Review of: "[Commentary] Are Large Language Models obsolete? And are multimodal solutions inappropriate?"

Shahid Butt[1]

1 National University of Science and Technology

Potential competing interests: No potential competing interests to declare.

Conclusion recommends, "principles of lateralization, sensory-motor integration, gestural syntax, and embodied cognition, these agents can achieve more human-like interactions, better adaptability, and more effective learning capabilities." These improvements, if implemented, can really support a more realistic artificial intelligence environment, which can be close to the human response and also save much time for humans.

Specific examples of how principles like sensory-motor integration could be applied in AI design: for example, if a real-life problem where SWOT analysis is required arises, this technique can give some guidance to make a calculated decision.

The proposed dual-network architecture, which incorporates two distinct neural networks (one for LLM and one for gestural language) with a metaphorical translation apparatus between them, might address current limitations of LLMs in several significant ways:

1. **Enhanced Multimodal Processing**

Current Limitation: LLMs primarily process text and struggle with integrating other modalities like gestures, images, or sounds. **Proposed Solution:** By having separate networks specialized for different modalities (LLM for text and another for gestures), the system can more effectively process and integrate information from diverse sources. This multimodal approach can lead to richer, more nuanced interactions and understanding.

2. **Improved Contextual Understanding**

Current Limitation: LLMs often lack the ability to understand context derived from non-verbal cues, which is crucial in human communication. **Proposed Solution:** The gestural network can process and interpret body language, facial expressions, and other non-verbal cues. When combined with textual information through the translation apparatus, this can provide a more comprehensive contextual understanding, leading to more accurate and human-like responses.

3. **Creative Problem Solving and Error Handling**

Current Limitation: LLMs can generate plausible-sounding but incorrect or nonsensical answers, and they have limited creativity in generating novel ideas. **Proposed Solution:** The metaphorical translation apparatus and the interplay

between the two networks allow for the reuse of input-output cycles in novel ways. This can lead to machines that are not only capable of making errors but also capable of learning from them and generating creative solutions, much like humans do when they encounter and solve new problems.

4. **Embodied Cognition and More Natural Interactions**

Current Limitation: LLMs do not have a physical presence or an understanding of the physical world, limiting their ability to interact naturally with humans. **Proposed Solution:** The integration of a gestural language network brings in elements of embodied cognition, where the AI's understanding is grounded in physical interactions and sensory experiences. This leads to more natural and intuitive interactions, as the AI can respond to and interpret gestures and physical cues more effectively.

5. **Self-Awareness and Adaptability**

Current Limitation: LLMs lack self-awareness and adaptability in physical interactions.**Proposed Solution:** The differentiation in sensory-motor perception and conscious interoception within the gestural network can lead to a form of self-awareness. This self-awareness allows the AI to better understand its physical state and interactions, leading to more precise and adaptable responses in dynamic environments.

6. **Flexible Communication Models**

Current Limitation: LLMs have rigid communication models, often failing to adapt to different communication styles or partial information. **Proposed Solution:** By recognizing and fragmenting gestures into meaningful units and understanding them paradigmatically, the dual-network architecture can develop more flexible communication models. This allows the AI to adapt to different gestures and communication styles dynamically, improving its overall communicative effectiveness.

7. **Grounded Language Processing**

Current Limitation: LLMs process language in a largely abstract manner, disconnected from sensory experiences. **Proposed Solution:** The integration of language processing with sensory-motor experiences ensures that the AI's language capabilities are grounded in real-world interactions. This leads to more natural language use and a better alignment with human communication patterns.

8. **Holistic Learning Systems**

Current Limitation: LLMs have limited capability to integrate learning across different sensory modalities.**Proposed Solution:** Concepts like cross-modality, embodied cognition, and mirror neurons can be incorporated into the dual-network system. This creates a more holistic learning system where the AI can learn by imitation and integrate information

across different sensory modalities, leading to a deeper and more versatile understanding.

By addressing these limitations, the proposed dual-network architecture holds the potential to significantly enhance the capabilities of LLMs, making them more robust, adaptable, and aligned with human-like processing and interaction.