v1: 14 May 2025

Commentary

Is DeepSeek a Metacognitive AI?

Peer-approved: 14 May 2025

Qeios, Vol. 7 (2025)
ISSN: 2632-3834

Ronaldo Mota[1]

1. Chair in Artificial Intelligence, Brazilian College of Advanced Studies (CBAE), Universidade Federal do Rio de Janeiro, Brazil

The relationship between metacognition and DeepSeek models represents a compelling and yet underexplored area of research. Metacognition refers to a system's capacity to monitor and comprehend its own cognitive processes, which includes the active regulation and adjustment of its procedures. In the DeepSeek-R1 and DeepSeek-R1-Zero, it is evident that the interactions between the system's monitoring and control processes are both present and crucial for achieving the coherent and often surprising levels of reasoning that define these models. In particular, the "aha moment" is discussed as the best example of a sort of metacognitive behaviour in the DeepSeek models. In this sense, DeepSeek could be paving a new avenue for reasoning in Artificial Intelligence (AI) by prioritizing reinforcement learning (RL) over the more conventional approach of supervised fine-tuning (SFT). This study aims to analyse the implications of such innovations for machines' abilities to simulate behaviours based on self-reflection and to act accordingly. We will explore the extent to which these elements can be associated with metacognition, a trait traditionally considered to be uniquely human. Furthermore, from an educational standpoint, the significance of the relationship between advancements in DeepSeek and metacognition is highlighted, particularly in relation to the importance of prioritizing metacognitive approaches in education.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Considerations on Large Language Models (LLMs)

OpenAI's ChatGPT was made available to the public at the end of 2022 [1]. Since then, LLMs have attracted everyone's attention and pointed towards a new phase of AI-driven entities. More recently, at the beginning of 2025, DeepSeek from the Chinese company Hangzhou DeepSeek AI Co. surprised the AI research community again [2]. These two recent moments have elevated discussions and concerns about the "new generation of AIs" to a new level.

Regarding DeepSeek, initially, one of the main observed advantages was that the training cost was significantly lower compared to ChatGPT. Additionally, it impressed with its performance in the tests it underwent and revolutionized the field by publicly releasing the code of its models, reinforcing the start-up's commitment to open-source AI.

However, the greatest originality of the DeepSeek system was yet to be revealed more clearly: an advancement in what we call possible metacognitive prerogatives. In this case, it is associated with moments when the model, as programmed, develops a pause, reevaluates, and optimizes its problem-solving approach, which has been conventionally termed an "aha moment" [3]. If confirmed to the depth suspected, we are facing a phenomenon previously considered exclusive to human reasoning. This marks a significant advancement resulting from what is called RL, greatly amplifying AI capabilities [4].

 doi.org/10.32388/PJ3POM.2

In other words, by integrating RL techniques, DeepSeek goes beyond static and pre-programmed responses and actively learns, in a differentiated manner, through testing and system feedback. This self-improvement mechanism allows the model to recognize when an initial approach can and should be optimized, leading to adjustments that enhance performance, adaptability, and reasoning capacity.

During training, DeepSeek reportedly demonstrated a fundamental behavioural shift: instead of following a fixed sequence of procedures and calculations, it allocated more resources to complex problems, exhibiting a self-awareness reminiscent of human thought processes, simulating something we might classify as metacognition [5].

This behaviour suggests that DeepSeek was not merely processing information but was actively engaged in the ability to reflect on its own problem-solving strategy and refine it appropriately. Researchers attribute this advancement to its RL structure, which optimizes decision-making processes based on past experiences rather than relying solely on pre-trained patterns. Thus, DeepSeek is characterized by an approach where multiple methods are simultaneously adopted, learning to solve problems using different approaches to confirm the answer. Furthermore, the model naturally learned to break down complex problems into smaller, verifiable steps, enabling the development of increasingly sophisticated reasoning chains. This also involves pattern recognition through the identification of parts of the problem that had been previously solved.

This combination of strategies allows for a hierarchical breakdown, that is, the decomposition of complex problems into simpler and more manageable parts. Coupled with initially adopted alternative approaches, it results in the consideration of various paths in search of the most efficient solution. In summary, the model learned to initially solve a simpler version of the problem, then identify extreme cases, subsequently generalize the solution, and finally optimize the implementation.

In essence, DeepSeek could be paving a new path for AI reasoning, emphasizing RL rather than the more traditional SFT approach, demonstrating that AI can learn reasoning without explicit human guidance. To what extent do these processes resemble the metacognitive capacity, which is presumed to be exclusively human?

2. On Metacognition

To explore this inquiry, let us delve a bit deeper into what metacognition entails. There are several definitions, the most general being: “a type of higher-order thinking in which the thinker has active control over the process.” Self-regulation, knowledge, monitoring, evaluation, and awareness of one’s own mental activities are essential elements of metacognition [6]. The term metacognition was coined by researcher John Flavell in[7], referring to the process of thinking about one’s own thinking and learning [7].

The mental action or process of acquiring knowledge and understanding through thought, experience, and senses is known as cognition. Thus, metacognition transcends cognition. Therefore, the hierarchical nature of the psychological processes involved in cognition places metacognition at the top, referring to the processes that supervise, manage, and coordinate cognitive activities. The relevance of researching to what extent AI entities might be capable of exploring metacognitive predicates, or not, arises from the fact that in the “competition” between humans and machines, mastery of metacognition seems crucial.

Some scholars argue that humans dominate three main predicates: physical strength, cognition, and metacognition [8]. Regarding physical strength, nearly three centuries ago, with the advent of James Watt's steam engine, machines began to surpass humans with significant advantages. In terms of cognition, the victory of Deep Blue/IBM over Garry Kasparov in 1997 [9], as well as that of AlphaZero/Google over Stockfish 8 in 2017, are emblematic events attesting that, in terms of simple cognition, machines, like physical strength, reaffirmed the trend of surpassing humans in this cognitive field as well. What remains for humans, as the final frontier, is metacognition, potentially serving as a differentiated competitive advantage [10].

It is noteworthy that in 2017, the significant novelty was that, unlike Stockfish 8, which had accumulated centuries of prior chess experiences along with other operational predicates, AlphaZero was characterized by having learned from scratch. In other words, it learned solely by playing against itself, utilizing the basic principles of machine self-learning. It is astonishing that AlphaZero transformed from a complete amateur to the best chess player in just four hours, entirely dispensing with any direct human collaboration or even other machines throughout its learning process.

3. DeepSeek-R1 and DeepSeek-R1-Zero

Returning to the present day, as previously noted, we are witnessing a significant advancement in AI with the launch of DeepSeek-R1-Zero and DeepSeek-R1, two models that disrupt the traditional paradigm of AI training.

The key differentiator of DeepSeek-R1-Zero is that it does not rely on human feedback to improve its responses. In other words, in contrast to the application of SFT, DeepSeek-R1-Zero was trained solely through RL, unleashing the power of reasoning through self-improvement. For years, AI training followed a standard script: start with SFT and continue with optimization through RL. DeepSeek-R1-Zero reverses this script by: i) completely ignoring SFT and training through pure RL; ii) allowing the model to develop its reasoning skills through a reward-driven mechanism; and iii) exhibiting emergent behaviours such as self-verification and reflection, without any instance of human curation. In summary, the main conclusion is that it may be possible for AI models to develop reasoning without human assistance, simply through exposure to their own environment [11].

In turn, DeepSeek-R1 adopts a hybrid approach based on DeepSeek-R1-Zero, but with an initial cold-start SFT phase. That is, it anchors itself in high-quality Chain-of-Thought (CoT) data before RL. Thus, by going through several stages of RL and rejection sampling, it ultimately achieves performance comparable to OpenAI's model o1-1217 in reasoning tasks. Another highlight of DeepSeek-R1 is its distillation power, which promotes reduction without sacrificing performance; that is, a technique for transferring its reasoning capabilities to smaller models. The relevance of this process is that smaller models are more efficient but generally lack reasoning capabilities. Researchers have demonstrated that knowledge distillation from a large model trained in RL significantly improves the performance of smaller models.

These concerns must be discussed together with the recent findings from Anthropic's paper *Reasoning Models Don't Always Say What They Think*, by Chen et al., [12] [12], which shows that chain-of-thought reasoning often fails to reflect the actual computations behind the model's answers. Besides that, in human psychology, metacognition involves both calibrated monitoring, knowing when one is likely right or wrong, and strategic control, choosing how to act based on that awareness.

Additionally, it is interesting to observe that, recently [13], OpenAI released OpenAI o3 and o4-mini, the latest in their o-series of models trained to think for longer before responding. These are the smartest models they've released to date, representing a step change in ChatGPT's capabilities for everyone from curious users to advanced researchers. For the first time, in accordance with them, their reasoning models can use and combine every tool within ChatGPT. Critically, these models are trained to reason about when and how to use tools to produce detailed and thoughtful answers in the right output formats, typically in under a minute, to solve more complex problems. This allows them to tackle multi-faceted questions more effectively, a step toward a more agentic ChatGPT that can independently execute tasks on your behalf. The combined power of state-of-the-art reasoning with full tool access translates into significantly stronger performance across academic benchmarks and real-world tasks, setting a new standard in both intelligence and usefulness. The improvements are on the way to continue to scale reinforcement learning. OpenAI have trained both models to use tools through reinforcement learning, teaching them not just how to use tools, but to reason about when to use them. Their ability to deploy tools based on desired outcomes makes them more capable in open-ended situations, particularly those involving visual reasoning and multi-step workflows. This improvement is reflected both in academic benchmarks and real-world tasks, as reported by early testers.

In conclusion, the next generation of AI training, DeepSeek-R1-Zero and DeepSeek-R1, as well others following the same paths, are paving a new path for AI reasoning. By emphasizing RL instead of SFT, the models demonstrate that AI can learn reasoning without explicit human guidance.

4. RL: The Ace in the Hole

As previously discussed, RL has been instrumental in advancing the reasoning capabilities of DeepSeek, thereby enhancing logical thinking and problem-solving in AI. As clearly demonstrated by Abdallah [14], RL is typically employed to align model outputs with desired behaviours, such as factual accuracy, logical consistency, and human-like reasoning. This methodology can significantly impact the training of models tasked with solving complex reasoning challenges, including mathematical problems, code generation, and logical inference.

In general, the training of LLMs involves various levels of supervised learning, where models learn from labelled datasets. For example, ChatGPT utilizes Reinforcement Learning from Human Feedback (RLHF), a machine learning technique in which a reward model is trained using direct human feedback and subsequently employed to optimize performance through RL. One of the most widely used RL algorithms for optimizing LLMs, developed by ChatGPT, is Proximal Policy Optimization (PPO). This algorithm aids in balancing performance and training stability while preventing drastic updates that could destabilize the learning process. However, PPO has several notable limitations, including high computational costs, dependency on reward models, instability in long sequences, and challenges related to the exploration-exploitation trade-off. Specifically, the balance between exploration—attempting new responses—and exploitation—refining known responses—can ultimately hinder the model's ability to discover new reasoning patterns.

In response to these limitations, DeepSeek has developed Group Policy Optimization (GRPO) as a more effective alternative for mathematical reasoning. GRPO eliminates the need for a separate critic network by estimating baseline rewards from grouped outputs, thereby reducing computational demands and maintaining consistent training signals. In contrast to PPO, which generates a single response for each input prompt, GRPO produces a set of outputs. These multiple responses are ranked within the sample group, and rewards are assigned based on comparative quality, ensuring that only the highest-quality outputs receive reinforcement. Such modifications lead to more stable and effective improvements.

DeepSeek-R1 adopts the GRPO strategy by employing pure RL to enhance cognitive tasks. Unlike traditional approaches that rely on SFT prior to RL, DeepSeek-R1 was trained exclusively using RL (DeepSeek-R1-Zero), fostering self-improving reasoning skills. However, this pure RL approach initially resulted in poor readability and language mixing. To mitigate these issues, a cold-start phase was introduced, incorporating a small amount of high-quality SFT before RL training to ensure more structured responses.

In summary, the application of RL in DeepSeek models represents a significant advancement in training LLMs for reasoning tasks. The adoption of GRPO, as opposed to PPO, enables these models to achieve greater efficiency, more stable training, and enhanced performance in mathematical problem-solving and logical reasoning. It is reasonable to anticipate that future developments in the use of RL and GRPO will lead to increasingly autonomous reasoning models capable of addressing more complex intellectual missions. This evolution will bring these systems closer to performing activities that simulate metacognitive behaviours.

Certainly, the best example of a sort of metacognitive behaviour in the DeepSeek models is the "aha moment," when the system learned that it needed to think more, there was an interesting moment during training. DeepSeek-R1-Zero learned to spend more time contemplating difficult problems, as if it were essentially learning to develop metacognition. This "aha moment" is a breakthrough because it demonstrates that RL-based training can result in self-improvement techniques that were previously thought to require human intervention [15]. It appears DeepSeek-R1-Zero has learned to spend more time thinking about a problem by reconsidering its initial strategy. It seems like if there is one moment when the system says: “Wait, wait. Wait. That’s an aha moment I can flag here”. This behaviour is a clear example of how RL can lead to unexpected and sophisticated outcomes. Then, in this sense, it is possible that RL can unlock new levels of intelligence in artificial systems, leading to more autonomous and adaptive models in the future.

5. Conclusions on DeepSeek and Metacognition

The relationship between metacognition and active control in the behaviour of the DeepSeek system is a compelling area of study. Metacognition appears to be integral to the system's ability to monitor and comprehend its own cognitive processes, as well as to the control mechanisms involved in the active regulation and adjustment of these processes. This is particularly evident in the DeepSeek-R1 model, which not only recognizes when its reasoning requires modification but also actively reallocates computational resources and alters its problem-solving strategies. In this way, the system demonstrates the capacity to inhibit initial responses, maintain focus on complex problems over extended sequences, and flexibly switch between various reasoning approaches. The interaction between monitoring and control processes seems fundamental to achieving coherent reasoning.

The emergence of metacognitive control through reinforcement learning (RL) indicates that it is a necessary component of coherent information processing, rather than merely an ancillary resource. A system must possess both awareness of its cognitive

processes and the ability to regulate them to maintain coherence in complex reasoning tasks. An even more striking demonstration of this capability was observed in DeepSeek-R1-Zero, an advanced iteration of the model. During the intermediate training phases, DeepSeek-R1-Zero exhibited an enhanced ability to dynamically allocate thinking time to problems, optimizing its responses in real-time. Rather than adhering to a rigid, rule-based training regimen, the system learned to autonomously adjust its problem-solving approach based on incentive structures. This implies that, instead of being explicitly programmed to recognize specific types of solutions, it received appropriate incentives and independently developed sophisticated reasoning strategies.

There is no doubt that the DeepSeek system represents a significant advancement in the ability of machines to simulate behaviours that involve reflecting on their own cognitive processes and acting accordingly. Such behaviours are typically associated with metacognitive predicates, which have traditionally been considered exclusive to humans. The extent to which machines may have surpassed humans in these characteristics remains an open question, as we currently lack sufficient evidence to draw definitive conclusions. However, it is undeniable that if machines ever achieve a state of metacognitive completeness, the DeepSeek models will have played a pivotal role in the historical trajectory marking the onset of this process, the timeline for which remains unpredictable.

Furthermore, one of the primary objectives of this article is to emphasize the importance of an education focused on metacognitive skills, having established a nascent connection between the recent advancements in DeepSeek and metacognition. From the perspective of future professional opportunities, mastery of metacognitive predicates, especially those not yet fully developed in AI-driven entities, may represent significant avenues for growth. In this context, elucidating the progress made by LLMs, particularly DeepSeek, should serve as a clarion call to educators and students regarding the priority and relevance of metacognitive educational approaches.

About the author

Ronaldo Mota holds the Chair in Artificial Intelligence at the Federal University of Rio de Janeiro at the Brazilian College of Advanced Studies CBAE/UFRJ.

References

1. ^ *OpenAI et al. (2023). "Technical Report on GPT-4." ar Xiv. https://arxiv.org/abs/2303.08774/.*
2. ^ *DeepSeek-AI (2025). "Technical Report on DeepSeek-V3." arXiv. https://arxiv.org/abs/2412.19437/.*
3. ^ *Heikkilä M (2025). "The 'Aha Moment' of DeepSeek Creates a New Way to Build Powerful AI with Less Mo ney." Financial Times. https://www.ft.com/content/ea 803121-196f-4c61-ab70-93b38043836e.*
4. ^ *Girotra A (2025). "How DeepSeek-R1 is Redefining AI Reasoning: The RL-First Paradigm Shift." LinkedIn. ht tps://www.linkedin.com/pulse/how-deepseek-r1-redef ining-ai-reasoning-rl-first-abhinav-girotra-zvrnc/.*
5. ^ *Gupta M (2025). "DeepSeek and Metacognitive Beha vior." Medium. https://medium.com/@manavg/deepse ek-and-meta-cognitive-behaviour-1c5cabedf632.*
6. ^ *Mota R (2019). "Learning to Learn is More than Lear ning." The Physics Educator. 1:1950001-1-6.*
7. a, b *Flavell JH (1979). "Metacognition and Cognitive Mo nitoring: A New Area of Investigation for Development al Psychology." American Psychologist. 34(10):906-91 1.*
8. ^ *Eysenck MW, Eysenck C (2023). Artificial Intelligence vs. Humans. Artmed Publishing.*
9. ^ *"Kasparov vs. Deep Blue: The Match that Changed Hi story." (2018). Chesscom. https://www.chess.com/articl e/view/deep-blue-kasparov-chess/.*
10. ^ *Goldemeier G, Mota R (2023). "Rationality and Scient ific Thinking as Foundations for Leadership in the Wo rkplace." Qeios. 50. https://www.qeios.com/read/BKHX OW.*
11. ^ *DeepSeek (2025). "DeepSeek-R1: Encouraging Reason ing Ability in LLMs via Reinforcement Learning." arXi v. https://arxiv.org/abs/2501.12948/.*
12. a, b *Chen Y, Benton J, Radhakrishnan A, Uesato J, Denis on C, Schulman J, et al. (2025). "Reasoning models do n't always say what they think." Anthropic. https://ww w.anthropic.com/research/reasoning-models-dont-sa y-think.*
13. ^ *Available at: https://openai.com/index/introducing-o 3-and-o4-mini/).*
14. ^ *Abdallah B (2025). "Understanding Reinforcement L earning in DeepSeek-R1." Medium. https://medium.co m/@la_boukouffallah/understanding-reinforcement-l earning-in-deepseek-r1-079d3360ca6c.*
15. ^ *Smith C. "Developers caught DeepSeek R1 having an 'aha moment' on its own during training." bgr.com. htt ps://bgr.com/tech/developers-caught-deepseek-r1-havi ng-an-aha-moment-on-its-own-during-training.*

Declarations

Funding: No specific funding was received for this work.
Potential competing interests: No potential competing interests to declare.