Qeios

Peer Review

Review of: "Is DeepSeek a Metacognition AI?"

Sunny, Chi Lik Au[1]

1. Tung Wah Eastern Hospital, Hong Kong, China

The article builds a progressive case: DeepSeek's RL-driven innovations lead to behaviors that resemble metacognition. The logic is generally coherent, moving from context to technical details to conclusions. However, it relies heavily on analogy rather than empirical or theoretical proof that these behaviors constitute metacognition. The absence of a critical distinction between advanced reasoning and true metacognitive awareness (e.g., consciousness) undermines the diagnosis, making it more speculative than definitive.

The author assumes DeepSeek's behaviors (e.g., pausing, optimizing) equate to metacognition without proving they involve conscious self-awareness. It presents DeepSeek's advancements positively but doesn't question whether these are truly metacognitive or just sophisticated optimization. While readability issues are noted, other potential drawbacks (e.g., scalability, biases) are underexplored.

Suggestions for Improvement

Clarify the Definition of Metacognition:

- **Suggestion**: Refine the definition by integrating cognitive science perspectives to distinguish between algorithmic self-optimization and human-like metacognition involving consciousness.

Expand Discussion of Limitations:

- **Suggestion**: Detail challenges like computational scalability, ethical risks (e.g., unintended biases), or limitations in generalizing RL-driven reasoning.

Enhance Educational Implications:

doi.org/10.32388/87KSSF

- **Suggestion**: Propose specific educational strategies (e.g., teaching metacognitive reflection via problem-solving exercises) to complement the conclusion.

Address Counterarguments:

- **Suggestion**: Engage with critiques, such as the view that AI lacks subjective awareness and thus cannot achieve true metacognition, and offer rebuttals or concessions.

Declarations

Potential competing interests: No potential competing interests to declare.