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

Qian Niu[1,2] Junyu Liu[2] Ziqian Bi[3] Pohsun Feng[4]
Benji Peng[5] Keyu Chen[6]

[2]Kyoto University, Kyoto, Japan
[3]Indiana University, Bloomington, IN, USA
[4]National Taiwan Normal University, Taipei, Taiwan
[5]Purdue University, West Lafayette, IN, USA

Table of Contents

Aim

- The rise of LLMs has prompted renewed interest in how machines can simulate or approximate human cognitive functions.
- This development has opened up new possibilities for cognitive scientists to explore the nature of intelligence, language, and thought through the lens of AI.
- LLMs are not only tools for performing language tasks but also serve as models to examine cognitive phenomena.
- Researchers use LLMs to investigate how complex behaviour and apparent understanding can emerge from large-scale statistical learning.
- These models raise compelling questions about whether intelligence must necessarily be embodied or symbolically grounded.

Aim of the Review

- This review aims to provide a comprehensive overview of current research at the intersection of LLMs and cognitive science.
- The review explores how LLMs compare with human cognitive processes across various dimensions, including language understanding, reasoning, and problem-solving.
- It examines how LLMs perform on tasks traditionally used in cognitive psychology and how their abilities are evaluated in comparison to human benchmarks.
- The review also discusses the potential for LLMs to act as cognitive models that inform research in linguistics, developmental psychology, and neuroscience.

Contributions

- This review contributes to a deeper interdisciplinary understanding of intelligence by bridging cognitive science and artificial intelligence.
- It identifies challenges and research opportunities at this intersection and highlights how such collaboration can support the development of more interpretable, trustworthy, and cognitively inspired AI systems.

Similarities in Language Processing and Representation

- LLMs demonstrate human-level word prediction performance in natural language contexts, suggesting strong parallels with human language processing.
- Research indicates that LLMs represent linguistic information in ways that align with brain encoding and decoding patterns observed during human language tasks.
- Larger LLMs produce internal representations that resemble neural response patterns measured in brain imaging studies.

Cognitive Effects and Sensory Judgements

- LLMs exhibit human-like cognitive effects, including priming, distance effects, SNARC effects, and size congruity effects.
- They also show content effects in logical reasoning tasks, such as syllogism judgments and the Wason selection task.
- LLMs correlate significantly with human data in sensory judgements across multiple modalities, including pitch, loudness, colour, taste, and timbre.

Key Differences in Reasoning and Memory

- Humans outperform LLMs in reasoning tasks, particularly under out-of-distribution or novel conditions that require generalisation beyond training data.
- LLMs exhibit higher error rates in inductive reasoning tasks and often fail to model basic statistical reasoning principles used by humans.
- Although LLMs show primacy and recency effects, their memory mechanisms differ substantially from human biological memory.
- Human conceptual structures remain consistent across contexts, whereas LLMs produce task-dependent and less coherent conceptual representations.

Limitations in Linguistic and Conceptual Competence

- LLMs exhibit near human-level formal linguistic competence but struggle with functional linguistic competence in context-rich settings.
- Their reasoning abilities often depend heavily on surface-level linguistic patterns rather than deep semantic understanding.
- The stability and consistency of conceptual knowledge in LLMs remain limited compared to human cognition.

Methods for Evaluating Cognitive Abilities in LLMs

- Researchers employ cognitive psychology tasks such as CogBench to assess LLM performance across behavioural metrics.
- Neuroimaging comparisons use fMRI and MEG data to assess similarity between LLM representations and human brain activity during language tasks.
- Classical psychological tasks, cognitive reflection tests, and semantic illusions are adapted to measure reasoning and biases in LLMs.

Emerging Benchmarks and Methodological Innovations

- Tools like MulCogBench integrate multimodal data including eye-tracking, subjective ratings, and neural responses to evaluate cognitive alignment.
- Developmental psychology-inspired tests assess generalisation and abstraction capabilities in LLMs using simplified stimuli.
- Advanced methods such as prototype analysis and proverb understanding have been proposed to test common sense reasoning in LLMs.
- Cognitive psychology frameworks are increasingly used to study decision-making, information search, and causal reasoning in language models.

Summary of Comparisons and Evaluation Methods

- LLMs exhibit remarkable human-like behaviour in language processing and perceptual tasks, but fall short in reasoning, generalisation, and conceptual consistency.
- A variety of methodologiesâranging from cognitive benchmarks to neuroimaging comparisonsâoffer a rich toolkit for evaluating LLMs as cognitive models.
- Future research must continue refining evaluation methods and model design to bridge the gap between artificial and human cognition.

Overview : Applications in Cognitive Science

- The integration of Large Language Models (LLMs) into cognitive science has opened new avenues for exploring human cognition and advancing AI systems.
- This section presents how LLMs are used as cognitive models, their theoretical contributions, and their applications across diverse cognitive domains.
- Synthesising recent research provides a clearer picture of their potential and limitations in understanding human cognition.

LLMs as Cognitive Models : General Modelling Potential

- LLMs have demonstrated the ability to model human behaviour through fine-tuning on psychological experiment data, often outperforming traditional cognitive models in decision-making tasks.
- Fine-tuned LLMs can capture individual behavioural differences and generalise across tasks, supporting their use as generalist cognitive models.
- Integrative frameworks, such as rational meaning construction, combine LLMs with probabilistic models to produce context-sensitive inference and commonsense reasoning.

LLMs as Cognitive Models : Language and Meaning

- Studies show that transformer-based LLMs can predict human neural and behavioural responses during language processing, supporting predictive processing theories in neuroscience.
- LLMs can model language acquisition through statistical learning without relying on innate grammar mechanisms.
- LLMs demonstrate the ability to process complex grammatical structures, even outperforming humans in specific recursive tasks.
- These findings challenge traditional assumptions about human-specific language faculties.

LLMs as Cognitive Models : Emerging Cognitive Capacities

- Research highlights emergent cognitive abilities in LLMs such as affordance recognition, theory of mind, and logical reasoning, despite training on next-word prediction.
- LLMs have achieved human-level performance across various cognitive tasks, lending support to associationist theories of cognition.
- Some scholars suggest parallels between LLM mechanisms and human memory theories, including Tulvingâs model, raising questions about consciousness and cognitive emergence.

LLMs as Cognitive Models : Ongoing Debates

- While LLMs offer new insights into human cognition, concerns remain about whether they truly capture human-like abstractions.
- Researchers highlight the need for caution in interpreting LLM outputs, given the architectural and learning process differences between models and the human brain.
- Methodologies such as Representational Similarity Analysis (RSA) have revealed how scaling and training data impact LLMâbrain alignment.

Insights from LLMs for Cognitive Science

- LLMs have prompted critical reflections on cognitive theories, challenging rule-based models but requiring careful interpretation.
- Scholars caution against anthropomorphising LLMs and stress the need for philosophical precision when relating AI to human cognition.
- The debate continues on whether LLMs are computational analogues of human language processing or functionally distinct systems.

LLMs Informing Broader Cognitive Research

- LLMs have been used as simulated economic agents to replicate behavioural economics experiments, offering cost-effective research tools.
- Frameworks have been developed to evaluate the cognitive plausibility of different language models based on learning mechanisms and data exposure.
- Some researchers propose that LLMs can serve as scientific models of external languages, informing sociolinguistics and cultural cognition.

Extending the Scope of Cognitive Science

- LLMs challenge conventional boundaries of cognition and encourage interdisciplinary approaches to intelligence studies.
- Researchers have used LLMs to explore cultural distinctions, leveraging statistical regularities in training data to model cultural cognition.
- LLMs enable large-scale analysis of language in psychological research, although ethical challenges must be addressed.

Applications in Specific Cognitive Domains : Causal Reasoning

- LLMs are being applied in causal inference tasks, enhancing reasoning capacity and supporting fairness and safety in decision-making.
- Benchmarks show that LLMs can outperform existing methods in generating causal arguments, though challenges remain in high-stakes applications.

Applications in Specific Cognitive Domains : Semantics and Creativity

- Studies in lexical semantics reveal that LLMs can model subtle sense distinctions and novel sense combinations, offering new perspectives on semantic theory.
- In creative writing, LLMs support translation and review tasks but often fall short in nuanced planning and originality, highlighting both promise and limitations.

Conclusion : Applications of LLMs in Cognitive Science

- LLMs have demonstrated significant potential in modelling language, reasoning, and decision-making, and in informing cognitive theories.
- Their versatility across domains makes them valuable tools, yet interpretability and ethical concerns must be actively addressed.
- Continued interdisciplinary research will be essential for aligning LLMs more closely with human cognitive processes and ensuring responsible use in cognitive science.

Overview : Limitations and Improvement of LLMs

- The rapid advancement of Large Language Models (LLMs) requires critical evaluation of their cognitive limitations and biases.
- This section examines key constraints in LLM performance and summarises recent efforts aimed at enhancing their capabilities and human alignment.
- The goal is to develop more reliable, cognitively aware, and robust AI systems.

Cognitive Biases and Limitations of LLMs (I)

- LLMs have shown limitations in Theory-of-Mind tasks, with studies highlighting failures on simple task variations that humans find trivial.
- Cognitive biases observed in LLMs include framing effects, overconfidence, and underconfidence, similar to well-documented biases in human reasoning.
- Some researchers suggest that LLMs display a "confidenceâcompetence gap" that mirrors the DunningâKruger effect in human judgement.

Cognitive Biases and Limitations of LLMs (II)

- LLMs have been criticised for lacking deeper linguistic and cognitive understanding, resulting in incomplete or biased language representations.
- Comparative evaluations reveal inconsistencies and irrationalities in LLM responses that differ from human reasoning patterns.
- Methods inspired by cognitive bias theory have been proposed to systematically identify predictable, high-impact errors in model outputs.
- Some scholars argue for using the term "confabulation" rather than "hallucination" to better conceptualise LLM-generated inaccuracies.

Improving LLM Performance : Cognitive-Inspired Methods

- Cognitive science-inspired approaches have been proposed to address LLM limitations, such as the bounded pragmatic speaker model to guide language generation.
- Models like CogGPT integrate iterative cognitive mechanisms to enhance role-specific cognitive dynamics and improve information processing over time.
- Chain-of-thought prompting, informed by probabilistic models, has been shown to improve metaphor understanding and paraphrasing abilities.

Improving LLM Performance : Aligning with Human Cognition

- Training LLMs on summarisation tasks has improved alignment with human brain activity, indicating enhanced language comprehension.
- Researchers advocate combining data-driven approaches with domain knowledge and introducing inductive biases to improve robustness.
- Addressing shortcut learning and model fragility remains a critical priority in advancing human-aligned AI systems.

Conclusion : Addressing Limitations and Enhancing Capabilities

- Addressing cognitive biases and improving the reliability of LLMs is essential for their integration into human-centric applications.
- Continued research must focus on evaluation techniques, cognitive alignment, and ethical development practices.
- The future of LLM development lies in deeper integration with cognitive science to produce AI systems that emulate more robust and nuanced human cognition.

Overview : Synergising LLMs and Cognitive Architectures

- Recent research explores the integration of Large Language Models (LLMs) with cognitive architectures to enhance AI systems.
- This integrated approach aims to leverage the complementary strengths of both systems, while mitigating their individual limitations.
- Integration methods vary in complexity and purpose, from modular frameworks to neuro-symbolic reasoning systems.

Integration Approaches and Frameworks

- Romero et al. proposed three integration approaches : modular, agency-based, and neuro-symbolic, each supported by distinct theoretical foundations.
- Kirk et al. demonstrated a six-step knowledge extraction method using template-based prompting and natural-language interaction with GPT-3.
- Joshi and Ustun integrated LLMs into cognitive architectures such as Soar and Sigma, treating them as prompt-able declarative memory systems.
- GonzÃ¡lez-Santamarta et al. used LLMs within the MERLIN2 architecture for autonomous robots, enhancing reasoning and humanârobot interaction.

Applications and Benefits of Integration

- Zhu and Simmons showed that combining LLMs with cognitive architectures improved task efficiency and reduced token requirements in kitchen task execution.
- Nakos and Forbus enhanced natural language understanding by integrating BERT into the Companion cognitive architecture, improving disambiguation and fact plausibility.
- Wray et al. outlined a strategy for integrating LLMs into cognitive agents to enhance task learning, performance, and knowledge generalisation.
- Zhou et al. proposed a Cognitive Personalised Search (CoPS) model, combining LLMs with memory mechanisms inspired by human cognition to improve user modelling and search relevance.

Challenges and Future Directions

- Key challenges include ensuring the accuracy, relevance, and interpretability of knowledge extracted from LLMs.
- Computational costs and scalability constraints remain critical considerations in large-scale integration.
- Future research should explore more advanced integration techniques, improve LLM-based reasoning efficiency, and expand applications across domains.
- Addressing the limitations of both LLMs and cognitive architectures will be crucial to building more robust and adaptable intelligent systems.

Discussion : Summary of Findings

- The intersection of Large Language Models (LLMs) and cognitive science marks a transformative moment in both artificial intelligence and our understanding of human cognition.
- This review has highlighted major advances in comparing LLMs to human cognitive processes, evaluating their cognitive abilities, and exploring their potential as cognitive models.
- While LLMs demonstrate notable similarities with human cognition, important differences persist and warrant further investigation.

Discussion : Cognitive Similarities and Differences

- LLMs exhibit human-like behaviours in language processing, logical reasoning effects, and sensory judgement across multiple modalities.
- These similarities extend to the neural level, with larger LLMs producing brain-like representational patterns.
- However, humans continue to outperform LLMs in reasoning tasks, especially under out-of-distribution and novel problem conditions.
- LLMs also display limitations in functional linguistic competence, highlighting the need for deeper, context-aware understanding.

Discussion : Enhancing Cognitive Capabilities

- Future research should prioritise improving LLM generalisation and reasoning in constrained and unfamiliar contexts.
- Enhancing functional linguistic competence is essential for more realistic cognitive modelling.
- Bridging the gap between surface-level fluency and deeper semantic understanding remains a critical goal.

Discussion : LLMs as Cognitive Models

- Fine-tuned LLMs can outperform traditional cognitive models in certain behavioural tasks, offering promising tools for cognitive science.
- These models allow for the investigation of individual variability and underlying cognitive mechanisms.
- However, fundamental differences in architecture and learning mechanisms necessitate cautious interpretation.

Future Challenge : Research Priorities

- Future research must focus on aligning LLMs more closely with human cognitive processes, both in architecture and evaluation methodology.
- Cross-disciplinary integration between cognitive science and AI engineering will be crucial to achieving this alignment.
- More diverse and comprehensive cognitive benchmarks are needed to assess the full spectrum of human-like cognition in LLMs.

Future Challenge : Domain-Specific Applications

- LLMs have shown promising applications in fields such as causal reasoning, lexical semantics, and creative writing.
- Further development of domain-specific models may improve their cognitive realism and task relevance.
- Specialised LLMs could enable more precise modelling of cognition in specific disciplines or problem areas.

Future Challenge : Addressing Cognitive Biases

- LLMs reflect biases analogous to those in human cognition, offering both challenges and research opportunities.
- Developing mitigation strategies is essential for ethical and accurate AI deployment.
- Controlled studies of LLMsâ biases can yield insights into the cognitive biases inherent in human reasoning.

Future Challenge : Integrating with Cognitive Architectures

- Integration of LLMs with cognitive architectures offers a pathway toward more robust, efficient, and generalisable AI systems.
- Future work should aim to refine integration techniques and optimise LLM-based reasoning within cognitive systems.
- Applications in real-world environments will test the scalability and adaptability of these integrated models.

Conclusion : Towards Human-Centric AI and Cognition

- The convergence of LLM research and cognitive science offers profound potential to advance understanding of both artificial and human intelligence.
- A balanced perspective is essentialâcelebrating capabilities while critically addressing limitations.
- Future development must be guided by ethical reflection, rigorous empirical testing, and interdisciplinary collaboration.