

Peer Review

Review of: "Is DeepSeek a Metacognition AI?"

Konda Mani Saravanan¹

1. Biotechnology, Bharath Institute Of Higher Education & Research, India

The study offers a coherent and compelling examination of DeepSeek's metacognitive abilities, contextualizing the analysis within the wider framework of reinforcement learning (RL) and its ramifications for AI autonomy. The conceptual connection between metacognition—historically regarded as a distinctly human trait—and the emerging behaviors in DeepSeek models is both insightful and persuasive.

Here are my observations aimed at enhancing your manuscript:

While the manuscript asserts that DeepSeek exhibits metacognitive-like behaviors, it may benefit from a more explicit framework or criteria for what constitutes metacognition in machines. For instance, how do we empirically differentiate between complex reasoning and true self-monitoring or self-awareness?

Consider adding a brief discussion on the potential limitations or counterarguments—for instance, concerns about anthropomorphizing machine behavior or the philosophical distinction between simulation and actual cognition.

The manuscript could be enriched by a more detailed comparison with other cutting-edge models (e.g., Gemini, Claude, GPT-4-turbo), especially regarding their metacognitive capabilities, if any.

Some readers may benefit from clearer definitions early in the paper (e.g., distinctions among RL, RLHF, PPO, and GRPO). Perhaps a short glossary or footnote system would be helpful.

While the references are current and relevant, including a few more peer-reviewed works from cognitive science and AI alignment literature could add depth and academic weight to the metacognitive discussion.

There are a few typographical inconsistencies (e.g., broken hyphenations and formatting in citations) likely due to PDF rendering—these can be resolved in the final editing stage.

You may wish to double-check and standardize terms like "self-improvement" and "self-reflection" throughout the paper for consistency.

Declarations

Potential competing interests: No potential competing interests to declare.