Large Language Models and Cognitive Science: A Comprehensive Review of Similarities, Differences, and Challenges

Qian Niu*†, Junyu Liu†, Ziqian Bi‡, Pohsun Feng§, Benji Peng¶, Keyu Chen¶, Ming Li¶,
†Kyoto University
‡Indiana University
§National Taiwan Normal University
¶Georgia Institute of Technology
Corresponding Email: niu.qian.f44@kyoto-u.jp

***Abstract*—This comprehensive review explores the intersection of Large Language Models (LLMs) and cognitive science, examining similarities and differences between LLMs and human cognitive processes. We analyze methods for evaluating LLMs cognitive abilities and discuss their potential as cognitive models. The review covers applications of LLMs in various cognitive fields, highlighting insights gained for cognitive science research. We assess cognitive biases and limitations of LLMs, along with proposed methods for improving their performance. The integration of LLMs with cognitive architectures is examined, revealing promising avenues for enhancing artificial intelligence (AI) capabilities. Key challenges and future research directions are identified, emphasizing the need for continued refinement of LLMs to better align with human cognition. This review provides a balanced perspective on the current state and future potential of LLMs in advancing our understanding of both artificial and human intelligence.**

***Keywords*—Large Language Models, Cognitive Science, Cognitive Psychology, Neuroscience**

I. Introduction

The emergence of Large Language Models (LLMs) has sparked a revolution in artificial intelligence (AI), challenging our understanding of machine cognition and its relationship to human cognitive processes. As these models demonstrate increasingly sophisticated capabilities in language processing, reasoning, and problem-solving, they have become a focal point of interest for cognitive scientists seeking to unravel the mysteries of human cognition. This intersection of LLMs and cognitive science has given rise to a new frontier of research, offering unprecedented opportunities to explore the nature of intelligence, language, and thought.

The relationship between LLMs and cognitive science is multifaceted and bidirectional. On one hand, insights from cognitive science have informed the development and evaluation of LLMs, inspiring new architectures and training paradigms that aim to more closely mimic human cognitive processes. On the other hand, the remarkable performance of LLMs on various cognitive tasks has prompted researchers to reevaluate existing theories of cognition and consider new perspectives on how intelligence emerges from complex systems.

This review aims to provide a comprehensive overview of the current state of research at the intersection of LLMs and cognitive science. We explore the similarities and differences between LLMs and human cognitive processes, examining how these models perform on tasks traditionally used to study human cognition. We also delve into the methods developed for evaluating LLMs cognitive abilities, highlighting the challenges and opportunities in assessing AI through the lens of cognitive science. Furthermore, we investigate the potential of LLMs to serve as cognitive models, discussing their applications in various domains of cognitive science research and the insights they provide into human cognition. The review also addresses the cognitive biases and limitations of LLMs, as well as the ongoing efforts to improve their performance and align them more closely with human cognitive processes. We examine recent developments in this area, discussing the potential synergies and challenges that arise from combining these approaches.

As LLMs continue to evolve and their capabilities expand, it becomes increasingly important to critically assess their relationship with human cognition and their potential impact on cognitive science research. This review offers a balanced and comprehensive examination of these issues, presenting insights into the current state of the field. It identifies key areas for future research and discusses the challenges and opportuni-

ties at the exciting intersection of LLMs and cognitive science. By bridging AI with cognitive science, this line of inquiry promises to deepen our understanding of human cognition and inform the development of more sophisticated, ethical, and human-centric AI systems. This comprehensive and critical examination not only highlights the current achievements but also maps out a path forward in this dynamic area of study.

II. Comparison of LLMs and Human Cognitive Processes

LLMs have revolutionized our understanding of AI and its potential to mimic human cognitive processes. These models have shown capabilities that resemble human cognition in various tasks, including language processing, sensory judgments, and reasoning. However, despite these similarities, there are fundamental differences between LLMs and human cognitive processes that merit close examination. This section explores these similarities and differences, evaluates the methods used to assess LLMs cognitive abilities, and discusses the potential of LLMs as cognitive models. By comparing LLMs with human cognition, we can better understand the strengths and limitations of these models in emulating human thought processes.

A. Similarities and differences between LLMs and human cognitive processes

LLMs have demonstrated remarkable capabilities in various cognitive tasks, often exhibiting human-like behaviors and performance. One of the key similarities observed is in the domain of language processing. LLMs can achieve human-level word prediction performance in natural contexts, suggesting a deep connection between these models and human language processing [1]. Studies have shown that LLMs represent linguistic information similarly to humans, enabling accurate brain encoding and decoding during language processing [2]. This similarity extends to the neural level, where larger neural language models exhibit representations that are increasingly similar to neural response measurements from brain imaging [3].

LLMs also demonstrate human-like cognitive effects in certain tasks. For instance, GPT-3 exhibits priming, distance, SNARC, and size congruity effects, which are well-documented phenomena in human cognition [4]. Additionally, LLMs show content effects in logical reasoning tasks similar to humans, particularly in challenging tasks like syllogism validity judgments and the Wason selection task [5]. Research has shown that LLMs can capture aspects of human sensory judgments across multiple modalities. Marjieh et al. [6] demonstrated that similarity judgments from GPT models are significantly correlated with human data across six sensory modalities, including pitch, loudness, colors, consonants, taste, and timbre. This suggests that LLMs can extract significant perceptual information from language alone.

However, significant differences exist between LLMs and human cognitive processes. Humans generally outperform LLMs in reasoning tasks, especially with out-of-distribution prompts, demonstrating greater robustness and flexibility [7]. LLMs struggle to emulate human-like reasoning when faced with novel and constrained problems, indicating limitations in their ability to generalize beyond their training data. Lamprinidis [8] found that LLMs' cognitive judgments are not human-like in limited-data inductive reasoning tasks, with higher errors compared to Bayesian predictors. This suggests that LLMs may not model basic statistical principles that humans use in everyday scenarios as effectively as previously thought.

Moreover, while LLMs exhibit near human-level formal linguistic competence, they show patchy performance in functional linguistic competence [9]. This suggests that LLMs may excel at surface-level language processing but struggle with deeper, context-dependent understanding and reasoning. Another notable difference lies in the memory properties of LLMs compared to human memory. Although LLMs exhibit some human-like memory characteristics, such as primacy and recency effects, their forgetting mechanisms and memory structures differ from human biological memory [10]. Suresh et al. [11] found that human conceptual structures are robust and coherent across different tasks, languages, and cultures, while LLMs produce conceptual structures that vary significantly depending on the task used to generate responses. This highlights a fundamental difference in the stability and consistency of conceptual representations between humans and LLMs.

B. Methods for evaluating LLMs cognitive abilities

Researchers have developed various methods to evaluate the cognitive abilities of LLMs, often drawing inspiration from cognitive science and psychology. These methods aim to provide a comprehensive assessment of LLMs' capabilities and limitations in comparison to human cognition.

One prominent approach is the use of cognitive psychology experiments adapted for LLMs. For example, CogBench, a benchmark with ten behavioral metrics from seven cognitive psychology experiments, has been developed to evaluate LLMs [12]. This benchmark allows for a systematic comparison of LLMs performance across various cognitive tasks. Another

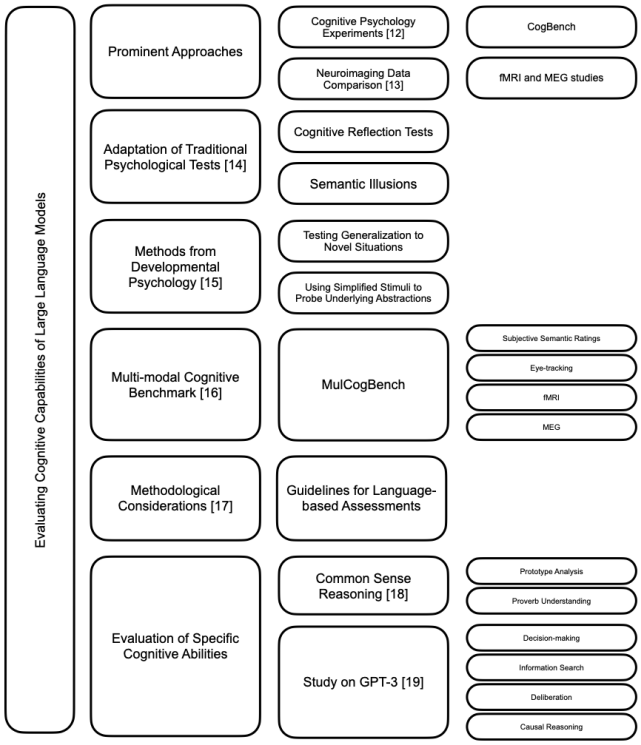

Figure 1: Evaluating Cognitive Capabilities of LLMs

method involves using neuroimaging data to compare LLMs representations with human brain activity. Studies have employed Functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) recordings to analyze the similarity between LLMs activations and brain responses during language processing tasks [13]. This approach provides insights into the neural-level similarities and differences between LLMs and human cognition.

Researchers have also adapted traditional psychological tests for use with LLMs. For instance, cognitive reflection tests and semantic illusions have been used to evaluate the reasoning capabilities of LLMs [14]. These tests help reveal the extent to which LLMs exhibit human-like biases and reasoning patterns. Additionally, methods from developmental psychology have been proposed to understand the capacities and underlying abstractions of LLMs [15]. These approaches focus on testing generalization to novel situations and using simplified stimuli to probe underlying abstractions.

In an effort to create more comprehensive evaluation tools, Zhang et al. [16] introduced MulCogBench, a multi-modal cognitive benchmark dataset for evaluating Chinese and English computational language models. This dataset includes various types of cognitive data, such as subjective semantic ratings, eye-tracking, fMRI, and MEG, allowing for a comprehensive comparison between LLMs and human cognitive processes. Ivanova [17] provided a set of methodological considerations for evaluating the cognitive capacities of LLMs using language-based assessments. The paper highlights common pitfalls and provides guidelines for designing high-quality cognitive evaluations, contributing to best practices in AI Psychology.

Delving deeper into specific cognitive abilities, Srinivasan et al. [18] proposed novel methods based on cognitive science principles to test LLMs' common sense reasoning abilities through prototype analysis and proverb understanding. These methods offer new ways to assess LLMs' cognitive capabilities in more nuanced and context-dependent tasks. Binz and Schulz [19] used tools from cognitive psychology to study GPT-3, assessing its decision-making, information search, deliberation, and causal reasoning abilities. Their approach demonstrates the potential of cognitive psychology in studying AI and demystifying how LLMs solve tasks.

In summary, Large Language Models exhibit remarkable parallels with human cognitive processes, particularly in language and sensory tasks, yet they fall short in several critical areas, such as reasoning under novel conditions and functional linguistic competence. The diverse methodologies employed to evaluate LLMs' cognitive abilities highlight both their potential and limitations as models of human cognition. As LLMs continue to evolve, they provide a valuable tool for exploring the nature of human intelligence, but their differences from human cognitive processes must be carefully considered. Future research should aim to refine these models further, improving their alignment with human cognition and addressing the gaps that currently exist. Understanding the complex interplay between LLMs and human cognitive processes will advance both AI and cognitive science, bridging the divide between machine and human intelligence.

III. Applications of LLMs in Cognitive Science

The integration of LLMs into cognitive science research has opened up new avenues for understanding human cognition and developing more sophisticated AI systems. This section explores the multifaceted applications of LLMs in cognitive science, examining their role as cognitive models, their contributions to theoretical insights, and their specific applications in various cognitive domains. By synthesizing recent research, we aim to provide a comprehensive overview of the current state and future potential of LLMs in advancing our understanding of human cognition.

A. LLMs as Cognitive Models

The potential of LLMs to serve as cognitive models has gained significant attention in recent research. Studies have demonstrated that LLMs can be turned into accurate cognitive models through fine-tuning on psychological experiment data, offering precise representations of human behavior and often outperforming traditional cognitive models in decision-making tasks [20]. These models have shown promise in capturing individual differences in behavior and generalizing to new tasks after being fine-tuned on multiple tasks, suggesting their potential to become generalist cognitive models capable of representing a wide range of human cognitive processes. Versatility of LLMs in various cognitive domains have been explored. Wong et al. [21] introduced a computational framework called rational meaning construction, integrating neural language models with probabilistic models for rational inference. This approach demonstrates LLMs' ability to generate context-sensitive translations and support commonsense reasoning across various cognitive domains. Piantadosi and Hill [22] highlighted LLMs' capacity to capture essential aspects of meaning through conceptual roles, challenging skepticism about their ability to possess human-like concepts.

In the realm of language processing, Schrimpf et al. [23] conducted a systematic integrative modeling study, revealing that transformer-based ANN models can predict neural and behavioral responses in human language processing. Their findings support the hypothesis that predictive processing shapes language comprehension mechanisms in the brain, aligning with contemporary theories in cognitive neuroscience. Kallens et al. [24] demonstrated that LLMs can produce human-like grammatical language without an innate grammar, providing valuable computational models for exploring statistical learning in language acquisition and challenging traditional views on language learning. Lampinen's [25] research further challenges our understanding of human language processing, demonstrating that with minimal prompting, LLMs can outperform humans in processing recursively nested grammatical structures. This raises questions about the cognitive mechanisms underlying both human and artificial language comprehension. Nolfi [26] explored the unexpected cognitive abilities developed by LLMs through indirect processes, including dynamical semantic operations, theory of mind, affordance recognition, and logical reasoning. These findings suggest that LLMs can develop integrated cognitive skills that work synergistically, despite being primarily trained on next-word prediction tasks. This research highlights the importance of understanding these emergent capabilities in relation to human cognition. Sartori and Orrú [27] provided empirical evidence that LLMs perform at human levels in a wide variety of cognitive tasks, including reasoning and problem-solving. Their findings support associationism as a unifying theory of cognition and demonstrate the potential for significant impact on cognitive psychology, suggesting new avenues for modeling human cognitive processes. Li and Li [28] proposed an intriguing duality between LLMs and Tulving's theory of memory, suggesting that consciousness may be an emergent ability based on this duality. This perspective offers a novel approach to understanding the relationship between LLMs and human cognition, potentially bridging artificial and biological intelligence research.

However, it is important to note that while LLMs can serve as plausible models of human language understanding, there are ongoing debates about the extent to which they truly capture human-like cognitive abstractions [29]. Some researchers argue that it is premature to make definitive claims about the abilities or limitations of LLMs as models of human language understanding, emphasizing the need for further empirical testing. Katzir [30] provided a balanced assessment of the strengths and weaknesses of LLMs, highlighting their sophisticated inductive learning capabilities while also addressing significant limitations such as opacity, data requirements, and differences from human cognitive processes. Besides, the use of LLMs as cognitive models offers new opportunities for understanding human cognition. By analyzing the internal representations and processes of these models, researchers can gain insights into potential mechanisms underlying human cognitive abilities. However, caution is necessary when interpreting these findings, as the fundamental differences in architecture and learning processes between LLMs and the human brain must be considered. Ren et al. [31] investigated how well LLMs align with human brain cognitive processing signals using Representational Similarity Analysis (RSA). Their findings suggest that factors such as pretraining data size, model scaling, and alignment training significantly impact the similarity between LLMs and brain activity, providing insights into how LLMs might be improved to better model human cognition.

In conclusion, while LLMs show great promise as cognitive models, further research is needed to fully understand their capabilities and limitations in representing human cognitive processes. The ongoing exploration of LLMs as cognitive models continues to provide valuable insights into both artificial and human cognition, potentially reshaping our understanding of language, reasoning, and cognitive processes.

B. Insights from LLMs for cognitive science research

LLMs have provided valuable insights for cognitive science research, challenging existing theories and offering new perspectives on human cognition. Veres [32] argued that while LLMs challenge rule-based theories, they do not necessarily provide deeper insights into the nature of language or cognition. This perspective highlights the need for careful interpretation of LLMs capabilities in the context of cognitive science and cautions against overinterpretation of model performance. Shanahan [33] emphasized the importance of understanding the true nature and capabilities of LLMs to avoid anthropomorphism and ensure responsible use and discourse around AI in cognitive science research. This cautionary approach underscores the need for precise language and philosophical nuance in AI discourse, particularly when drawing parallels between artificial and human cognition. Blank [34] explored whether LLMs can be considered computational models of human language processing, discussing different interpretations and implications for future research. This work highlights the ongoing debate about whether LLMs process language like humans and the significance of this question for cognitive science, emphasizing the need for rigorous empirical investigation. Grindrod [35] argued that LLMs can serve as scientific models of E-languages (external languages), providing insights into the nature of language as a social entity. This perspective offers a novel approach to using LLMs in linguistic inquiry and cognitive science research, potentially bridging computational linguistics and sociolinguistics.

The application of LLMs in cognitive science research has opened up new avenues for exploring human behavior and decision-making processes. Horton [36] demonstrated the potential of using LLMs as simulated economic agents to replicate classic behavioral economics experiments. This innovative approach suggests new possibilities for using LLMs to explore human behavior and decision-making processes in cognitive science, offering a cost-effective method for piloting studies and generating hypotheses. Connell and Lynott [37] evaluated the cognitive plausibility of different types of language models, emphasizing the importance of learning mechanisms, corpus size, and grounding in assessing their relevance to human cognition. Their work provides a framework for critically evaluating the applicability of LLMs to cognitive modeling. Mitchell and Krakauer [38] surveyed the debate on whether LLMs understand language in a humanlike sense, advocating for an extended science of intelligence to explore diverse modes of cognition. This perspective highlights the need for a broader understanding of intelligence and cognition in the context of LLMs, encouraging interdisciplinary collaboration in AI and cognitive science research. Buttrick [39] proposed using LLMs to study cultural distinctions by analyzing the statistical regularities in their training data, offering new avenues for exploring cultural cognition and representation. This approach demonstrates the potential of LLMs as tools for investigating complex sociocultural phenomena in cognitive science. Finally, Demszky et al. [40] reviewed the potential of LLMs to transform psychology by enabling large-scale analysis and generation of language data. They emphasized the need for further research and development to address ethical concerns and harness the full potential of LLMs in psychological research, highlighting both the opportunities and challenges in this emerging field.

In conclusion, LLMs have demonstrated significant potential as cognitive models and have provided valuable insights for cognitive science research. However, their limitations and the need for careful interpretation of their capabilities underscore the importance of continued research and interdisciplinary collaboration in this rapidly evolving field. Future work should focus on refining LLMs to better align with human cognitive processes, developing more rigorous evaluation methods, and addressing ethical considerations to ensure responsible and productive integration of LLMs in cognitive science research.

C. Application of LLMs in specific cognitive fields

LLMs have demonstrated significant potential in various cognitive domains, including causal reasoning, lexical semantics, and creative writing. In the realm of causal inference, Liu et al. [41] conducted a comprehensive survey exploring the mutual benefits between LLMs and causal inference, highlighting how causal perspectives can enhance LLMs' reasoning capacities, fairness, and safety. Similarly, Kıcıman et al. [42] benchmarked the causal capabilities of LLMs, finding that they outperform existing methods in generating causal arguments across various tasks, while also noting their limitations in critical decision-making scenarios. In the field of lexical semantics, Petersen and Potts [43] utilized LLMs to conduct a detailed case study of the English verb "break," demonstrating that LLM representations can capture known sense distinctions and identify new sense combinations. Their findings suggest a reconsideration of the commitment to discreteness in semantic theory, favoring a more fluid, usage-based approach. Extending to creative domains, Chakrabarty et al. [44] investigated the utility of LLMs in assisting professional writers through an empirical user study. Their research revealed that writers find

LLMs most helpful for translation and review tasks rather than planning, while also identifying significant weaknesses in current models, such as reliance on clichés and lack of nuance.

These studies collectively underscore the diverse applications of LLMs in cognitive fields, from enhancing causal reasoning to supporting creative processes, while also highlighting areas for improvement and future research directions. In conclusion, the application of Large Language Models in cognitive science research represents a significant advancement in our ability to model and understand human cognition. LLMs have demonstrated remarkable potential as cognitive models, offering insights into language processing, reasoning, and decision-making that challenge and expand existing theories. Their versatility in addressing diverse cognitive tasks, from causal inference to creative writing, underscores their value as research tools across multiple domains of cognitive science.

However, the integration of LLMs into cognitive research is not without challenges. Researchers must navigate issues of interpretability, ethical considerations, and the potential for overinterpretation of model capabilities. The ongoing debate about the nature of LLMs "understanding" and its relationship to human cognition highlights the need for continued critical examination and empirical investigation. As the field progresses, interdisciplinary collaboration will be crucial in refining LLMs to better align with human cognitive processes, developing more rigorous evaluation methods, and addressing ethical concerns. The future of LLMs in cognitive science research holds promise for transformative insights into the nature of intelligence, both artificial and biological, potentially bridging gaps between computational models and human cognition. By carefully leveraging the strengths of LLMs while acknowledging their limitations, researchers can continue to push the boundaries of our understanding of the mind and pave the way for more advanced AI systems that complement and enhance human cognitive abilities.

IV. Limitations and Improvement of LLMs Capabilities

The rapid advancement of LLMs has necessitated a comprehensive evaluation of their capabilities and limitations. This section examines the cognitive biases and constraints inherent in LLMs, as well as proposed methods for enhancing their performance. By critically analyzing these aspects, researchers aim to develop more robust and reliable AI systems that can better emulate human-like cognition and language understanding.

A. Cognitive biases and limitations of LLMs

Recent studies have extensively explored the cognitive biases and limitations of LLMs. Ullman [45] demonstrated that LLMs fail on trivial alterations to Theory-of-Mind tasks, suggesting a lack of robust Theory-of-Mind capabilities. Talboy and Fuller [46] identified multiple cognitive biases in LLMs similar to those found in human reasoning, highlighting the need for increased awareness and mitigation strategies. Thorstad [47] advocated for cautious optimism about LLMs performance while acknowledging genuine biases, particularly framing effects. Singh et al. [48] investigated the confidence-competence gap in LLMs, revealing instances of overconfidence and underconfidence reminiscent of the Dunning-Kruger effect. Marcus et al. [49] argued that LLMs currently lack deeper linguistic and cognitive understanding, leading to incomplete and biased representations of human language. Macmillan-Scott and Musolesi [50] evaluated seven LLMs using cognitive psychology tasks, finding that they display irrationality differently from humans and exhibit significant inconsistency in their responses. Jones and Steinhardt [51] presented a method inspired by human cognitive biases to systematically identify and test for qualitative errors in LLMs, uncovering predictable and high-impact errors. Smith et al. [52] proposed using the term "confabulation" instead of "hallucination" to more accurately describe inaccurate outputs of LLMs, emphasizing the importance of precise metaphorical language in understanding AI processes.

B. Methods for improving LLMs performance

Researchers have proposed various methods to improve LLMs performance and address their limitations. Nguyen [53] introduced the bounded pragmatic speaker model to understand and improve language models by drawing parallels with human cognition and suggesting enhancements to reinforcement learning from human feedback (RLHF). Lv et al. [54] developed CogGPT, an LLM-driven agent with an iterative cognitive mechanism that outperforms existing methods in facilitating role-specific cognitive dynamics under continuous information flows. Prystawski et al. [55] demonstrated that using chain-of-thought prompts informed by probabilistic models can improve LLMs' ability to understand and paraphrase metaphors. Aw and Toneva [56] found that training language models to summarize narratives improves their alignment with human brain activity, indicating deeper language understanding. Du et al. [57] reviewed recent developments addressing shortcut learning and robustness challenges in LLMs, suggesting the combination of

data-driven schemes with domain knowledge and the introduction of more inductive biases into model architectures.

These studies collectively highlight the importance of understanding and addressing cognitive biases and limitations in LLMs while exploring innovative methods to enhance their performance and alignment with human cognition. Future research should focus on developing more robust evaluation techniques, integrating insights from cognitive science, and creating LLMs that exhibit deeper linguistic and cognitive understanding.

In conclusion, the assessment and improvement of LLM capabilities remain critical areas of research in the field of AI. The studies reviewed in this section collectively highlight the importance of understanding and addressing cognitive biases and limitations in LLMs while exploring innovative methods to enhance their performance and alignment with human cognition. Future research should focus on developing more robust evaluation techniques, integrating insights from cognitive science, and creating LLMs that exhibit deeper linguistic and cognitive understanding. By addressing these challenges, researchers can pave the way for more advanced and reliable AI systems that can better serve human needs and contribute to various domains of knowledge and application.

V. Integration of LLMs with Cognitive Architectures

Recent research has explored various approaches to integrate LLMs with cognitive architectures, aiming to enhance AI systems' capabilities. This synergistic approach leverages the strengths of both LLMs and cognitive architectures while mitigating their respective weaknesses. Romero et al. [58] presented three integration approaches: modular, agency, and neuro-symbolic, each with its own theoretical grounding and empirical support. Kirk et al. [59] explored the direct extraction of task knowledge from GPT-3 by cognitive agents, using template-based prompting and natural-language interaction. They proposed a six-step process for knowledge extraction and integration into cognitive architectures. Joshi and Ustun [60] proposed a method to augment cognitive architectures like Soar and Sigma with generative LLMs, using them as prompt-able declarative memory within the architecture. González-Santamarta et al. [61] integrated LLMs into the MERLIN2 cognitive architecture for autonomous robots, focusing on enhancing reasoning capabilities and human-robot interaction.

Several studies have demonstrated the potential benefits of combining LLMs with cognitive architectures in various domains. Zhu and Simmons [62] presented a framework that combines LLMs with cognitive architectures to create an efficient and adaptable agent for performing kitchen tasks. Their approach demonstrated improved efficiency and fewer required tokens compared to using LLMs alone. Nakos and Forbus [63] discussed the integration of BERT into the Companion cognitive architecture, showing improvements in disambiguation and fact plausibility prediction for natural language understanding tasks. Wray et al. [64] reviewed the capabilities of LMs for cognitive systems and proposed a research strategy for integrating LMs into cognitive agents to improve task learning and performance. They emphasized the need for effective prompting, interpretation, and verification strategies. Zhou et al.[65] proposed a Cognitive Personalized Search (CoPS) model that integrates LLMs with a cognitive memory mechanism inspired by human cognition to enhance user modeling and improve personalized search results.

These studies collectively demonstrate the potential of integrating LLMs with cognitive architectures to create more robust, efficient, and adaptable AI systems. However, challenges remain, including ensuring the accuracy and relevance of extracted knowledge, managing computational costs, and addressing the limitations of both LLMs and cognitive architectures. Future research directions include exploring more sophisticated integration methods, improving the efficiency of LLM-based reasoning, and investigating the application of these integrated systems in various domains.

VI. Discussion

The intersection of LLMs and cognitive science has opened up a fascinating new frontier in AI and our understanding of human cognition. This review has highlighted the significant progress made in comparing LLMs and human cognitive processes, developing methods for evaluating LLMs cognitive abilities, and exploring the potential of LLMs as cognitive models. However, it also reveals several important areas for future research and consideration.

One of the most striking findings is the remarkable similarity between LLMs and human cognitive processes in certain domains, particularly in language processing and some aspects of reasoning. The ability of LLMs to exhibit human-like priming effects, content effects in logical reasoning, and even capture aspects of human sensory judgments across multiple modalities suggests a deep connection between these artificial systems and human cognition. This similarity extends to the neural level, with larger neural language models showing representations increasingly similar to neural response measurements from brain imaging.

However, the review also underscores significant differences between LLMs and human cognitive processes. Humans generally outperform LLMs in reasoning tasks, especially with out-of-distribution prompts, demonstrating greater robustness and flexibility. The struggle of LLMs to emulate human-like reasoning when faced with novel and constrained problems indicates limitations in their ability to generalize beyond their training data. Moreover, while LLMs exhibit near human-level formal linguistic competence, they show patchy performance in functional linguistic competence, suggesting a gap in deeper, context-dependent understanding and reasoning.

These findings highlight the need for future research to focus on enhancing the generalization capabilities of LLMs and improving their performance in functional linguistic competence. Developing methods to imbue LLMs with more robust and flexible reasoning abilities, particularly in novel and constrained problem spaces, could significantly advance their cognitive capabilities.

The review also reveals the potential of LLMs as cognitive models, with studies demonstrating that fine-tuned LLMs can offer precise representations of human behavior and often outperform traditional cognitive models in decision-making tasks. This suggests a promising avenue for using LLMs to gain insights into human cognitive processes. However, caution is necessary when interpreting these findings, as the fundamental differences in architecture and learning processes between LLMs and the human brain must be considered.

VII. Future Challenge

Future research should focus on developing more sophisticated methods for aligning LLMs with human cognitive processes. This could involve integrating insights from cognitive science into the architecture and training of LLMs, as well as exploring novel ways to evaluate and compare LLMs performance with human cognition across a wider range of cognitive tasks.

The application of LLMs in specific cognitive fields, such as causal reasoning, lexical semantics, and creative writing, demonstrates their potential to contribute to various areas of cognitive science research. However, it also highlights the need for continued refinement and specialization of LLMs for specific cognitive domains. Future work could focus on developing domain-specific LLMs that more accurately model human cognition in particular areas of expertise.

The review also addresses the cognitive biases and limitations of LLMs, revealing that these models can exhibit biases similar to those found in human reasoning. This finding presents both challenges and opportunities. On one hand, it underscores the need for increased awareness and mitigation strategies to address these biases in AI systems. On the other hand, it offers a unique opportunity to study cognitive biases in a controlled, artificial environment, potentially leading to new insights into the nature and origins of these biases in human cognition.

The integration of LLMs with cognitive architectures represents a promising direction for future research. This approach aims to leverage the strengths of both LLMs and cognitive architectures while mitigating their respective weaknesses. Future work in this area could focus on developing more sophisticated integration methods, improving the efficiency of LLM-based reasoning within cognitive architectures, and exploring the application of these integrated systems in various real-world domains.

In conclusion, the intersection of LLMs and cognitive science offers exciting possibilities for advancing our understanding of both artificial and human intelligence. However, it also presents significant challenges that require careful consideration and further research. As we continue to explore this frontier, it is crucial to maintain a balanced perspective, acknowledging both the remarkable capabilities of LLMs and their current limitations. By doing so, we can work towards developing AI systems that not only perform well on specific tasks but also contribute to our understanding of cognition itself.

References

[1] A. Goldstein, Z. Zada, E. Buchnik, M. Schain, A. Price, S. A. Nastase, A. Feder, D. Emanuel, A. Cohen, A. Jansen, H. Gazula, G. Choe, A. Rao, C. Kim, C. Casto, A. Flinker, S. Devore, W. Doyle, D. Friedman, P. Dugan, A. Hassidim, M. Brenner, Y. Matias, K. A. Norman, O. Devinsky, and U. Hasson, "Thinking ahead: prediction in context as a keystone of language in hum ans and machines," *arXiv [cs.CL]*, 2020.

[2] G. Tuckute, N. Kanwisher, and E. Fedorenko, "Language in brains, minds, and machines," *Annu. Rev. Neurosci.*, 2024.

[3] G. Mischler, Y. A. Li, S. Bickel, A. D. Mehta, and N. Mesgarani, "Contextual feature extraction hierarchies converge in large language M odels and the brain," *arXiv [cs.CL]*, 2024.

[4] J. Shaki, S. Kraus, and M. Wooldridge, "Cognitive effects in large language models," *European Conference on Artificial Intelligence*, 2023.

[5] I. Dasgupta, A. K. Lampinen, S. C. Y. Chan, H. R. Sheahan, A. Creswell, D. Kumaran, J. L. McClelland, and F. Hill, "Language models show human-like content effects on reasoning tasks," *arXiv [cs.CL]*, 2022.

[6] R. Marjieh, I. Sucholutsky, P. van Rijn, N. Jacoby, and T. L. Griffiths, "Large language models predict human sensory judgments across six modal ities," *arXiv [cs.CL]*, 2023.

[7] K. M. Collins, C. Wong, J. Feng, M. Wei, and J. B. Tenenbaum, "Structured, flexible, and robust: benchmarking and improving large lan guage models towards more human-like behavior in out-of-distribution r easoning tasks," *arXiv [cs.CL]*, 2022.

[8] S. Lamprinidis, "LLM cognitive judgements differ from human," *arXiv [cs.CL]*, 2023.

[9] K. Mahowald, A. A. Ivanova, I. A. Blank, N. Kanwisher, J. B. Tenenbaum, and E. Fedorenko, "Dissociating language and thought in large language models," *arXiv [cs.CL]*, 2023.

[10] R. A. Janik, "Aspects of human memory and large language models," *arXiv [cs.CL]*, 2023.

[11] S. Suresh, K. Mukherjee, X. Yu, W.-C. Huang, L. Padua, and T. Rogers, "Conceptual structure coheres in human cognition but not in large langu age models," *Conference on Empirical Methods in Natural Language Processing*, 2023.

[12] J. Coda-Forno, M. Binz, J. X. Wang, and E. Schulz, "CogBench: a large language model walks into a psychology lab," *arXiv [cs.CL]*, 2024.

[13] C. Caucheteux and J. King, "Brains and algorithms partially converge in natural language processin g," *Commun. Biol.*, 2022.

[14] T. Hagendorff and S. Fabi, "Human-like intuitive behavior and reasoning biases emerged in language models – and disappeared in GPT-4," *ArXiv*, 2023.

[15] M. C. Frank, "Baby steps in evaluating the capacities of large language models," *Nat. Rev. Psychol.*, 2023.

[16] Y. Zhang, X. Zhang, C. Li, S. Wang, and C. Zong, "MulCog-Bench: A multi-modal cognitive benchmark dataset for evaluating chinese and english computational language models," *arXiv [cs.CL]*, 2024.

[17] A. A. Ivanova, "Running cognitive evaluations on large language models: The do's and t he don'ts," *arXiv [cs.CL]*, 2023.

[18] R. Srinivasan, H. Inakoshi, and K. Uchino, "Leveraging cognitive science for testing large language models," *International Conference on Artificial Intelligence Testing*, 2023.

[19] M. Binz and E. Schulz, "Using cognitive psychology to understand GPT-3," *Proc. Natl. Acad. Sci. U. S. A.*, 2022.

[20] ——, "Turning large language models into cognitive models," *arXiv [cs.CL]*, 2023.

[21] L. Wong, G. Grand, A. K. Lew, N. D. Goodman, V. K. Mansinghka, J. Andreas, and J. B. Tenenbaum, "From word models to world models: Translating from natural language to the probabilistic language of thought," *arXiv [cs.CL]*, 2023.

[22] S. T. Piantadosi and F. Hill, "Meaning without reference in large language models," *arXiv [cs.CL]*, 2022.

[23] M. Schrimpf, I. Blank, G. Tuckute, C. Kauf, E. A. Hosseini, N. Kanwisher, J. Tenenbaum, and E. Fedorenko, "The neural architecture of language: Integrative modeling converges on predictive processing," *Proc. Natl. Acad. Sci. U. S. A.*, 2020.

[24] P. C. Kallens, R. Kristensen-Mclachlan, and M. H. Christiansen, "Large language models demonstrate the potential of statistical learnin g in language," *Cogn. Sci. (Hauppauge)*, 2023.

[25] A. K. Lampinen, "Can language models handle recursively nested grammatical structures? a case study on comparing models and humans," *arXiv [cs.CL]*, 2022.

[26] S. Nolfi, "On the unexpected abilities of large language models," *arXiv [cs.CL]*, 2023.

[27] G. Sartori and G. Orrú, "Language models and psychological sciences," *Front. Psychol.*, 2023.

[28] J. Li and J. Li, "Memory, consciousness and large language model," *arXiv [cs.CL]*, 2024.

[29] E. Pavlick, "Symbols and grounding in large language models," *Philosophical Transactions of the Royal Society A*, 2023.

[30] R. Katzir, "Why large language models are poor theories of human linguistic cognit ion: A reply to piantadosi," *Biolinguistics*, 2023.

[31] Y. Ren, R. Jin, T. Zhang, and D. Xiong, "Do large language models mirror cognitive language processing?" *arXiv [cs.CL]*, 2024.

[32] C. Veres, "A precis of language models are not models of language," *arXiv [cs.CL]*, 2022.

[33] M. Shanahan, "Talking about large language models," *arXiv [cs.CL]*, 2022.

[34] I. Blank, "What are large language models supposed to model?" *Trends Cogn. Sci.*, 2023.

[35] J. Grindrod, "Modelling language," *arXiv [cs.CL]*, 2024.

[36] J. Horton, "Large language models as simulated economic agents: What can we learn from homo silicus?" *Social Science Research Network*, 2023.

[37] L. Connell and D. Lynott, "What can language models tell us about human cognition?" *Curr. Dir. Psychol. Sci.*, 2024.

[38] M. Mitchell and D. Krakauer, "The debate over understanding in AI's large language models," *Proc. Natl. Acad. Sci. U. S. A.*, 2022.

[39] N. Buttrick, "Studying large language models as compression algorithms for human cul ture," *Trends Cogn. Sci.*, 2024.

[40] D. Demszky, D. Yang, D. S. Yeager, C. J. Bryan, M. Clapper, S. Chandhok, J. Eichstaedt, C. A. Hecht, J. P. Jamieson, M. Johnson, M. Jones, D. Krettek-Cobb, L. Lai, N. JonesMitchell, D. C. Ong, C. Dweck, J. J. Gross, and J. W. Pennebaker, "Using large language models in psychology," *Nat. Rev. Psychol.*, 2023.

[41] X. Liu, P. Xu, J. Wu, J. Yuan, Y. Yang, Y. Zhou, F. Liu, T. Guan, H. Wang, T. Yu, J. McAuley, W. Ai, and F. Huang, "Large language models and causal inference in collaboration: A compreh ensive survey," *arXiv [cs.CL]*, 2024.

[42] E. Kıcıman, R. Ness, A. Sharma, and C. Tan, "Causal reasoning and large language models: Opening a new frontier for causality," *arXiv [cs.CL]*, 2023.

[43] E. A. Petersen and C. Potts, "Lexical semantics with large language models: A case study of english "break"," *Findings*, 2023.

[44] T. Chakrabarty, V. Padmakumar, F. Brahman, and S. Muresan, "Creativity support in the age of large language models: An empirical S tudy involving emerging writers," *arXiv [cs.CL]*, 2023.

[45] T. Ullman, "Large language models fail on trivial alterations to theory-of-mind ta sks," *arXiv [cs.CL]*, 2023.

[46] A. N. Talboy and E. Fuller, "Challenging the appearance of machine intelligence: Cognitive bias in LLMs and best practices for adoption," *arXiv [cs.CL]*, 2023.

[47] D. Thorstad, "Cognitive bias in large language models: Cautious optimism meets anti- panglossian meliorism," *arXiv [cs.CL]*, 2023.

[48] A. K. Singh, S. Devkota, B. Lamichhane, U. Dhakal, and C. Dhakal, "The confidence-competence gap in large language models: A cognitive st udy," *arXiv [cs.CL]*, 2023.

[49] G. Marcus, E. Leivada, and E. Murphy, "A sentence is worth a thousand pictures: Can large language models und erstand human language?" *arXiv [cs.CL]*, 2023.

[50] O. Macmillan-Scott and M. Musolesi, "(ir)rationality and cognitive biases in large language models," *arXiv [cs.CL]*, 2024.

[51] E. Jones and J. Steinhardt, "Capturing failures of large language models via human cognitive biases," *Neural Information Processing Systems*, 2022.

[52] A. L. Smith, F. Greaves, and T. Panch, "Hallucination or confabulation? neuroanatomy as metaphor in large lang uage models," *PLOS Digit. Health*, 2023.

[53] K. Nguyen, "Language models are bounded pragmatic speakers: Understanding RLHF fro m a bayesian cognitive modeling perspective," *arXiv [cs.CL]*, 2023.

[54] Y. Lv, H. Pan, R. Fu, M. Liu, Z. Wang, and B. Qin, "CogGPT: Unleashing the power of cognitive dynamics on large language M odels," *arXiv [cs.CL]*, 2024.

[55] B. Prystawski, P. Thibodeau, C. Potts, and N. D. Goodman, "Psychologically-informed chain-of-thought prompts for metaphor underst anding in large language models," *arXiv [cs.CL]*, 2022.

[56] K. L. Aw and M. Toneva, "Training language models to summarize narratives improves brain alignm ent," *arXiv [cs.CL]*, 2022.

[57] M. Du, F. He, N. Zou, D. Tao, and X. Hu, "Shortcut learning of large language models in natural language underst anding," *Commun. ACM*, 2022.

[58] O. J. Romero, J. Zimmerman, A. Steinfeld, and A. Tomasic, "Synergistic integration of large language models and cognitive archite ctures for robust AI: An exploratory analysis," *arXiv [cs.CL]*, 2023.

[59] J. R. Kirk, R. E. Wray, and J. E. Laird, "Exploiting language models as a source of knowledge for cognitive agen ts," *arXiv [cs.CL]*, 2023.
[60] H. Joshi and V. Ustun, "Augmenting cognitive architectures with large language models," *Proceedings of the AAAI Symposium Series*, 2024.
[61] M. A. Gonzalez-Santamarta, F. J. Rodríguez-Lera, A. M. Guerrero-Higueras, and V. Matellan-Olivera, "Integration of large language models within cognitive architectures for autonomous robots," *arXiv [cs.CL]*, 2023.
[62] F. Zhu and R. Simmons, "Bootstrapping cognitive agents with a large language model," *AAAI Conference on Artificial Intelligence*, 2024.
[63] C. Nakos and K. D. Forbus, "Using large language models in the companion cognitive architecture: A case study and future prospects," *Proceedings of the AAAI Symposium Series*, 2024.
[64] R. Wray, J. R. Kirk, and J. Laird, "Language models as a knowledge source for cognitive agents," *arXiv.org*, 2021.
[65] Y. Zhou, Q. Zhu, J. Jin, and Z. Dou, "Cognitive personalized search integrating large language models with a n efficient memory mechanism," *arXiv [cs.CL]*, 2024.