Advances in Consumer Research
https://acr-journal.com/
Volume-2 | Issue-6 | Dec 2025
Original Researcher Article

Meta-Cognitive Architectures For Self-Monitoring And Error Awareness In Artificial Agents

Abinaya. K[1], Dhamayanthi. P[2], Sivanathan. M[3], Dr. S. Sakthivel Padaiyatchi[4], D. Sujeetha[5], N. Tamilarasi[6]
[1]Assistant professor CSE Jai Shriram Engineering College
Email:ID: abinayakrishnadevarayan@gmail.com
[2]Assistant Professor Computer Science and Engineering Kgisl institute of technology
Email:ID: dhamayanthi.p@kgkite.ac.in
[3]Assistant professorInformation Technology Department,V. S. B. Engineering College, Karur, India
Email:ID: shivamohan2007@gmail.com
[4]Professor,Department of Electrical and Electronics Engineering, Nehru Institute of Engineering and Technology, Coimbatore, India
Email:ID: sithansakthi@gmail.com
[5]Assistant Professor Computer science and Engineering DepartmentNehru Institute of Engineering and Technology
Email:ID: sujeetha.venkatachalam@gmail.com
[6]Assistant Professor Department of CCE Nehru institute of Technology
Email:ID: nittamilarasi@nehrucolleges.com

ABSTRACT
N/A.

Keywords: N/A

INTRODUCTION:

Artificial intelligence (AI) has been expanding rapidly over recent years, thus rendering the creation of intelligent systems capable of handling complicated tasks in the sphere of robotics, medicine, finance, and autonomous systems. The more autonomous AI systems become, the more important it becomes to make sure that they monitor their performance and find potential errors. The standard artificial agents lack the ability to consider their internal operations and rely on predetermined algorithms and regulations. To address this weakness, researchers have offered meta-cognitive architectures that allow artificial agents to track, evaluate and regulate their mental activities.

Lastly, meta-cognition is rooted in cognitive psychology, and it is described as having the capacity to reflect upon the way one thinks. When used with artificial intelligence, it enables systems to trace internal operations, analyse task performance, and detect the deviations between the expected outcomes. Meta-cognitive architecture is therefore crucial to improving the reliability, adaptability and robustness of artificial agents, in dynamic and unpredictable environments.

2. Concept of Meta-Cognition in Artificial Intelligence

Meta-cognition was the initial concept in the field of psychology, and it involves the ability of a person to observe and regulate their thinking processes. It involves two major aspects and these are tracking mental activities and controlling the same activities according to the feedback. Meta-cognition allows human beings to evaluate personal learning, recognize mistakes, and make adjustments to improve the performance (Langdon et al., 2022). Such an idea has been actively pursued in the study of artificial intelligence to enhance the capabilities of the artificial agents.

In AI systems, meta-cognition implies that an artificial agent can observe and provide analysis of the internal decision-making processes. Unlike the rigid enforcement of the actions based on the issued rules or machine learning models, meta-cognitive systems can self-assess their performance and adjust their behaviour when needed (Sternberg, 2021). This will involve monitoring of task progress, detection of inconsistencies and taking corrective action.

As an indicator, an AI with meta-cognitive abilities can recognize when its predictions are not very certain or when a given strategy is not performing well (Han, 2025). It can then alter its tactics, seek greater information or adopt other tactics. By incorporating meta-cognitive processes, artificial agents will be more adaptive, aware and competent in enhancing their performance over time and need not be externally controlled.

3. Meta-Cognitive Architectures in Artificial Agents

Meta-cognitive architectures are structural frameworks by which artificial intelligence agents can monitor and manage their cognitive activities. These architectures tend to consist of multiple layers where each of them is executing different functions in the system. The most celebrated framework is composed of a low-level cognitive layer and a meta-cognitive layer.

It is the basic level of cognition, and it performs the primary functions of perception, reasoning, learning, and decision-making (Ukov and Tsochev, 2025). This layer deciphers information of the environment and generates actions based on programmed algorithms or learned

models. And even this layer by itself is not able to estimate whether its actions are effective or correct.

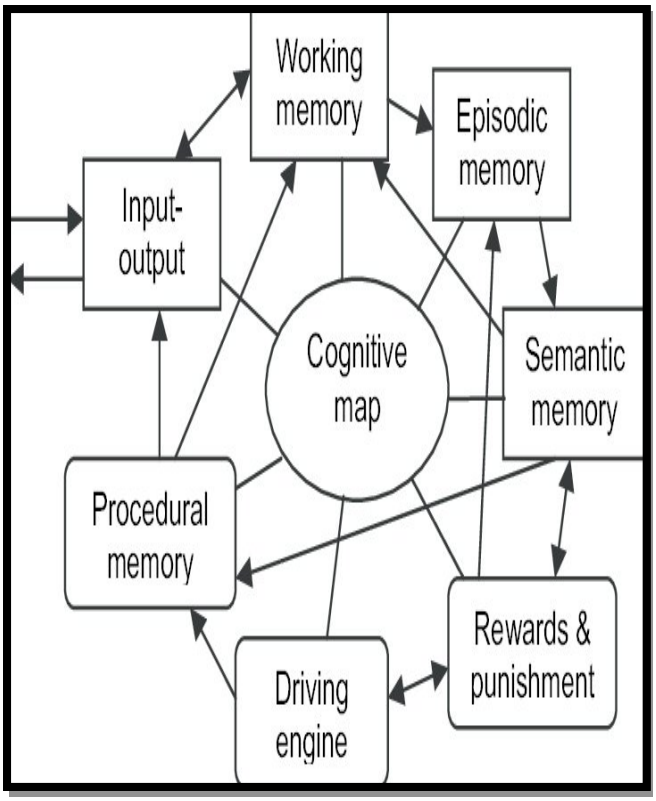

Figure 1: The-Hybrid-Cognitive-Architecture-At-A-Large-Scale-Shapes-Reflect-The-Nature

(Source: researchgate.net,2026)

The bottom cognitive layer will be overridden by the meta-cognitive layer which will be a supervisory system. It continually tracks the behaviour of the base layer and analyses performance measures, such as the tasks accomplished, accuracy and system confidence. In such a way, the meta-cognitive layer can identify the inefficiencies or potential errors in the decision-making process by means of monitoring and evaluation (Drigas et al., 2023).

The meta-cognitive system can provoke corrective actions when a problem is detected. These activities may include changing the parameters, selecting alternative policies, requesting additional information, and even stopping until the next study process is completed (Martinengo et al., 2025). The result of such stratified form is smarter customized behavior of artificial agents that have self-regulation and adaptation learning.

The meta-cognitive architectures come in handy in the ambiguous environment where the instability and dynamism are high order. Such architecture allows artificial agents to monitor and self-correct their behaviour resulting in a far more robust, high-performance, and long-term system performance.

4. Self-Monitoring Mechanisms in Artificial Agents

One of the fundamental roles of the meta-cognitive artificial agents is self-observing. The ability of a system to study and evaluate its own internal processes and output behaviour is the ability. Self-monitoring features enable artificial agents to keep track of their activity, evaluate their performance, and detect cases of performance deviation.

One of the primary techniques that is used in self-monitoring is the use of feedback loops. These loops continuously receive data about the system performance and compare it to already established goals or performance indicators (Du et al., 2025). Where discrepancies are realised, the system can take corrective measures to improve its behaviour.

Artificial agents may also rely on internal performance indicators, such as prediction confidence, error rates, and time to complete a task (Wang et al., 2026). Indicatively, in a machine learning system, the agent can have access to the precision of its predictions and be aware of the fact that it requires changing its model or retraining.

Self-monitoring is a very important factor in autonomous systems such as robots or self-driving cars. Here, agents must constantly examine sensor data, decision and environmental changes outcomes to ensure that the operation is safe and effective (Braun et al., 2024). In case the system can detect any abnormal patterns or inconsistencies, it can initiate diagnostic mechanisms or alter its behaviour to avoid risk.

By introducing self-monitoring capability, artificial agents can become more autonomous and trustworthy. This not only enables the system to perform better but also enhances transparency since the agent can offer explanations or insights into the process of internal decision-making in the system.

5. Error Awareness and Detection in Artificial Agents

The other significant aspect of the meta-cognitive architecture is error awareness. It is the ability of a virtual agent to understand that something has gone amiss or that a decision may have some undesired effects. Traditional AI systems are often identified with errors by human operators or post-processing evaluation. However, when provided with meta-cognitive systems, artificial agents can identify mistakes in their inner system and respond to them.

Anomaly detection, uncertainty estimation and predictive monitoring are among the mechanisms that can be used to detect errors (Yuttachai et al., 2024). Anomaly detection algorithms enable systems to identify unusual trends in data or behaviour that are not normal per expectation. Similarly, uncertainty estimation would help an agent to determine the precision of its predictions or decisions.

To illustrate, a medical diagnosis AI-based solution can test the accuracy of its predictions. In other words, the result can be tagged as uncertain by the system and human verification is needed when the confidence is near zero (Bao et al., 2023). Such error awareness reduces the risks of making erroneous decisions, besides enhancing the overall system safety.

The agents can also learn out of their mistakes and out of the meta-cognitive architecture. Should there be an error, the system can learn why an issue has taken place and

adjust its internal models or strategies. This lifelong learning improves future performance and reduces the possibility of repeating such mistakes.

The particular space of high error awareness is the area of autonomous vehicles and medical and financial decision-making systems (Bickley and Torgler, 2023).

6. Challenges and Limitations of Meta-Cognitive Architectures

On the one hand, meta-cognitive architectures are characterized by a few limitations and challenges despite their advantages. The former is more complexity in calculations required to set up monitoring and control mechanisms. More processing and system design complexity may be required by the introduction of meta-cognitive layers (Bao et al., 2023).

The other challenge is the modeling of meta-cognitive mechanisms that can be compared to the human self awareness (Bickley and Torgler, 2023). Artificial systems can trace the information and performance indicators, but the reflective thinking that is human-like is still difficult to reproduce. In addition, the technical aspect is that it might be difficult to come up with effective sound monitoring systems that work in real-time.

Scalability is also another problem, in which large AI systems are able to generate huge volumes of information which need processing by meta-cognitive components. Another set of ethical considerations is that of the cases of autonomous system making independent decisions by internal considerations.

To address these problems, the research on AI architecture development, computational efficiency, and responsible AI development needs to be continued.

7. CONCLUSION

The meta-cognitive architectures are an important move towards the development of intelligent artificial agents. Such architectures promote the safety, adaptability, and reliability of AI technologies because they can manage what happens inside it and detect potential errors. Self-monitoring mechanisms allow the agent to examine his performance as compared to the error awareness which allows the agent to correct his mistakes immediately.

Although there are problems associated with meta-cognitive techniques such as computational complexity, and system design, studies continue to improve meta-cognitive techniques. As ever more autonoetic AI systems are developed, meta-cognition will be required to develop intelligent agents to act responsibly and efficiently in complex environments.

.

REFERENCES

1. Bao, F., Wang, X., Sureshbabu, S.H., Sreekumar, G., Yang, L., Aggarwal, V., Boddeti, V.N. and Jacob, Z., 2023. Heat-assisted detection and ranging. Nature, 619(7971), pp.743-748.
2. Bickley, S.J. and Torgler, B., 2023. Cognitive architectures for artificial intelligence ethics. Ai & Society, 38(2), pp.501-519.
3. Braun, M., Greve, M., Brendel, A.B. and Kolbe, L.M., 2024. Humans supervising artificial intelligence–investigation of designs to optimize error detection. Journal of Decision Systems, 33(4), pp.674-699.
4. Drigas, A., Mitsea, E. and Skianis, C., 2023. Meta-learning: A Nine-layer model based on metacognition and smart technologies. Sustainability, 15(2), p.1668.
5. Du, S., Wen, Q., Han, T., Ren, J., Wang, M., Dai, Y., Ge, X., Li, L., Liu, J. and Gao, S., 2025. Nanoscale Metal-Organic Framework-Based Self-Monitoring Oxygen Economizer and ROS Amplifier for Enhanced Radiotherapy-Radiodynamic Therapy. Advanced Science, 12(35), p.e03582.
6. Han, H., 2025. Meta-learning contributes to cultivation of wisdom in moral domains: Implications of recent artificial intelligence research and educational considerations. International Journal of Ethics Education, 10(1), pp.79-101.
7. Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M. and Kanai, R., 2022. Meta-learning, social cognition and consciousness in brains and machines. Neural Networks, 145, pp.80-89.
8. Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M. and Kanai, R., 2022. Meta-learning, social cognition and consciousness in brains and machines. Neural Networks, 145, pp.80-89.
9. Martinengo, L., Ha, N.H.L., Tay, E., Tong, S.C. and Sevdalis, N., 2025. Implementation of a digital health intervention (CHAMP) for self-monitoring of hypertension: protocol for 3 interlinked implementation studies. JMIR Research Protocols, 14(1), p.e72942.
10. Sternberg, R.J., 2021. Meta-intelligence: Understanding, control, and coordination of higher cognitive processes. Heidelberger Jahrbücher Online, 6, pp.487-502.
11. Ukov, T. and Tsochev, G., 2025. Reviewing a model of metacognition for application in cognitive architecture design. Systems, 13(3), p.177.
12. Ukov, T. and Tsochev, G., 2025. Reviewing a model of metacognition for application in cognitive architecture design. Systems, 13(3), p.177.
13. Wang, S., Huang, Z., Xu, Z., Guo, F., Zhang, T., Hong, M., Song, P. and Zhao, Y., 2026. A Robust Dual-mode Self-Monitoring Battery Thermal Management System via Bilayer Structural Design. Advanced Functional Materials, 36(8), p.e07825.
14. Yuttachai, H., Arbaoui, B. and Yusraw, O., 2024. Agent-based model for situational awareness in the workplace: enhancing neural

networks with direct feedback alignment. TEM Journal, 13(3), p.1786.

15. Website
16. researchgate.net,2026. The-hybrid-cognitive-architecture-at-a-large-scale-Shapes-reflect-the-nature. [Online]. Available at: https://www.researchgate.net/figure/The-hybrid-cognitive-architecture-at-a-large-scale-Shapes-reflect-the-nature-of_fig1_228348380 [Accessed on: 04.02.2026].
.