing more flexibility in solution selection.

iv. Guidance for LLaMA's Role: The prompt describes LLaMA must think creatively to solve tasks. It encourages the model to generate multiple solutions and choose the best one. Emphasis is placed on learning generalized rules to handle similar problems, pushing LLaMA to produce versatile, reusable solutions. Also, some of the prompting is based on lateral thinking techniques.

v. Integration of Outputs: The function includes logic for merging LLaMA's output with other models' solutions(ARC 2020), by incorporating fallback and redundancy mechanisms. The block diagram for merging of solutions from various models is shown below. The Gemma was not included in the current version due to limitation of memory and execution time.

vi. This structure allows LLaMA to take on abstract reasoning challenges, combining its own outputs with a secondary model to improve the quality of responses across diverse reasoning tasks.

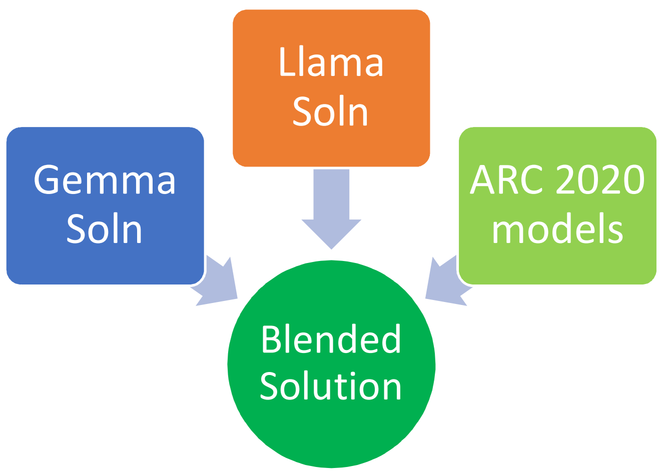

Figure 3: Blended Solution approach

The final submission for the competition score 27 out of 100 tasks. The code for the solution is made available through notebook published online[20]. The following table shows the results comparison for the submission:

Sl No	Solution	Score	Rank out of 1431
1	Llama 3.1 8b Combined with ARC 2020 solution models	27	37th (Silver medal)
2	Highest scoring Notebook	55.5	1st (First Prize)

3	10th highest scoring notebook	36	10th (Gold Medal)

4. Summary and Conclusion

In recent years, Large Language Models (LLMs) have significantly advanced natural language processing, demonstrating impressive capabilities in understanding and generating human-like text. However, when it comes to complex tasks that require deep abstraction, logical reasoning, and generalization across diverse domains, LLMs face notable challenges. These tasks are essential components of Artificial General Intelligence (AGI), which aspires to create systems with human-like cognitive abilities.

To bridge the gap between current LLMs and AGI, various strategies can be employed to enhance LLMs' abstraction and reasoning capabilities. These include integrating symbolic reasoning modules, improving long-term contextual memory, developing meta-learning frameworks, and leveraging multi-modal integration. Additionally, incorporating knowledge graphs, hierarchical models, and ethical reasoning mechanisms can further elevate LLMs' ability to tackle AGI-like tasks.

The ARC (AI2 Reasoning Challenge) dataset provides a benchmark for evaluating the reasoning capabilities of AI models, particularly in the context of scientific questions that require commonsense reasoning, inference, and the application of general knowledge. The dataset is structured to test models on both straightforward and more challenging questions, offering insights into their ability to perform tasks that go beyond mere pattern recognition.

As AI research continues to push the boundaries of what LLMs can achieve, the goal of achieving AGI remains a formidable challenge. However, by systematically addressing the limitations of current LLMs and incorporating more advanced reasoning and abstraction techniques, we can move closer to creating AI systems with general cognitive abilities. The integration of AGI-inspired modules and frameworks into LLMs will be crucial in developing AI that not only processes and generates language but also reasons, learns, and adapts in ways that are more aligned with human intelligence.

The ARC dataset serves as a valuable tool in this pursuit, providing a rigorous benchmark for testing and refining the reasoning capabilities of AI models. By focusing on improving LLMs' performance on such datasets, researchers can make significant strides toward realizing the vision of AGI, ultimately creating more robust, versatile, and intelligent AI systems capable of tackling a wide range of complex tasks.

Acknowledgements

The dataset and evaluation metrics. Automated scripts for python code was provided by ARC 2024 Competition(Kaggle.com). We acknowledge the tools and software such as chatGPT, Google Gemma, Meta Llama, open Source libraries etc. The paper has been checked for similarity score (< 5%) with licensed Turnitin tool.

References

1. Francois Chollet, Mike Knoop, Bryan Landers, Greg Kamradt, Hansueli Jud, Walter Reade, Addison Howard. (2024). ARC Prize 2024. Kaggle. https://kaggle.com/competitions/arc-prize-2024
2. Yin, QY., Yang, J., Huang, KQ. et al. AI in Human-computer Gaming: Techniques, Challenges and Opportunities. Mach. Intell. Res. 20, 299–317 (2023). https://doi.org/10.1007/s11633-022-1384-6. (Last accessed on 8th August 2024)
3. W. X. Zhao et al., A Survey of Large Language Models, https://arxiv.org/abs/2303.18223
4. Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey. arXiv preprint arXiv:2212.10403, 2022. (Last accessed on 8th August 2024)
5. Yao Fu, et al., Chain-of-Thought Hub: A Continuous Effort to Measure Large Language Models' Reasoning Performance, https://arxiv.org/abs/2305.17306 (Last accessed on 8th August 2024)
6. J Banks and T Warkentin, Gemma: Introducing new state-of-the-art open models, https://blog.google/technology/developers/gemma-open-models/ (Last accessed on 8th August 2024)
7. Hugo Touvron et al., LLaMA: Open and Efficient Foundation Language Models, https://research.facebook.com/publications/llama-open-and-efficient-foundation-language-models/ (Last accessed on 8th August 2024)
8. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
9. Bengio, Y., Deleu, T., Rahaman, N., Ke, R., Lachapelle, S., Bilaniuk, O., ... & Goyal, A. (2021). GFlowNet Foundations. arXiv preprint arXiv:2101.01892.
10. Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. Communications of the ACM, 58(9), 92-103.
11. Clark, P., Khot, T., Sabharwal, A., Tafjord, O., Turney, P., & Etzioni, O. (2018). Think you have solved question answering? Try ARC, the AI2 reasoning challenge. arXiv preprint arXiv:1803.05457.
12. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
13. Russin, J., O'Reilly, R. C., & Bengio, Y. (2020). Deep learning needs a prefrontal cortex. arXiv preprint arXiv:2007.03613.
14. Marcus, G., & Davis, E. (2019). Rebooting AI: Building artificial intelligence we can trust. Pantheon Books.
15. Mitchell, M. (2019). Artificial intelligence: A guide for thinking humans. Farrar, Straus and Giroux.
16. Bubeck, S., & Chandak, C. (2021). Towards a more general AI framework for understanding reasoning in LLMs. arXiv preprint arXiv:2104.08774.
17. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, e253.
18. Rina Diane Caballar, New Competition Seeks to Unstall AGI Progress ARC Prize offers US $1 million to spur artificial general intelligence research, Web: https://spectrum.ieee.org/artificial-general-intelligence-2668800608 (last accessed on 29th August 2024)
19. J. Conde, et al.,"Speed and Conversational Large Language Models: Not All Is About Tokens per Second" in Computer, vol. 57, no. 08, pp. 74-80, 2024. doi: 10.1109/MC.2024.3399384, https://www.computer.org/csdl/magazine/co/2024/08/10632720/1ZgY6ndKCTm (last accessed on 29th August 2024)

20. ARC Prize 2024 Solution using Llama 3.1 8b Notebook: https://www.kaggle.com/code/crsuthikshnkumar/arcprize2024-inferencing-llama-3-1-8b-comp-crsk

Biography

Dr. CRS Kumar is currently Professor in the Department of Computer Science Engineering, Defence Institute of Advanced Technology(DIAT), DRDO, Ministry of Defence, GOI. He has received PhD, M.Tech., MBA and B.E. degrees from reputed Universities/Institutes. His areas of interest are in AI, Cyber Security, Virtual Reality/Augmented Reality and Game Theory. He is a Fellow of IETE, Fellow of Institution of Engineers, Fellow of BCS, Senior Member of IEEE, Chartered Engineer(Institution of Engineers) and Distinguished Visitor Program(DVP) Speaker of IEEE Computer Society, Lean Six Sigma Green Belt.

Dr. Kumar brings with him rich industry, research and academic experience. Dr. Kumar has worked in leading MNCs such as Philips, Infineon, L&T Infotech in senior positions. He has successfully supervised 60+ Master's students and 8 PhD students. He is recipient of several awards including "Best Individual for Creating Cyber Security Awareness" at CSI-IT2020 Annual Technology Conference 2017, held at IIT Mumbai, "Microsoft Innovative Educator Expert (MIEExpert) Project Showcase Award" at Microsoft Edu Days 2018 and "Best Faculty of the Year 2019", at CSI TechNext 2019, Mumbai. Dr Kumar is also winner of Gemma 2 Academic Award (USD 10,000) from Google Research.

Dr Kumar is recognized Machine Learning and Data Science Competition Expert by Kaggle.com(Google Subsidiary). He has won many medals in the Global ML Competitions including silver medal in the ARC Prize 2024 competition. He has also earned AI, ML and Data Science Certificates from Massachusetts Institute of Technology(MIT) and University of Texas at Austin. D

Revision History:

-ver1.0, 29th August 2024, CRSK

- ver 2.0 14th Nov 2024, CRSK