Large Language Models(LLM) for Automating 20 Questions Game

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

Abstract:
The 20 Questions game, a classical problem in artificial intelligence and natural language processing, involves one player thinking of an object while the other player asks yes/no questions to guess the object within 20 questions. Traditional approaches to this game have relied on decision trees, heuristic algorithms, and rule-based systems. However, the advent of Large Language Models (LLMs) such as GPT-3 and GPT-4 has opened new avenues for enhancing the performance and interactivity of AI players in this game.

This paper explores the application of LLMs in playing the 20 Questions game, focusing on the model's ability to generate and interpret natural language questions and responses. We detail the architecture and training process of an LLM tailored for this task, highlighting modifications made to optimize the model's performance in understanding and narrowing down possible objects. Through extensive testing, we demonstrate that LLMs can achieve a higher success rate and engage in more coherent and contextually appropriate dialogues compared to traditional methods.
Additionally, we analyze the model's performance across various object categories, noting the strengths and limitations of LLMs in dealing with different types of objects. Our results show that while LLMs excel in categories with well-defined and common attributes, they face challenges with abstract or less common objects. We propose several strategies for overcoming these limitations, including hybrid approaches that combine LLMs with specialized knowledge bases and reinforcement learning techniques.

The findings of this study suggest that LLMs hold significant potential for not only improving the 20 Questions game but also advancing broader applications in conversational AI, interactive entertainment, and educational tools. Future work will focus on refining the model's questioning strategy, enhancing its knowledge representation, and exploring its application in other interactive scenarios.

Key Words: *Large Language Models (LLMs), 20 Questions game, Natural Language Processing (NLP), Decision trees, Heuristic algorithms, Rule-based systems, Dialogue coherence, Object categories, Reinforcement learning*

1. Introduction

The 20 Questions game is a well-known problem in artificial intelligence (AI) and natural language processing (NLP) that involves one player selecting an object and the other player attempting to guess the object by asking up to 20 yes/no questions[1]. This game presents a unique challenge for AI systems due to its requirement for efficient information gathering, logical reasoning, and natural language understanding[2]. Traditionally, AI systems have approached this problem using decision trees, heuristic algorithms, or rule-based systems that map out potential objects and the corresponding questions that might lead to their identification. However, these methods often struggle with the complexity and variability inherent in human language, especially when dealing with ambiguous or less common objects.

With the advent of Large Language Models (LLMs), such as OpenAI's GPT-3 and GPT-4, there has been a significant shift in how AI systems approach natural language tasks[3]. LLMs are trained on vast amounts of text data, allowing them to generate and interpret language with a high degree of fluency and contextual understanding[5]. This makes them particularly well-suited for interactive tasks like the 20 Questions game, where the AI must engage in a dynamic conversation, asking pertinent questions and interpreting the responses in real-time[6,7].

This paper investigates the application of LLMs to the 20 Questions game, exploring how these models can enhance the AI's ability to generate effective questions and accurately guess the object in question. Unlike traditional methods that rely on pre-defined decision-making processes, LLMs can adapt their questioning strategy based on the context and prior responses, leading to a more flexible and human-like interaction. We hypothesize that LLMs can outperform conventional approaches by leveraging their deep language understanding and ability to process complex linguistic patterns.

In this study, we first review the limitations of existing approaches to the 20 Questions game and then present our methodology for adapting LLMs to this task. We conduct extensive experiments to evaluate the performance of LLMs in different scenarios, analyzing their strengths and weaknesses in various object categories. The results of our experiments provide insights into the potential of LLMs to revolutionize not only the 20 Questions game but also broader applications in conversational AI and interactive entertainment.

Ultimately, this research aims to contribute to the ongoing development of more sophisticated and interactive AI systems, pushing the boundaries of what is possible in natural language processing and AI-driven communication. This paper specifically addresses the challenges and issues faced by LLMs for games such as 20 Questions through the datasets, ideas, code shared based on Machine learning Competition[1].

2. 20 Questions Game

The 20 Questions game is a classic parlor game that involves two players. One player (the "answerer") thinks of an object, person, or place, while the other player (the "questioner") attempts to guess what the answerer is thinking of by asking up to 20 yes/no questions[8]. The goal of the questioner is to narrow down the possibilities and identify the correct answer within 20 questions. The game is a simple yet effective exercise in logical reasoning, deduction, and categorization.

Basic Rules of the 20 Questions Game are as follows:

- Selection of the Subject: The answerer thinks of an object, person, or place but does not reveal it to the questioner.

- Asking Questions: The questioner asks yes/no questions to gather information about the subject. The answerer can only respond with "yes," "no," or sometimes "I don't know" or "maybe" if applicable.
- Narrowing Down: Each question should help the questioner eliminate possible answers and narrow down the field of possibilities. For example, a questioner might start by asking, "Is it a living thing?" to divide the possibilities into living and non-living categories.

- Guessing the Subject: At any point, the questioner can make a guess about the subject. If the guess is correct, the game ends, and the questioner wins. If incorrect, the game continues until 20 questions have been asked.
- Winning or Losing: If the questioner guesses the subject correctly within 20 questions, they win the game. If not, the answerer wins.

Variations and Applications:

- Reverse 20 Questions: In this version, the roles are reversed; the answerer provides a description of an object, and the questioner tries to guess it based on the description.
- Digital and AI Versions: The 20 Questions game has been adapted into digital formats, including websites and mobile apps, where an AI or computer algorithm plays the role of the questioner or answerer.
- Educational Use: The game is often used in educational settings to teach students logical thinking, categorization, and language skills.

In the context of artificial intelligence, the 20 Questions game is often used as a benchmark for testing AI systems' ability to process natural language, perform logical reasoning, and interact with humans in a meaningful way[1]. AI systems, especially those based on Large Language Models (LLMs), can play the game by generating relevant questions, understanding responses, and making educated guesses about the subject. The challenge for AI in this game lies in efficiently narrowing down possibilities and handling ambiguous or unexpected answers.

3. 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.

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.

Applications of Large Language Models:

- **Chatbots and Virtual Assistants**: LLMs power chatbots and virtual assistants, enabling them to engage in natural and meaningful conversations with users, answer queries, and provide assistance in real-time.
- **Content Generation**: LLMs can generate high-quality written content, including articles, reports, and creative writing, which can be used in various industries such as journalism, marketing, and entertainment.
- **Translation and Localization**: LLMs facilitate machine translation by converting text from one language to another while preserving the meaning, tone, and context of the original content.
- **Sentiment Analysis**: Businesses use LLMs to analyze customer feedback, reviews, and social media posts to gauge public sentiment and make data-driven decisions.
- **Coding Assistance**: LLMs can assist developers by generating code snippets, debugging code, and providing recommendations based on the context of the code being written.
- **Educational Tools**: LLMs are used in educational platforms to provide personalized tutoring, generate practice questions, and help students understand complex topics.

Challenges and Ethical Considerations:

- **Bias and Fairness**: LLMs can inadvertently learn and propagate biases present in the training data, leading to biased or unfair outcomes in their responses.
- **Misinformation**: Due to their ability to generate human-like text, LLMs can be used to create convincing fake news or misinformation, raising concerns about their impact on public discourse.
- **Resource Intensive**: Training and deploying LLMs require significant computational resources, making them accessible primarily to large organizations with substantial budgets.
- **Privacy Concerns**: LLMs trained on publicly available data might inadvertently generate responses that reveal sensitive or private information.

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.

4. LLM for 20 Questions Game

Using Large Language Models (LLMs) for the 20 Questions game involves leveraging their advanced natural language understanding and generation capabilities to improve the game's interactive experience. The goal is for the LLM to play the role of either the questioner or the answerer, depending on the specific implementation. Here's how LLMs can enhance the 20 Questions game.

Question Generation:

- Adaptive Questioning: LLMs can generate contextually appropriate and adaptive questions based on previous responses. Unlike traditional decision trees or rule-based systems, which may follow a predefined sequence of questions, LLMs can dynamically adjust their questioning strategy to maximize information gain and quickly narrow down the possibilities.
- Understanding Ambiguity: LLMs can handle ambiguous responses more effectively by asking follow-up questions to clarify uncertain or vague answers. For instance, if a player responds with "maybe" or "I don't know," the LLM can generate a question that probes the area of uncertainty further.

Answer Interpretation:

- Contextual Understanding: LLMs excel at understanding the context of answers provided by the player. They can interpret nuanced language and consider synonyms, related concepts, or implied meanings in the responses, allowing for a more flexible and human-like interaction.
- Handling Miscommunication: If a player provides an unexpected answer, LLMs can recognize potential miscommunication and adjust the questioning strategy accordingly, improving the overall flow of the game.

Object Categorization:

- Knowledge Integration: LLMs are trained on vast amounts of data, which allows them to access a broad knowledge base about various objects, people, places, and concepts. This extensive knowledge enables the LLM to consider a wide range of possibilities and ask more precise questions.
- Categorical Reasoning: LLMs can categorize objects based on features such as physical attributes, functions, or common uses. For example, if the object is a "banana," the LLM might ask questions related to its category ("Is it a fruit?"), color ("Is it yellow?"), or shape ("Is it curved?").

Learning from Interaction:

- Reinforcement Learning: LLMs can be fine-tuned using reinforcement learning, where the model learns from previous games to improve its questioning strategy. Over time, the model can become more efficient at guessing the correct object within 20 questions by refining its approach based on feedback and outcomes from past interactions.
- Error Correction: If the LLM fails to guess the object or makes an incorrect guess, it can learn from the mistake by adjusting its internal model, leading to better performance in future games.

Human-Like Interaction:

- Natural Conversation: LLMs can engage in more natural and engaging conversations, making the game feel less mechanical and more akin to playing with a human opponent. This can enhance the user experience, making the game more enjoyable and immersive.

- Creative Questioning: LLMs can introduce creative or unexpected questions that might lead to faster resolution of the game or introduce new ways of thinking about the object in question.

Challenges and Limitations:

- Handling Abstract Objects: While LLMs are strong in handling common objects, they may struggle with abstract or highly specific objects that are less represented in their training data. For example, if the object is an obscure scientific instrument, the LLM might have difficulty generating relevant questions.
- Bias in Questioning: LLMs can inadvertently introduce biases in their questioning strategy based on patterns learned during training. Ensuring that the model asks fair and unbiased questions across different categories is an ongoing challenge.
- Efficiency: While LLMs can generate high-quality questions, they might not always choose the most efficient path to the correct answer, potentially asking more questions than necessary.

Potential Use Cases:

- Educational Tools: The 20 Questions game powered by LLMs can be used as an educational tool to teach students about various topics in an interactive and engaging way.
- Entertainment: LLM-driven versions of the 20 Questions game can be integrated into virtual assistants, mobile apps, or online platforms as a fun and interactive game for users.
- Training AI Systems: The 20 Questions game can serve as a testing ground for improving AI systems' natural language understanding, reasoning, and interaction capabilities.

In summary, using LLMs for the 20 Questions game can significantly enhance the gameplay experience by providing a more flexible, intelligent, and human-like interaction. The ability of LLMs to generate adaptive questions, interpret nuanced answers, and learn from previous interactions makes them well-suited for this task, pushing the boundaries of what AI can achieve in interactive games and conversational AI. There is significant research in this field as indicated by literature review[4, 6, 9, 11, 12].

5. Implementation and Results

For the 20 Questions Game Dataset, we utilize the Kaggle Competition LLM 20 Questions [1]. The implementation will consist of one guesser LLM, responsible for asking questions and making guesses, and one answerer LLM, responsible for responding with "yes" or "no" answers. Through strategic questioning and answering, the goal is for the guesser to correctly identify the secret word in as few rounds as possible.

Using the LLM 20 Questions competition metric, the implementation is to evaluate LLMs on key skills such as deductive reasoning, efficient information gathering through targeted questioning, and collaboration between paired agents. It also presents a constrained setting requiring creativity and strategy with a limited number of guesses. Success will demonstrate

LLMs' capacity for not just answering questions, but also asking insightful questions, performing logical inference, and quickly narrowing down possibilities.

The evaluation process involves agents playing episodes (games) against other bots developed by participating teams in the competition that have a similar skill rating. Over time, skill ratings will go up with wins, down with losses, or evened out with ties. The competition is configured to run in a cooperative, 2 vs. 2 format. The bots will be randomly paired with a bot of similar skill in order to face off against another random pairing. On each pair, one bot will be randomly assigned as questioner and the other as answerer.

Each bot has an estimated skill rating which is modeled by a Gaussian $N(\mu,\sigma 2)$ where μ is the estimated skill and σ represents the uncertainty of that estimate which will decrease over time. The first play is for validation episode where that new bots plays against copies of itself to make sure it works properly. The new bot is initialized with $\mu 0=600$ and it joins the pool of bots for ongoing evaluation.

Ranking System: After an episode finishes, the rating estimate for all bots in the episode are updated. If one bot pair won, its μ will be increased and decrease the opponent's μ, if the result was a tie, then move the μ values closer towards their mean. The updates will have magnitude relative to the deviation from the expected result based on the previous μ values, and also relative to each bot's uncertainty σ.

6. Summary and Conclusion

The integration of Large Language Models (LLMs) into the 20 Questions game represents a significant advancement in AI-driven interactive experiences. Leveraging advanced natural language processing and multimodal capabilities, the game can evolve from a simple parlor activity into a sophisticated, dynamic interaction between AI and humans.

Key Enhancements for LLM:

- Adaptive Questioning: LLMs can generate contextually relevant and adaptive questions, refining its strategy based on the player's responses. This makes the gameplay more efficient and engaging.
- Nuanced Answer Interpretation: LLM excels at understanding complex, ambiguous, or nuanced answers, maintaining a coherent and natural conversation throughout the game.
- Multimodal Integration: LLM's ability to process both text and images allows for a richer gameplay experience, where players can provide visual clues that the model can incorporate into its reasoning process.
- Learning and Personalization: Over time, LLM can refine its questioning strategy based on past interactions, potentially personalizing the game to suit individual player tendencies.
- Human-Like Interaction: The model's advanced language capabilities ensure that the interaction feels more human, with natural conversation flow, creativity, and even humor.

Challenges:

- Handling Abstract Concepts: LLM may still face challenges with highly abstract or specific objects not well-represented in its training data.
- Balancing Complexity: Ensuring that the game remains accessible while leveraging advanced AI capabilities is crucial for a broad user base.

Integrating LLM into the 20 Questions game marks a significant leap forward in the evolution of AI-driven games and conversational AI. LLM's ability to blend natural language understanding with visual processing enables a more sophisticated and immersive gameplay experience, offering players a more engaging and personalized interaction. The model's adaptive questioning, nuanced understanding, and potential for learning make it an ideal candidate for transforming traditional interactive games into cutting-edge AI experiences.

However, as with any advanced AI, challenges remain, particularly in handling abstract concepts and ensuring accessibility. Despite these challenges, the potential for LLM to revolutionize the 20 Questions game—and other similar interactive AI applications—is substantial, opening new avenues for educational tools, entertainment, and beyond.

Acknowledgements

The dataset and evaluation metrics. Automated scripts for python code was provided by LLM 20 Questions Competition(Kaggle.com). We acknowledge the tools and software such as chatGPT, Google Gemma, Meta Llama, Turnitin, open Source libraries etc.

References

1. Zoe Mongan, Luke Sernau, Will Lifferth, Bovard Doerschuk-Tiberi, Ryan Holbrook, Will Cukierski, Addison Howard. (2024). LLM 20 Questions. Kaggle. https://kaggle.com/competitions/llm-20-questions(Last accessed on 8th August 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. Yizhe Zhang, Jiarui Lu, Navdeep Jaitly, Probing the Multi-turn Planning Capabilities of LLMs via 20 Question Games. https://arxiv.org/abs/2310.01468v3 (Last accessed on 8th August 2024)
5. 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)
6. Peter Clark, et al., Think you have Solved Question Answering? Try ARC, the AI2 Reasoning Challenge, https://arxiv.org/abs/1803.05457 (Last accessed on 8th August 2024)
7. 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)
8. 20 Questions Wiki: https://en.wikipedia.org/wiki/Twenty_questions (Last accessed on 8th August 2024)

9. Noever, D. & McKee, Forrest. (2022). Chatbots as Problem Solvers: Playing Twenty Questions with Role Reversals. 10.48550/arXiv.2301.01743. (Last accessed on 8th August 2024)
10. Y. Chen et al., Learning-to-Ask: Knowledge Acquisition via 20 Questions, https://arxiv.org/abs/1806.08554 (Last accessed on 8th August 2024)
11. H. Hu et al., Playing 20 Question Game with Policy-Based Reinforcement Learning, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3233–3242, Brussels, Belgium, October 31 - November 4, 2018.
12. A. Dey et al., All It Takes is 20 Questions!: A Knowledge Graph Based Approach, https://arxiv.org/abs/1911.05161 (Last accessed on 8th August 2024)
13. 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)
14. 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)

Biography

Dr. CRS Kumar is currently Professor in the School of Computer Engineering & Mathematical Sciences, 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 recognized Machine Learning and Data Science Competition Expert by Kaggle.com(Google Subsidiary). He has won many medals in the Global ML Competitions. He has also earned AI, ML and Data Science Certificates from Massachusetts Institute of Technology(MIT) and University of Texas at Austin.

Revision History:

-ver1.0, 9th August 2024, CRSK