Using Human Brain Insights to Tackle LLM Challenges

Aakash Roy

aakashroyiitb@gmail.com

31st October, 2024

The human brain stands as our most complete example of artificial general intelligence (AGI), marked by unparalleled efficiency and speed in cognitive processing. Its distinct architecture offers critical advantages over large language models (LLMs) in reasoning, memory retrieval, and learning efficiency. Analyzing these differences provides insight into how AI systems can be optimized to better align with human cognition and performance.

Reasoning: Adaptive Intelligence vs. Pattern Matching

Human reasoning leverages adaptive, contextual insights, drawing on vast experiential networks and abstract principles to infer and deduce from limited data. In contrast, LLMs rely on statistical associations from massive datasets, often lacking the capacity to generalize outside their training scope. Apple's research highlights this limitation, showing that LLM performance is sensitive to even slight variations in questions, as seen in the GSM-Symbolic benchmark[1]. Here, changing numerical values caused substantial accuracy declines, revealing LLMs' reliance on memorized patterns rather than authentic logical reasoning. This discrepancy becomes more pronounced with the inclusion of irrelevant clauses, causing performance drops up to 65% due to the absence of genuine inferential capabilities.

Theory of Mind (ToM) is a critical component of human reasoning, encompassing the ability to understand others' intentions, emotions, and beliefs. Research from the University of Washington demonstrates that while LLMs excel in specific tasks, such as summarization and question answering, they often struggle with open-ended questions requiring ToM[2]. While LLM-generated responses may appear coherent, they often lack the depth and contextual insight of human responses, reflecting a fundamental gap in their ability to simulate genuine ToM reasoning.

The brain also employs analogical reasoning and trial-and-error strategies. When approaching a new problem, it can draw parallels with previous experiences and iteratively refine its approach based on feedback, leading to highly effective problem-solving skills. LLMs, while encoding language associations similar to the brain's association cortices, lack the goal-directed learning processes present in human cognition, which are guided by reward prediction systems. This divergence results in LLMs lacking the goal-oriented behaviour that characterizes human problem-solving.

Information Structuring and Memory Retrieval

The brain's organizational structure facilitates rapid information retrieval through a hierarchical framework of associative links, enabling humans to situate new data within broader networks of meaning. This hierarchical organization contrasts with the LLM's data storage in high-dimensional matrices, which limits its ability to dynamically contextualize information. The brain's memory retrieval process is highly selective, employing associative recall and contextual retrieval cues, which help prioritize relevant information and avoid unnecessary details.

Research indicates that human memories are encoded in synaptic connections, particularly within the dorsal pallium, where new synapse formation and loss, rather than simple adjustment of existing synapses, are fundamental to memory encoding[3]. This selective memory mechanism allows the brain to efficiently retrieve and organize pertinent information, a feature currently absent in LLMs, which rely on direct pattern matching across vast datasets, often resulting in slower and less accurate retrieval.

Efficiency and Parameter Utilization

The brain, with its approximately 86 billion neurons and trillions of synaptic connections, exemplifies parameter efficiency, achieving complex cognition with vastly fewer parameters than the trillions required by LLMs. Brain imaging studies reveal task-specific activations in distinct brain regions, such as those for mathematical reasoning, language, and visual processing, underscoring the brain's compartmentalized functionality. Neural networks like ChatGPT, however, exhibit broad, undifferentiated activation across the entire network regardless of task. This lack of specialized regions suggests inefficiency in comparison to the brain's targeted signal processing, where selective activation and inhibition mechanisms prioritize critical information, optimizing cognitive performance with minimal energy[4].

Eric Kandel's study on Aplysia shows that long-term memory formation involves gene expression and neuronal growth, including increased excitability, new synapses, and more neurotransmitter storage. This process requires a target neuron, linking learning with neural development. During memory formation, cell-adhesion molecules, which normally limit growth, are briefly inhibited to allow synaptic strengthening[5].

To bridge this gap, bio-inspired spiking neural networks (SNNs) implemented on neuromorphic chips use spiking neurons that fire only at specific thresholds, mimicking the brain's efficient signalling [6]. Unlike traditional networks, SNNs leverage multiscale synaptic plasticity to adapt dynamically, concentrating resources on relevant data and filtering noise. This selective activation aligns closely with the brain's approach, offering a path toward energy-efficient AI that approximates human-like processing.

LLMs also demand large, high-quality datasets, whereas the brain excels at learning from limited, noisy data, utilizing error-correcting feedback and inference to construct robust mental models. This difference underscores the brain's superior adaptability and efficiency in processing information with minimal computational resources.

Conclusion

To advance AI systems toward human-like cognition, researchers can incorporate principles inspired by brain architecture. This might involve integrating hierarchical processing, memory-based context retrieval, and dynamic synaptic-like updates, potentially enhancing AI's contextual awareness, inferential capabilities, and adaptive learning. For instance, applying concepts from self-organizing maps and cause-effect relationships could improve transparency and generalization, helping mitigate issues like overfitting and resource inefficiency.

The unique architecture of the human brain provides a roadmap for addressing current challenges in LLM development. By emulating its efficient processing, hierarchical organization, and adaptive learning strategies, future AI models may achieve enhanced reasoning, memory, and learning efficiencies, aligning them more closely with the robust and nuanced processing observed in human cognition. The integration of brain-inspired principles could ultimately foster AI systems capable of more sophisticated reasoning, improved resource use, and a closer alignment with human cognitive processes, marking a significant step toward human-like AGI.

References

[1] Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S., & Farajtabar, M. (2024, October 7). *GSM-Symbolic: Understanding the limitations of mathematical reasoning in large language models*. arXiv.org. https://arxiv.org/abs/2410.05229v1

[2] Amirizaniani, M., Martin, E., Sivachenko, M., Mashhadi, A., & Shah, C. (2024, October). Can LLMs Reason Like Humans? Assessing Theory of Mind Reasoning in LLMs for Open-Ended Questions. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management* (pp. 34-44). https://doi.org/10.1145/3627673.3679832

[3] Don Arnold, "Where Are Memories Stored in the Brain? New Research Suggests They May Be in the Connections Between Your Brain Cells," The Conversation. Available at: https://theconversation.com/where-are-memories-stored-in-the-brain-new-research-suggests-they-may-be-in-the-connections-between-your-brain-cells-174578.

[4] Jimmy Tidey, "New Research Is Beginning to Unravel How LLMs Think. What Are the Practical Consequences?" Medium, Available at: https://jimmytidey.medium.com/can-you-do-a-brain-scan-on-an-llm-anthropic-and-openai-are-trying-7bc82ef7a2a4.

[5] Ackerman S. Discovering the Brain. Washington (DC): National Academies Press (US); 1992. 8, Learning, Recalling, and Thinking. Available from: https://www.ncbi.nlm.nih.gov/books/NBK234153/

[6] Xu B, Poo MM. Large language models and brain-inspired general intelligence. Natl Sci Rev. 2023 Nov 3;10(10):nwad267. doi: 10.1093/nsr/nwad267. PMID: 37942481; PMCID: PMC10630093.