

Article

Not peer-reviewed version

Road to AGI: The Importance of Clear Objectives and Reward Functions

[Mohammad Rasoolinejad](#) *

Posted Date: 9 July 2024

doi: 10.20944/preprints202407.0687.v1

Keywords: Artificial Intelligence; Artificial General Intelligence; Reinforcement Learning; Machine learning; Neural network



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Road to AGI: The Importance of Clear Objectives and Reward Functions

Mohammad Rasoolinejad ^{1,2}

¹ Department of Neurosurgery, David Geffen School of Medicine, University of California, Los Angeles, CA; Rasoolinejad@u.northwestern.edu

² Ph.D. Graduate, Robert R. McCormick School of Engineering and Applied Science, Northwestern University, Evanston, IL

Abstract: Training artificial intelligence (AI) for reasoning presents significant challenges due to the lack of clear objectives and reward functions. In structured environments such as chess and Go, objectives are well-defined, and the optimal moves can be determined through search algorithms that maximize the reward function. This clarity enables AI to learn and improve autonomously through random trials and recursive calculation of rewards, mirroring human learning processes that depend on prior experiences and state-dependent actions. However, in real-world applications, AI lacks access to direct interaction and feedback, making it difficult to define proper reward functions. This limitation is particularly evident in the training of large language models (LLMs), where the quality of the output is constrained by the training data and the absence of a dynamic reward system. Simulated environments offer some utility but are inherently limited by their design and scope. Achieving artificial general intelligence (AGI) requires AI to function and receive feedback in real-world settings, similar to human cognitive and strategic development through interactive experiences. This paper explores the critical role of clear objectives and reward functions in AI training, the limitations posed by the current lack of real-world interaction, and the implications for future advancements in AI reasoning capabilities.

Keywords: artificial intelligence; artificial general intelligence; reinforcement learning; machine learning; neural network

Introduction

Artificial General Intelligence (AGI) represents a pivotal goal in the field of artificial intelligence research, characterized by the development of machines that possess the ability to understand, learn, and apply knowledge across a wide range of tasks at a level comparable to human intelligence. Unlike narrow AI, which is designed to excel in specific areas, AGI aims to create systems capable of performing any intellectual task that a human can, demonstrating versatility, adaptability, and a deep understanding of complex concepts. The significance of AGI lies in its potential to revolutionize numerous fields, from healthcare and education to robotics and beyond, by enabling machines to think and reason in a manner akin to humans.

The development of artificial intelligence (AI) has made significant strides in recent years, particularly in areas such as image recognition, natural language processing, and game playing. However, training AI to perform reasoning tasks remains a formidable challenge. This difficulty largely stems from the lack of clear objectives and reward functions, which are essential for guiding the learning process. In structured environments like chess and Go, the objectives are well-defined, and optimal moves can be identified through search algorithms designed to maximize specific goals. These games provide a clear and measurable reward: winning. This allows AI systems to learn and improve autonomously by conducting random trials, continually searching for the optimal path, and recursively calculating the reward function. As a result, AI can refine its actions by discovering the maximum reward path through exhaustive search and using each stage of the best path to train on the optimal next action.

In contrast, many real-world tasks lack this level of clarity. In these scenarios, defining objectives and reward functions becomes significantly more complex due to the inherent abstractions and uncertainties. Human reasoning and strategy development are shaped through a lifetime of trial and error, receiving constant feedback from the environment. Similarly, AI has the potential to

learn through random trials and measuring the environment's response, but the current limitation is the absence of real-world interaction and feedback for AI systems. Large language models (LLMs) highlight this issue. During training, these models rely on data states fed to them with objectives set by human writers. Consequently, the quality of LLM outputs is constrained by the training data and the lack of a dynamic reward system. Improvement and reasoning in AI require trial and error, searching for optimal paths, and defining reward functions, which is nearly impossible in real-world contexts.

This paper explores the critical role of clear objectives and reward functions in AI training, the limitations posed by the current lack of real-world interaction, and the implications for future advancements in AI reasoning capabilities. By examining these challenges, we aim to shed light on the path forward for developing AI systems capable of more advanced reasoning and decision-making akin to human intelligence.

Literature Review

Reinforcement learning (RL) has emerged as a critical area of study within artificial intelligence, focusing on how agents can learn optimal behaviors through interactions with their environment. In recent years, deep reinforcement learning (DRL) has gained significant attention due to its ability to handle high-dimensional state spaces and complex tasks. LeCun et al. (2015) describe the rise of deep learning and its integration with RL, leading to breakthroughs in areas like game playing and robotic control. The introduction of deep neural networks into RL has enabled the development of more sophisticated models that can learn from raw sensory inputs and achieve human-level performance in various tasks [1]. Li (2017) provides an overview of deep reinforcement learning, highlighting the advancements and challenges in the field. This work emphasizes the importance of neural network architectures, the role of experience replay, and the development of efficient algorithms for training deep RL models. The survey also discusses the limitations of current DRL approaches, including the instability of training and the sensitivity to hyperparameters [2].

François-Lavet et al. (2018) further explore the intricacies of deep reinforcement learning, providing a detailed introduction to the field and discussing the various methods and techniques used to train DRL models. This comprehensive review covers topics such as policy gradient methods, value-based methods, and model-based approaches, offering insights into the strengths and weaknesses of each [3]. Henderson et al. (2018) address the reproducibility issues in deep reinforcement learning, emphasizing the importance of standardized evaluation protocols and reporting practices. The paper highlights the variability in reported results due to differences in implementation, hyperparameter settings, and experimental conditions. This work underscores the need for rigorous and transparent research practices to ensure the reliability and validity of DRL research [4].

The real-world application of reinforcement learning presents unique challenges that differ significantly from those encountered in simulated environments. Dulac-Arnold et al. (2019) discuss the challenges of real-world reinforcement learning, including the difficulties in obtaining accurate and timely feedback, the need for safe exploration, and the handling of sparse and delayed rewards. This paper provides a framework for addressing these challenges and highlights the importance of developing algorithms that can operate effectively in real-world settings [5]. One of the promising approaches to bridge the gap between simulation and real-world application is sim-to-real transfer. Zhao et al. (2020) provide a survey on sim-to-real transfer in deep reinforcement learning for robotics, discussing the techniques used to transfer knowledge acquired in simulated environments to real-world tasks. This survey highlights the importance of domain adaptation, transfer learning, and the development of robust policies that can generalize across different environments [6].

The advancement of search algorithms, deep learning techniques, and the development of AlphaGo represent significant milestones in artificial intelligence. Search algorithms have long been a cornerstone of artificial intelligence, providing methods for problem-solving and decision-making in complex environments. Monte Carlo Tree Search (MCTS) has been particularly influential in game AI. Chaslot et al. (2008) introduced MCTS as a novel framework for game AI, combining tree search

algorithms with Monte Carlo simulations to handle vast search spaces. This approach allows for efficient exploration and exploitation of potential moves, making it highly effective for games with high branching factors [7]. Gelly and Silver (2011) extended the application of MCTS to the game of Go, integrating rapid action value estimation to enhance the algorithm's performance. Their work demonstrated that MCTS could rival human expertise by efficiently evaluating potential moves and learning optimal strategies [8]. The broader implications of MCTS and its extensions were further explored by Gelly et al. (2012), who discussed the grand challenges of computer Go and how MCTS could address these challenges through adaptive playout policies and parallelization techniques [9].

The development of AlphaGo marked a significant breakthrough in AI, showcasing the power of combining deep neural networks with tree search algorithms. Silver et al. (2016) described the architecture of AlphaGo, which integrates deep convolutional neural networks with MCTS to master the game of Go. This combination enabled AlphaGo to evaluate positions and simulate potential moves more effectively than previous approaches, leading to its victory over human champions [10]. In a subsequent study, Silver et al. (2017) presented AlphaGo Zero, an advanced version of AlphaGo that learned to play Go without human knowledge. By relying solely on self-play and reinforcement learning, AlphaGo Zero surpassed its predecessor, demonstrating the potential of deep learning and tree search to achieve superhuman performance without human input [11]. The integration of deep learning with tree search has also been explored in other contexts. Anthony et al. (2017) proposed a model that combines deep learning with MCTS to mimic human cognitive processes of thinking fast and slow. This model leverages the rapid evaluation capabilities of deep learning and the thorough search capabilities of MCTS, achieving state-of-the-art performance in various decision-making tasks [12]. The development of MCTS has revolutionized game AI, enabling efficient decision-making in complex environments. The success of AlphaGo demonstrates the power of combining deep neural networks with tree search algorithms to achieve superhuman performance. Attention mechanisms, as introduced by the Transformer model, have further enhanced AI capabilities, enabling more efficient processing and prioritization of information. These innovations collectively contribute to the ongoing progress toward achieving more advanced and versatile AI systems.

Attention mechanisms have emerged as a crucial innovation in deep learning, significantly enhancing the ability of AI systems to process and prioritize information. Vaswani et al. (2017) introduced the Transformer model, which relies entirely on attention mechanisms to handle sequential data. The model's architecture, which eschews recurrent layers in favor of self-attention, allows for efficient parallelization and improved performance on tasks such as language translation and text generation [13]. The concept of attention has since been widely adopted in various AI applications, contributing to substantial improvements in natural language processing and other fields.

Literature Review on Large Language Models

The introduction of BERT and the Transformer architecture has significantly advanced the field of natural language processing (NLP). BERT (Bidirectional Encoder Representations from Transformers) marked a substantial breakthrough in NLP by enabling the pre-training of deep bidirectional representations from unlabeled text. Devlin et al. (2018) introduced BERT, demonstrating its ability to achieve state-of-the-art performance across a wide range of NLP tasks. BERT's architecture relies on Transformers, specifically utilizing multi-layer bidirectional Transformer encoders to capture context from both directions in a sentence. This approach allows BERT to understand the meaning of words in context, leading to improved performance in tasks such as question answering, sentiment analysis, and named entity recognition [14].

The effectiveness of BERT has been extensively studied and visualized to understand its performance better. Hao et al. (2019) explored the loss landscapes and optimization trajectories of fine-tuning BERT on specific datasets. Their findings indicated that BERT's pre-training reaches a good initial point across downstream tasks, leading to wider optima and easier optimization compared to training from scratch. The visualization results also showed that BERT's fine-tuning procedure is robust to

overfitting, which contributes to its strong generalization capabilities [15]. Koroteev (2021) provided a comprehensive review of BERT's applications in natural language processing and understanding. This review highlighted BERT's versatility and effectiveness across various NLP tasks, emphasizing its ability to capture complex linguistic patterns and semantic relationships. The study also discussed the ongoing challenges and potential improvements in leveraging BERT for more advanced NLP applications [16].

The underlying Transformer model, introduced by Vaswani et al. (2017), has been a pivotal innovation in deep learning. Transformers use self-attention mechanisms to process input sequences in parallel, rather than sequentially, which allows for more efficient handling of long-range dependencies in text. This architecture forms the backbone of BERT and other advanced language models. The Transformer model's ability to scale and its success in various tasks, such as machine translation and text generation, underscore its significance in AI research [13]. Lin et al. (2022) conducted a survey on Transformers, discussing their applications and advancements beyond NLP. The study explored how Transformers have been adapted for tasks in computer vision, speech processing, and other domains, highlighting their flexibility and generalizability. The survey also addressed the challenges in training large-scale Transformer models and the ongoing research to optimize their performance and efficiency [17].

Bengesi et al. (2024) provided a comprehensive review of advancements in generative AI, including GANs, GPT, autoencoders, diffusion models, and Transformers. This review highlighted the transformative impact of these technologies on AI capabilities, particularly focusing on the role of Transformers in enabling sophisticated generative tasks. The study also discussed the future directions in generative AI research, emphasizing the importance of continued innovation in model architectures and training techniques [18]. The literature on BERT and Transformers underscores their profound impact on advancing AI capabilities in natural language processing and beyond. BERT's bidirectional context representation, enabled by the Transformer architecture, has set new benchmarks in NLP tasks. The ongoing research and visualization studies continue to enhance our understanding of these models, paving the way for future innovations in AI.

The development and deployment of Large Language Models (LLMs) have brought transformative changes to various fields, particularly in natural language processing (NLP) and artificial intelligence (AI). Kasneci et al. (2023) explore the opportunities and challenges posed by LLMs like ChatGPT in the field of education. Their study highlights how LLMs can enhance personalized learning experiences, provide instant feedback, and support educators by automating administrative tasks. However, they also caution against potential issues such as the accuracy of generated content, the need for human oversight, and the ethical implications of deploying AI in educational settings [19].

The comprehensive survey by Chang et al. (2024) offers an in-depth evaluation of LLMs, examining various metrics and methodologies used to assess their performance. This work underscores the importance of standardized benchmarks and the need for robust evaluation frameworks to ensure that LLMs are reliable and effective across different applications. The survey also points out gaps in current evaluation practices, suggesting areas for future research [20]. Kirchenbauer et al. (2023) present an innovative approach to enhancing the transparency and accountability of LLMs through the use of watermarks. They propose a framework for embedding watermarks into the outputs of LLMs, which can help trace the origins of generated content and detect unauthorized use. This method aims to address the growing concerns about the misuse of AI-generated text and the proliferation of deepfakes [21].

Zhao et al. (2023) provide a broad survey of LLMs, discussing their architectural innovations, training methodologies, and the diverse range of applications they support. The authors emphasize the significance of LLMs in advancing the state-of-the-art in NLP and their potential to revolutionize fields such as healthcare, finance, and customer service. They also highlight the challenges related to scalability, ethical considerations, and the environmental impact of training large models [22]. Xu et al. (2022) conduct a systematic evaluation of LLMs specifically designed for coding tasks. Their

study compares several models, including Codex, GPT-J, GPT-Neo, GPT-NeoX-20B, and CodeParrot, across various programming languages. The evaluation reveals that while open-source models like PolyCoder show promising results, there is still a significant performance gap compared to proprietary models such as Codex. The authors advocate for more open-source initiatives to democratize access to powerful LLMs in the coding domain [23].

Wei et al. (2022) delve into the emergent abilities of LLMs, exploring how these models can exhibit behaviors and capabilities that were not explicitly programmed. Their research investigates phenomena such as in-context learning, zero-shot performance, and the ability of LLMs to generalize across tasks and domains. The findings suggest that as LLMs grow in scale and complexity, they begin to demonstrate unexpected and sophisticated behaviors, raising both exciting possibilities and new challenges for AI research [24]. The literature on LLMs highlights their transformative potential and applications they support. From enhancing education and healthcare to advancing software development and customer service, LLMs are proving to be invaluable tools in various domains. However, the challenges related to evaluation, transparency, ethical considerations, and scalability remain significant. Continued research and collaboration are essential to address these issues and harness the full potential of LLMs in a responsible and sustainable manner.

The Role of Objectives and Reward Functions in AI

Objectives and reward functions are fundamental components in the development and training of artificial intelligence (AI). Objectives provide a clear goal for the AI to achieve, while reward functions offer a measurable way to evaluate the AI's progress toward that goal. These elements are crucial because they guide the learning process, helping the AI to understand what constitutes success and how to adjust its actions accordingly to maximize this success. In structured environments such as games, the importance of well-defined objectives and reward functions becomes readily apparent. Take, for example, games like chess and Go. In these games, the objective is explicitly clear: to win. This clarity allows for the creation of precise reward functions that can quantify success in a straightforward manner, such as gaining points, capturing pieces, or achieving checkmate. The reward functions in these games can be meticulously designed to guide the AI's learning process, enabling it to evaluate the outcomes of its moves, learn from them, and improve over time.

The structured nature of these environments facilitates the use of search algorithms to identify optimal moves. These algorithms work by exploring possible moves and their outcomes, calculating the rewards for each potential path. Through exhaustive search and recursive calculation of the reward function, the AI can determine the path that maximizes the reward, thereby discovering the optimal strategy to win the game. This process of trial and error, guided by clear objectives and reward functions, allows the AI to refine its actions and continuously improve its performance. In contrast, real-world tasks often lack the clarity and structure found in games. Defining objectives and reward functions in these scenarios is significantly more complex due to the inherent abstractions and uncertainties. Real-world environments are dynamic and multifaceted, with objectives that may not always be clearly defined or easily measurable. For instance, an AI system designed to navigate a social situation or make ethical decisions faces a multitude of variables and potential outcomes, many of which cannot be neatly quantified or predicted.

The ambiguity in real-world tasks poses a significant challenge for AI development. Without clear objectives, it becomes difficult for the AI to understand what it is supposed to achieve. Similarly, without well-defined reward functions, the AI cannot accurately measure its progress or learn from its actions. This lack of clarity impedes the AI's ability to refine its behavior through trial and error, as it cannot reliably assess the outcomes of its actions or determine which paths are most beneficial. Moreover, the complexity of real-world environments means that AI systems must contend with a broader range of variables and unpredictable interactions. Unlike games, where the rules are fixed and the environment is controlled, real-world scenarios are influenced by numerous external factors that can change unpredictably. This variability makes it challenging to design reward functions that can

consistently and accurately evaluate the AI's performance. Consequently, AI systems may struggle to adapt their behavior and improve their reasoning capabilities in these less structured environments.

Learning through Trial and Error

Human learning processes are inherently grounded in trial and error, a method through which individuals continuously refine their understanding and capabilities by interacting with their environment and receiving feedback. From early childhood, humans engage in exploratory behaviors, testing hypotheses about their surroundings, and adjusting their actions based on the outcomes they observe. This iterative process of trying different approaches, observing the results, and learning from successes and failures is fundamental to cognitive development and the acquisition of complex skills. Feedback mechanisms, both intrinsic and extrinsic, play a crucial role in this process. Intrinsic feedback comes from the internal recognition of errors and successes, while extrinsic feedback is provided by the environment, including social cues from other individuals, direct consequences of actions, and structured learning environments such as educational institutions.

Artificial intelligence (AI) systems also have the potential to learn through trial and error, using the environment's response to guide their development. This method, known as reinforcement learning, involves training an AI agent to make decisions by rewarding it for successful actions and penalizing it for unsuccessful ones. Over time, the AI agent learns to maximize its cumulative reward by refining its decision-making process. Random trials, in which the AI explores various actions without prior knowledge of their outcomes, are a critical aspect of this learning process. By exploring a wide range of possible actions and receiving feedback from the environment, the AI can develop strategies that lead to successful outcomes. However, current AI systems face significant limitations in accessing real-world feedback, which hampers their ability to learn effectively through trial and error. One major limitation is the lack of direct interaction with the real world. Most AI training occurs in simulated environments, which, although useful, cannot fully replicate the complexity and unpredictability of the real world. Simulated environments are designed with specific parameters and constraints, limiting the range of experiences and interactions an AI can have. As a result, the feedback AI receives in these environments is often less varied and complex compared to real-world feedback, constraining the depth and breadth of learning.

Another limitation is the ethical and practical concerns associated with allowing AI systems to engage in trial and error in real-world settings. For instance, in applications such as autonomous driving or healthcare, the consequences of errors can be severe, posing risks to human safety and well-being. These concerns necessitate stringent controls and safeguards, restricting the extent to which AI can experiment and learn from real-world interactions. Consequently, AI systems are often trained in highly controlled environments where the scope for trial and error is limited, impeding the development of robust and adaptable reasoning capabilities. Moreover, the complexity and variability of real-world environments introduce challenges in defining appropriate reward functions. Unlike games or controlled simulations, real-world scenarios often involve multiple, overlapping objectives and outcomes that are difficult to quantify. This makes it challenging to design reward functions that accurately reflect the desired goals and provide meaningful feedback to the AI. Without clear and consistent reward signals, AI systems struggle to learn effectively from their interactions with the environment. The limitations of current AI in accessing real-world feedback underscore the need for innovative approaches to training. To bridge the gap between simulated and real-world learning, researchers are exploring hybrid methods that combine the strengths of both. These approaches involve training AI in increasingly complex and realistic simulations while gradually introducing controlled real-world interactions. By leveraging advances in simulation technology and carefully managing real-world exposure, it may be possible to provide AI systems with richer and more varied feedback, enhancing their ability to learn through trial and error.

The Role of Games in Cognitive Development

Games have long been recognized as powerful tools for human cognitive and strategic development. From early childhood, games provide an engaging and interactive means for individuals to develop critical thinking, problem-solving skills, and strategic planning abilities. Games stimulate the mind by presenting challenges that require players to think ahead, adapt to changing conditions, and learn from their mistakes. This dynamic form of learning is instrumental in honing cognitive abilities, fostering creativity, and enhancing memory and concentration. The strategic nature of many games also plays a crucial role in developing advanced reasoning skills. Games such as chess, Go, and various strategy-based video games require players to formulate plans, anticipate opponents' moves, and make decisions under pressure. These activities mirror real-world scenarios where strategic thinking and quick decision-making are essential. Through repeated gameplay, individuals learn to recognize patterns, devise effective strategies, and improve their ability to think several steps ahead, all of which are valuable skills in both personal and professional contexts.

The application of game-based learning in artificial intelligence (AI) has leveraged these insights into human cognitive development. Games provide a structured yet complex environment where AI can be trained to develop similar cognitive and strategic skills. By engaging in game-based learning, AI systems can explore a variety of scenarios, receive immediate feedback, and refine their strategies over time. This approach has been particularly successful in training AI to master complex games like chess and Go, where clear objectives and well-defined rules allow for precise measurement of success and continuous improvement. Game-based learning allows AI to experiment with different strategies in a controlled setting, learning from successes and failures without real-world consequences. This iterative process enables AI to develop robust decision-making capabilities and adaptability. For example, the development of AlphaGo, an AI program that defeated human champions in the game of Go, demonstrated the potential of game-based learning to push the boundaries of AI performance. AlphaGo's success was attributed to its ability to learn from millions of game simulations, refining its strategies through deep reinforcement learning and neural networks.

Despite the successes of game-based learning in AI, significant challenges arise when translating these skills to real-world tasks. The structured nature of games, with their clear rules and objectives, contrasts sharply with the ambiguity and complexity of real-world environments. Real-world tasks often involve multiple overlapping objectives, unpredictable variables, and the need for contextual understanding that goes beyond the scope of most games. This makes it difficult for AI systems trained primarily in game environments to generalize their learning to real-world applications. One major challenge is the difference in the feedback mechanisms. In games, feedback is immediate and unambiguous, allowing AI to quickly learn the consequences of its actions. In the real world, feedback can be delayed, indirect, or influenced by a multitude of factors, making it harder for AI to discern the impact of its decisions. Additionally, real-world tasks often require a deeper understanding of context and the ability to interpret complex information, skills that are not easily developed through game-based learning alone.

Another challenge is the scalability of strategies learned in games to the broader, more variable conditions of real-world environments. Strategies that are effective in the constrained and predictable settings of games may not be directly applicable to the dynamic and often chaotic nature of real-world scenarios. This limitation underscores the need for AI to be exposed to a wider range of experiences and learning environments to develop truly adaptable and generalizable skills. Addressing these challenges requires innovative approaches that combine the strengths of game-based learning with real-world interactions to create AI systems capable of robust and adaptable performance across a wide range of scenarios.

Training Large Language Models (LLMs)

Large Language Models (LLMs) represent a significant advancement in the field of artificial intelligence, particularly in natural language processing. These models, such as GPT-3 and BERT, are

trained on vast corpora of text data, enabling them to generate human-like text, understand context, and perform a variety of language-related tasks. The training process for LLMs involves feeding the model with extensive datasets containing diverse examples of language usage. The model learns to predict the next word in a sentence, identify relationships between words, and understand syntactic and semantic intricacies through this exposure. This process relies heavily on supervised learning, where the model is trained on pre-labeled data, and unsupervised learning, where the model identifies patterns within the data without explicit labels.

Despite the impressive capabilities of LLMs, their training processes are constrained by the quality and scope of the training data. The datasets used to train these models are curated from a wide range of sources, including books, websites, and academic papers. However, the quality of the data can vary significantly, and biases present in the training data can be inadvertently learned by the models. Additionally, because the training data is static, the models do not have the ability to dynamically update their knowledge base with new information post-training. This static nature of training data limits the ability of LLMs to adapt to new contexts or learn from real-time interactions. Another significant constraint in the training of LLMs is the lack of dynamic reward systems. In reinforcement learning, dynamic reward systems provide continuous feedback to guide the learning process, allowing the model to adjust its behavior based on the outcomes of its actions. However, LLMs typically do not utilize such systems. Instead, they rely on pre-defined training objectives, such as minimizing prediction errors or maximizing the likelihood of generating grammatically correct sentences. While these objectives can produce models that are proficient at certain tasks, they do not provide the context-sensitive feedback necessary for advanced reasoning and adaptation.

The impact of these constraints on the quality of AI reasoning and output is profound. Because LLMs are trained on static datasets with fixed objectives, their ability to reason, adapt, and improve over time is inherently limited. The reasoning capabilities of these models are confined to the patterns and knowledge present in their training data. Consequently, they may generate responses that are contextually appropriate but lack deeper understanding or fail to address complex, nuanced questions effectively. Furthermore, the inability to incorporate real-time feedback and learning means that LLMs cannot improve their performance based on user interactions or new information, leading to a static performance level that does not evolve post-training. The lack of dynamic reward systems also means that LLMs cannot effectively evaluate the quality of their outputs in real-time. Without the ability to receive and integrate feedback, these models cannot refine their responses to better meet user needs or correct mistakes based on past interactions. This limitation restricts the potential for continuous improvement and limits the applicability of LLMs in dynamic, real-world scenarios where adaptability and ongoing learning are crucial.

The Necessity of Objectives and Goals in AI Reasoning

Reasoning, in the context of artificial intelligence (AI), is a process that involves making decisions, drawing inferences, and solving problems based on available information. However, this process becomes meaningless in the absence of clearly defined objectives and goals. Objectives provide the framework within which reasoning occurs, offering a purpose and direction for the AI's decision-making processes. Without a clear objective, the actions taken by an AI system lack coherence and purpose, rendering its reasoning capabilities ineffective and arbitrary. Although it is not always necessary for an intelligent agent to search for the optimal path in every state, having clear, objective-oriented next best action moves can significantly enhance its ability to achieve locally optimal decisions. This approach allows the AI to focus on making the best possible decision at each step, guided by an overarching objective that aligns its actions with the ultimate goal. By consistently taking the next best action, the AI can navigate complex environments more effectively, adapting to changing circumstances and refining its strategies in real time.

Training AI to focus on the next best action towards a goal involves providing it with a clear understanding of how individual actions contribute to achieving the ultimate objective. This method

not only simplifies the decision-making process but also improves the AI's reasoning and strategic planning capabilities. By breaking down the path to the ultimate goal into a series of manageable, goal-oriented steps, the AI can develop a more structured and purposeful approach to problem-solving. To illustrate, consider an AI system designed to navigate a complex maze. Instead of searching for the optimal path from the start to the finish, the AI can be trained to focus on the next best action at each intersection, guided by the objective of reaching the maze's exit. By evaluating the immediate outcomes of each possible move and selecting the one that best aligns with the goal of progressing through the maze, the AI can make incremental progress toward the ultimate objective. Over time, this approach allows the AI to develop a comprehensive understanding of the maze's structure, improving its ability to navigate similar challenges in the future. This objective-oriented training enhances the AI's ability to adapt to new and unforeseen situations. By learning to evaluate and select the next best action in various contexts, the AI becomes more versatile and capable of handling a broader range of tasks. This adaptability is crucial for developing robust reasoning and decision-making skills, as it enables the AI to generalize its learning to new domains and apply its knowledge in diverse scenarios.

The emphasis on clear objectives and next best actions also facilitates the integration of dynamic reward mechanisms. By providing immediate feedback on the outcomes of each action, these mechanisms help the AI to refine its strategies and improve its performance over time. This iterative process of action, feedback, and adjustment is fundamental to developing advanced reasoning capabilities, as it allows the AI to learn from its experiences and continuously optimize its decision-making processes. Reasoning in AI is inextricably linked to the presence of clear objectives and goals. Without these guiding principles, the actions taken by an AI system lack coherence and purpose, undermining its ability to make meaningful decisions. By training AI to focus on objective-oriented next best actions and explaining how these actions contribute to achieving the ultimate goal, we can significantly enhance its reasoning and decision-making capabilities. This approach not only simplifies the AI's decision-making process but also fosters adaptability and continuous improvement, laying the foundation for more advanced and versatile AI systems. Moreover, incorporating chain-of-thought prompting into AI models can further enhance their reasoning capabilities. As shown by Wei et al. (2022), chain-of-thought prompting enables large language models to perform complex reasoning tasks by generating intermediate reasoning steps, thereby improving the AI's ability to tackle arithmetic, commonsense, and symbolic reasoning problems [25].

Challenges in Mathematical Reasoning

Mathematical reasoning presents a unique set of challenges for artificial intelligence (AI) due to its inherent complexity and abstract nature. Unlike physical or language tasks, mathematical reasoning involves operating within highly abstract spaces where trial and error processes are not as straightforward. The abstract nature of mathematics means that reasoning often involves manipulating symbols, following formal rules, and constructing logical proofs, all of which require a deep understanding of underlying principles that are not directly observable in the physical world. The complexity of trial and error in abstract mathematical spaces stems from the need to explore numerous potential solutions without direct feedback from a physical environment. In traditional AI applications, feedback is typically derived from environmental responses to actions, allowing the AI to adjust and refine its strategies. However, in mathematical reasoning, the "environment" is an abstract conceptual space where feedback must be derived from logical consistency and adherence to mathematical rules rather than physical outcomes. This makes the process of learning through trial and error significantly more challenging, as the AI must rely on internal validation mechanisms to gauge the correctness of its reasoning.

Projecting abstract mathematical results to real-world scenarios adds another layer of difficulty. In many cases, mathematical reasoning is used to model and predict phenomena in the physical world. This requires not only solving abstract mathematical problems but also translating these solutions into meaningful real-world applications. For example, solving a complex differential equation may yield

a theoretical result, but applying this result to predict physical behavior in engineering or physics requires understanding of both the mathematical model and the real-world system it represents. AI currently struggles with this type of reasoning because it lacks the ability to interact with and observe real-world systems directly, limiting its capacity to bridge the gap between abstract mathematical results and practical applications.

Current limitations of AI in mathematical reasoning are evident in several key areas. First, AI lacks the intuitive understanding that humans often develop through years of education and experience. Human mathematicians draw upon a wealth of tacit knowledge and heuristic techniques that are difficult to encode in an AI system. This includes the ability to recognize patterns, make educated guesses, and apply creative problem-solving strategies that go beyond formal algorithmic processes. Second, the formalism required in mathematical reasoning poses a significant challenge. AI systems must follow strict logical rules and ensure the consistency and correctness of their solutions. Any deviation from these rules can lead to incorrect conclusions, and the abstract nature of mathematics means that errors are not always immediately apparent. This requires rigorous validation and verification processes, which are challenging to automate and integrate into AI systems.

Moreover, the lack of real-world interaction limits AI's ability to develop a practical understanding of how mathematical models relate to physical phenomena. Human mathematicians often use intuition and empirical observations to guide their reasoning and validate their models. In contrast, AI systems are confined to the data and formal rules they have been trained on, which can limit their ability to generalize and apply their reasoning to new, unseen problems. The challenges in mathematical reasoning for AI are multifaceted and deeply rooted in the abstract and formal nature of mathematics. The complexity of trial and error in abstract spaces, the difficulties in projecting mathematical results to real-world scenarios, and the current limitations of AI systems in developing intuitive and heuristic understanding all contribute to the hurdles faced in this domain. Overcoming these challenges will require advancements in both the theoretical foundations of AI and its practical capabilities in interacting with and learning from the real world.

Simulated Environments versus Real-World Interaction

Simulated environments have become a cornerstone in the training and development of artificial intelligence (AI), providing a controlled setting where AI systems can safely and efficiently learn and refine their capabilities. These environments offer numerous benefits, including the ability to conduct extensive testing without the risks associated with real-world trials. For instance, in the context of autonomous driving, simulations allow AI models to encounter and navigate a wide range of scenarios, from common traffic situations to rare and hazardous conditions, all within a virtual space. This capability enables rapid iteration and experimentation, fostering the development of robust AI systems. However, the utility of simulated environments comes with inherent limitations. One major drawback is that these simulations are constrained by the parameters set by their designers. This means that the range of scenarios an AI can experience is limited to what has been anticipated and programmed into the simulation. Consequently, AI systems trained exclusively in simulated environments may struggle to handle unexpected or novel situations that fall outside the scope of the simulation. This lack of exposure to the full diversity and unpredictability of the real world can result in AI that is less adaptable and resilient when deployed in real-world settings.

The importance of real-world trials and feedback cannot be overstated in the pursuit of advanced AI capabilities, including the aspiration of reaching artificial general intelligence (AGI). Real-world interaction provides AI systems with rich, dynamic feedback that is crucial for developing robust reasoning and decision-making skills. Unlike simulated environments, the real world is characterized by its complexity, variability, and the presence of unforeseen circumstances. This exposure is essential for AI systems to learn how to adapt, generalize, and make sound decisions in a wide array of contexts. Real-world feedback is also vital for improving the accuracy and effectiveness of AI systems. In practical applications, the intricacies of real-world data and interactions can reveal limitations and

flaws in AI models that simulations might overlook. By continuously interacting with the real world, AI systems can receive immediate, context-specific feedback that informs and refines their learning process. This iterative cycle of action, feedback, and adjustment is fundamental to the development of AI that can perform reliably and intelligently in diverse and unpredictable environments. While simulated environments provide a valuable foundation for initial training and development, they are not sufficient on their own to achieve the level of sophistication required for AGI. The limitations of simulations, particularly their inability to capture the full breadth of real-world variability and complexity, mean that AI systems must be exposed to real-world interactions to truly understand and navigate their environments. This exposure is necessary for developing the kind of adaptive, generalizable intelligence that characterizes human cognition and is the hallmark of AGI.

Towards Artificial General Intelligence (AGI)

Achieving AGI necessitates several fundamental requirements, the foremost being real-world interaction. For an AI system to develop comprehensive understanding and robust reasoning capabilities, it must engage with real environments, experiencing diverse scenarios and receiving dynamic feedback. This interaction is crucial as it allows the AI to learn from the complexities and unpredictability of real-world settings, refining its decision-making processes based on tangible outcomes. By performing tasks in real environments and utilizing dynamic reward mechanisms, AI can train itself based on the feedback received, rewarding states that contribute to paths of maximum reward. This iterative process enables the AI to build sophisticated reasoning abilities, considering end goals and planning several steps ahead.

The problem in training AGI for reasoning largely stems from the lack of clear objectives and reward functions. Unlike games such as chess and Go, where the goal is well-defined and the optimal move can be determined to maximize this objective through search algorithms, many real-world tasks lack this clarity. In games, the objective is to win, and the reward is the clear outcome of the game, making it easy to measure success. This allows AI to learn and improve autonomously through random trials, continually searching for the optimal path and recursively calculating the reward function. In such structured environments, AI can use these reward functions to refine its actions. By discovering the maximum reward path through exhaustive search, each stage of the best path can be used to train the AI on the optimal next action. This mirrors how humans learn; they make the next best action based on prior experiences and feedback rather than constantly foreseeing every consequence. As a result, actions become state-dependent, rooted in learned experiences rather than continuous search. Humans develop reasoning and strategy through a lifetime of trial and error, receiving constant feedback from their environment. This iterative process shapes their ability to take optimal actions in various contexts. Similarly, AI has the potential to learn through random trials and measuring the environment's response. However, the current limitation is that AI lacks access to the real world for direct interaction and feedback.

When training large language models (LLMs), for instance, the training data states are fed to the model with objectives set by human writers. Consequently, the quality of the LLM's output is constrained by the quality of its training data, and it cannot improve upon this without a proper reward function. Improvement and reasoning in AI require trial and error, searching for optimal paths, but defining a reward function in real-world scenarios is nearly impossible due to the inherent abstractions and uncertainties. Robotics plays a pivotal role in the pursuit of AGI. It provides a tangible platform for AI to interact with the physical world, integrating sensory inputs, motor actions, and cognitive processing. Through robotic systems, AI can perform complex tasks requiring coordination, dexterity, and real-time decision-making. Interaction with the environment is essential for developing adaptive and generalizable intelligence characteristic of AGI. By embedding AI in robots capable of navigating and manipulating the world, researchers can study how machines learn from direct experience, adapt to new situations, and develop long-term planning and reasoning skills.

Achieving artificial general intelligence (AGI) hinges on a paradigm shift from training AI models for every possible state to focusing on taking the next best action to achieve an ultimate goal. This approach recognizes that the complexity and unpredictability of real-world scenarios make it impractical to prepare an AI for every conceivable situation. Instead, AGI is cultivated through a process of continual learning and adaptation, where the system constantly evaluates its current state and determines the optimal action to move closer to its overarching objective. In this framework, the AI system is endowed with a set of high-level goals and a robust mechanism for assessing its progress towards these goals. Rather than being pre-programmed with specific responses to specific inputs, the system dynamically generates actions based on its current understanding of the environment and its end goals. This requires a sophisticated decision-making process that can weigh various potential actions, predict their outcomes, and select the one most likely to advance towards the ultimate goal.

Trial and failure are inevitable parts of the process in achieving artificial general intelligence (AGI). This inevitability stems from the intrinsic complexity and unpredictability of real-world environments, which cannot be entirely captured or anticipated in advance. Unlike narrowly focused artificial intelligence systems, AGI aims to operate across a wide range of tasks and adapt to novel situations, necessitating a learning process that embraces mistakes and learns from them. In this context, trial and error serve as fundamental mechanisms for learning. By experimenting with different actions and observing their outcomes, an AGI system can gradually refine its understanding of the environment and improve its decision-making capabilities. Each failure provides valuable information, highlighting the limitations of current strategies and offering insights into better approaches. This iterative process of trying, failing, analyzing, and adjusting is essential for developing the resilience and flexibility needed for AGI. Reinforcement learning, a key technique in developing AGI, inherently relies on trial and error. In reinforcement learning, an agent interacts with its environment, takes actions, and receives feedback in the form of rewards or penalties. Through repeated interactions, the agent learns to associate certain actions with positive or negative outcomes and adjusts its behavior to maximize cumulative rewards. This process is inherently experimental, involving many unsuccessful attempts before effective strategies emerge. Moreover, embracing trial and failure is crucial for fostering creativity and innovation in AGI systems. By allowing the AI to explore a wide range of possibilities, including those that lead to failure, we enable it to discover novel solutions and strategies that may not be immediately obvious. This exploratory aspect of learning is essential for AGI to handle the diverse and unforeseen challenges it will encounter in the real world.

The inevitability of trial and failure in the AGI development process reflects the way humans learn and adapt. Human intelligence is characterized by a continuous cycle of experimentation, feedback, and improvement. By mirroring this process, AGI systems can develop more human-like adaptability and problem-solving abilities, enhancing their effectiveness across a broad spectrum of tasks. Trial and failure are not just unavoidable but integral to the development of AGI. They drive the learning process, promote resilience and flexibility, and enable the discovery of innovative solutions. By embracing these elements, we can create AGI systems that are capable of continuous improvement and adaptation, ultimately achieving the goal of true general intelligence. Moreover, this method relies heavily on the concept of reward functions, which guide the AI by providing feedback on the desirability of different outcomes. By continuously optimizing its actions to maximize cumulative rewards, the AI system can develop sophisticated strategies for achieving its goals, even in complex and changing environments. Ultimately, the achievement of AGI through this approach involves creating systems that are not just reactive but proactive, capable of anticipating future states and planning accordingly. This shift from state-based training to action-oriented learning represents a fundamental change in how we design and understand intelligent systems, paving the way for the development of truly general artificial intelligence.

The Importance of Multimodal AI in Achieving AGI

The pursuit of Artificial General Intelligence (AGI) is driven by the goal of creating AI systems that can perform a wide range of tasks with the flexibility and adaptability of human intelligence. One of the key strategies in achieving AGI is the development of multimodal AI, which integrates multiple forms of data and sensory inputs to enhance the AI's understanding and reasoning capabilities. Multimodal AI systems combine visual, auditory, textual, and other types of information, enabling them to process and interpret complex, real-world scenarios more effectively than unimodal systems that rely on a single type of data. Multimodal AI is crucial for achieving AGI because it allows the AI to operate across a broader spectrum of tasks, rather than being confined to specific domains defined by narrow constraints. By leveraging diverse types of data, multimodal AI can develop a more holistic understanding of its environment, akin to human perception and cognition. This comprehensive understanding is essential for performing complex tasks that require the integration of different kinds of information, such as navigating a busy street, understanding spoken language in various accents, or interpreting visual cues in a social interaction.

The integration of multiple modalities enhances the robustness and versatility of AI systems. For instance, a multimodal AI system designed for autonomous driving can simultaneously process visual data from cameras, spatial data from LIDAR sensors, and auditory data from microphones. This combined input allows the system to build a richer and more accurate representation of its surroundings, improving its ability to make safe and effective driving decisions. Similarly, in healthcare, a multimodal AI can analyze medical images, patient records, and genetic data to provide more comprehensive and accurate diagnoses and treatment plans. Multimodal AI also plays a pivotal role in improving the AI's ability to generalize knowledge across different domains. By learning from diverse types of data, the AI can develop transferable skills that are applicable to various tasks and environments. This generalization capability is a cornerstone of AGI, as it enables the AI to adapt to new and unforeseen challenges without requiring extensive retraining. For example, an AI that has been trained to understand visual and textual data can apply its knowledge to tasks ranging from image captioning to sentiment analysis, demonstrating flexibility and adaptability. Moreover, the incorporation of multimodal data facilitates the development of more sophisticated reasoning and decision-making processes. By combining different types of information, the AI can draw richer and more informed inferences, leading to better outcomes. For instance, in a scenario where the AI needs to identify an object, it can use visual data to recognize the object's shape and color, while also using textual data to understand context-specific information about the object. This multimodal approach enables the AI to make more accurate and contextually relevant decisions.

The importance of multimodal AI in achieving AGI is further underscored by its potential to enhance human-AI interaction. Multimodal AI systems can understand and respond to human inputs in a more natural and intuitive manner, facilitating smoother and more effective communication. A key aspect of this development is the incorporation of human feedback to dynamically assign reward functions, which are essential for guiding AI learning and decision-making. By interpreting both verbal and non-verbal cues, multimodal AI systems can provide more personalized and context-aware responses, improving user experience. This real-time human feedback allows AI to refine its behavior continually, ensuring actions align with user preferences and ethical standards. Additionally, involving humans in the feedback loop fosters transparency and accountability, building trust and enabling the identification and correction of biases. As multimodal AI evolves, the role of human feedback in shaping AI behavior will be critical in advancing towards AGI, enhancing the AI's ability to learn, adapt, and perform reliably in diverse and complex real-world scenarios. By integrating multiple forms of data and sensory inputs, multimodal AI systems can operate across a broad range of tasks, demonstrating flexibility, robustness, and adaptability. This comprehensive approach not only enhances the AI's understanding and reasoning capabilities but also improves its ability to generalize knowledge and interact with humans effectively. As research and development in multimodal AI

continue to advance, it will play an increasingly vital role in the realization of AGI, unlocking new possibilities and transforming the capabilities of intelligent systems.

The Importance of Competing Models and Second Guessing

The journey toward achieving advanced artificial intelligence (AI) and, ultimately, artificial general intelligence (AGI) requires a fundamental shift in how we design and understand intelligent systems. One crucial aspect of this shift is the importance of competing models and second-guessing in decision-making processes. Determinism, while seemingly efficient and straightforward, often leads to inflexible systems that fail to adapt to the complexities and unpredictabilities of the real world. Instead, fostering a probabilistic approach, where models compete and continually reassess their decisions, mirrors the multifaceted nature of human cognition and behavior.

Human behavior is inherently non-deterministic. Our minds are continuously engaged in a dynamic interplay of competing thoughts, doubts, and reassessments. This internal competition among various cognitive processes is a fundamental characteristic of human intelligence. For instance, when making a decision, different parts of the brain, such as the left and right hemispheres, may process information differently, contributing diverse perspectives and competing strategies. This cognitive diversity allows humans to consider multiple potential outcomes, weigh probabilities, and make more robust decisions that can adapt to new information and changing circumstances. In AI development, embracing this paradigm of internal competition and second-guessing can lead to more resilient and adaptable systems. Instead of striving for a single, deterministic outcome, AI models should be designed to generate and evaluate multiple hypotheses. This approach encourages models to consider a range of possibilities, each with an associated probability, and to continually update these probabilities as new data becomes available. By doing so, AI systems can better handle uncertainty and make more informed decisions.

The role of competing models in AI can be likened to a marketplace of ideas, where different algorithms or sub-models propose various solutions to a given problem. Each model brings its own strengths, weaknesses, and perspectives, contributing to a richer overall decision-making process. This competition ensures that no single model's biases or limitations dominate the outcome. Instead, the system synthesizes input from multiple sources, leading to more balanced and well-rounded decisions. Second-guessing, or the ability to reassess and revise decisions, is equally important. In human cognition, second-guessing allows individuals to reconsider their initial judgments in light of new information or changing contexts. This flexibility is crucial for navigating complex and dynamic environments. Similarly, AI systems that incorporate mechanisms for second-guessing can avoid the pitfalls of rigid decision-making. By continually questioning and refining their choices, these systems can improve their performance over time and adapt to unforeseen challenges.

Probabilistic reasoning is at the heart of this approach. Unlike deterministic models that provide a single definitive answer, probabilistic models assign likelihoods to various outcomes, reflecting the inherent uncertainties of the real world. This probabilistic framework enables AI systems to make decisions based on expected values and risk assessments, rather than fixed outcomes. For instance, an AI system designed for medical diagnosis can evaluate the likelihood of various conditions based on patient data, considering the probabilities of each potential diagnosis rather than committing to a single conclusion. This approach allows the system to adapt its recommendations as more information becomes available, thereby improving its accuracy and reliability. The emphasis on probabilistic reasoning also enhances the robustness of AI systems in dealing with incomplete or noisy data. In real-world scenarios, data is often imperfect, and a probabilistic approach allows the AI to weigh different pieces of information according to their reliability, leading to more nuanced and accurate decisions. This flexibility is crucial for applications in fields such as finance, healthcare, and autonomous systems, where the ability to make sound decisions under uncertainty can have significant implications. Moreover, fostering an environment of competing models and second-guessing can drive innovation and continuous improvement in AI development. By encouraging multiple models

to propose solutions and challenge each other's assumptions, we can identify and address weaknesses more effectively. This iterative process of hypothesis generation, testing, and refinement is fundamental to scientific inquiry and can similarly enhance the development of intelligent systems.

Conclusion

The pursuit of Artificial General Intelligence (AGI) is one of the most ambitious and challenging goals in the field of artificial intelligence. This paper has explored various aspects crucial to the development of AGI, including the fundamental role of clear objectives and reward functions, the importance of real-world interaction and feedback, the challenges inherent in mathematical reasoning, the benefits and limitations of simulated environments, and the potential of game-based learning. The critical insight drawn from this discussion is that AGI cannot be achieved merely by training AI on specific states or tasks. Instead, AGI requires a more holistic approach where the AI continually takes the next best action to achieve its ultimate goal, adapting dynamically to new information and changing conditions. Real-world interaction is essential for AGI as it allows AI systems to experience the full complexity and unpredictability of their environment, receiving dynamic feedback that refines their reasoning and decision-making capabilities. By performing tasks in real environments and utilizing dynamic reward mechanisms, AI can learn to navigate the path of maximum reward, developing advanced reasoning abilities that consider long-term goals and consequences.

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *nature* **2015**, *521*, 436–444.
2. Li, Y. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274* **2017**.
3. François-Lavet, V.; Henderson, P.; Islam, R.; Bellemare, M.G.; Pineau, J.; others. An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning* **2018**, *11*, 219–354.
4. Henderson, P.; Islam, R.; Bachman, P.; Pineau, J.; Precup, D.; Meger, D. Deep reinforcement learning that matters. Proceedings of the AAAI conference on artificial intelligence, 2018, Vol. 32.
5. Dulac-Arnold, G.; Mankowitz, D.; Hester, T. Challenges of real-world reinforcement learning. *arXiv preprint arXiv:1904.12901* **2019**.
6. Zhao, W.; Queralta, J.P.; Westerlund, T. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. 2020 IEEE symposium series on computational intelligence (SSCI). IEEE, 2020, pp. 737–744.
7. Chaslot, G.; Bakkes, S.; Szita, I.; Spronck, P. Monte-carlo tree search: A new framework for game ai. Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 2008, Vol. 4, pp. 216–217.
8. Gelly, S.; Silver, D. Monte-Carlo tree search and rapid action value estimation in computer Go. *Artificial Intelligence* **2011**, *175*, 1856–1875.
9. Gelly, S.; Kocsis, L.; Schoenauer, M.; Sebag, M.; Silver, D.; Szepesvári, C.; Teytaud, O. The grand challenge of computer Go: Monte Carlo tree search and extensions. *Communications of the ACM* **2012**, *55*, 106–113.
10. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; others. Mastering the game of Go with deep neural networks and tree search. *nature* **2016**, *529*, 484–489.
11. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; others. Mastering the game of go without human knowledge. *nature* **2017**, *550*, 354–359.
12. Anthony, T.; Tian, Z.; Barber, D. Thinking fast and slow with deep learning and tree search. *Advances in neural information processing systems* **2017**, *30*.
13. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*.
14. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* **2018**.
15. Hao, Y.; Dong, L.; Wei, F.; Xu, K. Visualizing and understanding the effectiveness of BERT. *arXiv preprint arXiv:1908.05620* **2019**.
16. Koroteev, M.V. BERT: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943* **2021**.

17. Lin, T.; Wang, Y.; Liu, X.; Qiu, X. A survey of transformers. *AI open* **2022**, *3*, 111–132.
18. Bengesi, S.; El-Sayed, H.; Sarker, M.K.; Houkpati, Y.; Irungu, J.; Oladunni, T. Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. *IEEE Access* **2024**.
19. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günemann, S.; Hüllermeier, E.; others. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* **2023**, *103*, 102274.
20. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y.; others. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* **2024**, *15*, 1–45.
21. Kirchenbauer, J.; Geiping, J.; Wen, Y.; Katz, J.; Miers, I.; Goldstein, T. A watermark for large language models. International Conference on Machine Learning. PMLR, 2023, pp. 17061–17084.
22. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; others. A survey of large language models. *arXiv preprint arXiv:2303.18223* **2023**.
23. Xu, F.F.; Alon, U.; Neubig, G.; Hellendoorn, V.J. A systematic evaluation of large language models of code. Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming, 2022, pp. 1–10.
24. Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; others. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682* **2022**.
25. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D.; others. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* **2022**, *35*, 24824–24837.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.