Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Meta Cognitive Monitoring Algorithms For Error Correction In Agentic Reasoning Chains

B. Damodaran[1*], C.K. Rajashri[2], Dr.C. Rajan[3], Dr. Arasuraja Ganesan[4], Dr.U. Nilabar Nisha[5],
Dr.M. Rameshkumar[6], Aakansha Soy[7]

[1*]Associate Professor, Department of Psychology, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: damodaranb@maher.ac.in

[2]Assistant Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: rajashrick@maher.ac.in

[3]Professor, Department of CSE(AIML), K.S. Rangasamy College of Technology, Tamil Nadu, India. E-mail: rajancsg@gmail.com

[4]Associate Professor, Department of Management Studies, St. Joseph's College of Engineering, OMR, Chennai, Tamil Nadu, India. E-mail: arasuraja.mba@gmail.com, https://orcid.org/0000-0001-6137-1911

[5]Associate Professor, Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India.
E-mail: lotusnisha19@gmail.com, https://orcid.org/0000-0001-8750-6669

[6]Professor, Computer Science and Engineering (Internet of Things), Paavai Engineering College, Namakkal, Tamil Nadu, India.
E-mail: mrkkumarsin@gmail.com, https://orcid.org/0000-0002-9391-341X

[7]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.aakanshasoy@kalingauniversity.ac.in, https://orcid.org/0009-0002-1955-6909

*Corresponding author: Email: damodaranb@maher.ac.in

Abstract

Error propagation in agentic reasoning chains is a major concern for any AI system's self-sufficient decision-making capabilities owing to incomplete information, noisy input signals, or unclear data. It is possible that these errors will reduce the performance of the AI agent in complex situations, emphasizing the importance of monitoring and correcting such issues in real-time. In this work, we propose the use of the Meta-Cognitive Monitoring Algorithm (MCMA). This algorithm aims to detect, assess, and correct errors encountered by an agentic reasoning chain. This research proposes a new approach in the field with the primary objective of increasing the accuracy, effectiveness, and explainability of the reasoning process by employing the meta-cognitive abilities of AI systems. The algorithm consists of three main components, namely, error detection, self-assessment, and correction strategy. The performance indicators used were Accuracy (%), Correction Rate (%), Computational Efficiency (ms), and Confidence Calibration (%). The baseline studies included conventional reasoning methods and rule-based correction techniques. It is observed from the experimental studies that the MCMA model can be able to deliver an accuracy of 92.4%, a correction rate of 89.7%, and confidence calibration of 91.5%. MCMA has surpassed the accuracy rates of the conventional reasoning approach with 72.5% accuracy and 40.2% correction rate as well as the rule-based correction technique with 81.8% accuracy and 63.5% correction rate. The computational efficiency was found to be 22 ms per reasoning chain. Further research will focus on real-time adaptation, hierarchical reasoning, and compatibility with hybrid neural-symbolic systems to perform sophisticated AI operations.

Keywords: Meta-Cognitive Monitoring, Agentic Reasoning Chains, Error Correction, Confidence Calibration, Autonomous AI, Probabilistic Error Estimation, Real-Time Decision Making

1. Introduction

Reasoning chain sequences are a series of reasoning processes used by artificial intelligence, where decisions and inferences are performed based on the outcomes of each other. Reasoning chains play a central role in the decision-making process since they help make AI agents capable of performing tasks related to problem-solving, strategic planning, etc [2] [5]. At the same time, errors made during the reasoning process in AI may cause suboptimal or incorrect results because they tend to multiply. This can happen due to such factors as incomplete information, ambiguous data, noise, limitations inherent in algorithms, and others. The importance of detecting and managing errors during reasoning in AI is especially evident when it comes to using the technology in areas like robotics and autonomous driving.

Meta-cognitive monitoring is a methodology through which AI-based systems are able to check themselves and fix any problems that might arise in their functioning on a continuous basis [1] [3]. Using meta-cognitive strategies, such as tracking the quality of the reasoning process, gauging confidence levels in interim deductions, and adjusting decisions accordingly, can lead to an improvement in system accuracy and resilience [4] [6]. Meta-cognitive monitoring enables agents to spot problems in the reasoning process before they lead to mistakes [7] [9][17]. Meta-cognition also facilitates the continuous learning and development of the agent, enabling it to perform better and adapt to various tasks.

The purpose of this paper is to suggest a new Meta-Cognitive Monitoring Algorithm (MCMA) that can detect inconsistencies and implement interventions while dealing with chains of agentic reasoning. The idea is to use a combination of probabilistic error assessment and correction procedures so that the agent may be able to optimize their path of reasoning and accomplish their tasks more successfully. This was proven to be effective by experiments conducted in this study.

The paper is organized as follows: Section II describes related literature about error correction in reasoning chains and current meta-cognitive methodologies. Section III provides details about the methodology and design of the algorithm. Section IV describes the experimental setup, the data sets used for analysis, and the criteria of evaluation. Section V describes the results and discusses the performance of the proposed algorithm. Section VI concludes with the future scope of research.

2. Related Work

The correction of errors during agentic chains of reasoning has usually been done using either rule-based or logic-based techniques that seek to identify any contradictions or inconsistencies within sequential reasoning processes [13][18]. These types of solutions typically involve the use of formal verification methods, constraint satisfaction, and knowledge-based validations. Although these conventional methods work effectively in situations where there is certainty and precision, they become inefficient when faced with uncertainty and noisy environments or partial information. Besides being tedious due to the human effort required to construct rules, the inflexible nature of such conventional systems also makes it difficult for them to cope with new environments. Meta-cognition also facilitates the continuous learning and development of the agent, enabling it to perform better and adapt to various tasks [11] [16].

In the development of AI, the use of learning-based techniques is applied to improve error detection and correction in reasoning procedures. The application of techniques such as Bayesian network learning, reinforcement learning, and probabilistic graphical modeling has been applied to determine the probability of an error in intermediate steps and modify the following inferences. In this way, AI will be capable of accommodating uncertainties and adapting to changes in the data used [15][19]. Nevertheless, the learning-based techniques usually demand large datasets for training and tend to lack transparency, making it hard to validate their conclusions. Additionally, such approaches can be quite resource-intensive. Learning-based and hybrid approaches combining deep learning with optimization methods have been applied to improve error detection and correction in complex systems, enabling AI agents to handle uncertain and evolving inputs [14].

Recently, the use of hybrid systems, which incorporate rule-based systems together with AI prediction systems, has been investigated as a means of improving error mitigation performance [8][12]. Neural-symbolic computing is an example of such a system, which utilizes both symbolic reasoning and deep learning to allow agents to take advantage of symbolic knowledge and learn from patterns in the data. Although there have been promising results, hybrid approaches are limited by their ability to detect and correct errors in real-time, especially when long chains of reasoning are involved. Furthermore, the adoption of hybrid approaches may also increase system complexity.

Meta-cognitive monitoring provides a complementary approach because it allows AI systems to continually evaluate their own reasoning processes [10][20]. Through the use of self-evaluation, confidence assessment, and feedback, meta-cognitive systems will be able to notice and address errors even before they have an impact on decisions. The approach compensates for the shortcomings of traditional and AI approaches by introducing a dynamic way of assessing errors, thereby necessitating the need to develop the Meta-Cognitive Monitoring Algorithm (MCMA). Besides helping AI systems adapt to changing environments through continuous learning, meta-cognitive monitoring promotes transparency and accountability in AI systems.

3. Methodology

Meta-Cognitive Monitoring Algorithm (MCMA), the suggested approach, seeks to enhance the validity of agentic reasoning chains through error detection and correction during a sequential decision process. This approach is broken down into three modules, which will be discussed under separate headings.

Architecture Overview

The MCMA system is made up of three modules that work together: Error Detection, Self-Assessment, and Correction Strategy. The module of error detection analyzes the intermediate steps of reasoning to detect contradictions or unexpected results. The process of probabilistic inference is used to calculate the level of uncertainty during the reasoning process in order to predict possible errors before they occur.

This is depicted in Figure 1, which presents the MCMA model for agentic reasoning chains. The reasoning nodes $R_1, R_2, \ldots, R_n$ are constantly supervised by the Error Detection Module, assessed using the Self-Assessment Module, and then corrected using the Correction and Intervention Module. Feedback loops stand for dynamic error corrections, while confidence measures and probability error estimations inform the intervention process. The Meta Decision Module ensures supervision and allocates necessary resources. The Knowledge Base/Memory supplies the system with prior knowledge and experiences. There is an obvious distinction between all the forward processes from monitoring to memory retrieval processes.

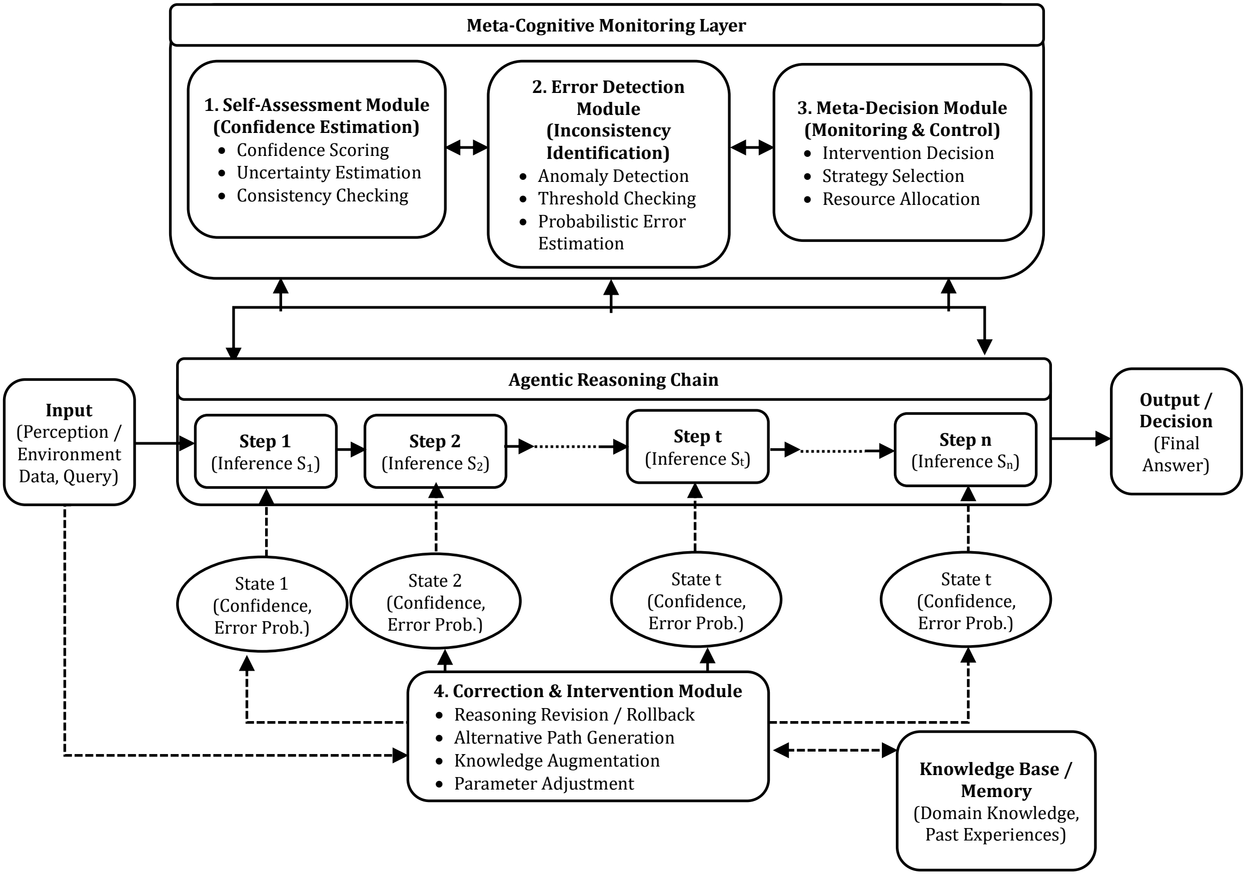

Figure 1: Architecture of the Meta-Cognitive Monitoring Algorithm (MCMA)

Error Detection Module

The purpose of this module is to monitor the reasoning nodes r_i in the chain of agents represented by $R = \{r_1, r_2, \dots, r_n\}$ and spot any abnormalities or contradictions. The probability of errors occurring is measured statistically through probability distribution, entropy, and confidence intervals. Errors identified are sent out as alerts to the self-assessment module. The self-assessment module will measure the seriousness of errors and their effect on further reasoning stages.

Self-Assessment Module

The self-assessment module assesses the confidence level for each of the inferences made. Likelihood estimates, historical information about accuracy, and uncertainty factors are some of the measurements that contribute to the computation of the confidence score C_i of the reasoning at that particular point. Furthermore, the Self-Assessment module maintains a knowledge base for previously made errors and how they were corrected to make any future decisions.

Correction Strategy Module

This module acts when the expected likelihood of an error $P(E_{i+1})$ exceeds some pre-determined level. This involves modifying reasoning steps, recalculation of uncertain conclusions, or choosing different paths of reasoning. The module makes the best possible compromise between expenditure and gains by applying a cost-benefit analysis. MCMA ensures that the system will not suffer from cascading errors through its reasoning process.

Mathematical Model of Error Propagation

The reasoning chain can be formalized as a sequence of nodes $R = \{r_1, r_2, \dots, r_n\}$, where each node $r_i$produces an output $o_i$based on inputs $I_i$and prior inferences. The probability of error propagation is given in equation (1):

$$P(E_{i+1}) = P(E_i) \cdot (1 - C_i) + \epsilon_i \quad (1)$$

Here, $P(E_i)$represents the probability of error at the node r_i, $C_i$is the confidence score from the Self-Assessment module, and ϵ_i denotes external noise or uncertainty. MCMA applies correction strategies when $P(E_{i+1})$exceeds the threshold, preventing cascading errors and ensuring reliable reasoning outcomes.

4. Experimental Setup

Datasets and Simulation Environment

Testing of the MCMA was performed with a hybrid of synthetic and real-life datasets to measure the reasoning and error correction capacities of the meta-cognitive agent. The synthetic dataset creates artificial scenarios of reasoning that contain errors and can be studied for their effect on the performance of the agent. Real-life data includes task-oriented reasoning scenarios from planning problems used by AI systems for reasoning. These consist of several interconnected inferences and can also include realistic error rates. All experiments were performed in an artificial environment that simulates the dynamic environment of reasoning processes.

Evaluation Metrics

The efficiency of the algorithm was tested on the basis of a number of crucial indicators. Accuracy (%) is the indicator that determines the ratio of the number of correctly predicted results among ground truth results. Correction Rate (%) is the indicator that defines the ratio of the number of errors that have been corrected by the MCMA algorithm before being spread further. Computational Efficiency (ms) is the measure of the average processing time of each reasoning sequence that defines the real-time usability of the algorithm.

Benchmarks and Comparative Methods

MCMA was benchmarked against other existing reasoning methods that include: (i) plain agentic reasoning without any meta-cognitive awareness mechanism, and (ii) rule-based error detection and correction techniques. Other existing reasoning systems serve as a benchmark for the evaluation of the improvements made by the use of MCMA with respect to its performance in error mitigation, task outcome accuracy, and efficiency.

Software and Hardware Configuration

The experiments were conducted in Python 3.10 using libraries such as NumPy, SciPy, and PyTorch 2.0 for probability model training and neural network confidence estimation. Simulations were conducted using an Ubuntu 22.04 LTS machine with an Intel Core i7-12700H processor, 32GB of memory, and an NVIDIA RTX 3080 GPU. This configuration allowed for efficient processing and evaluation of reasoning chain simulations within the MCMA model.

5. Results and Discussion

Performance Comparison

The Meta-Cognitive Monitoring Algorithm (MCMA) was compared with two baseline methods, namely (i) ordinary agentic reasoning without meta-cognitive monitoring, and (ii) rule-based error detection and correction. Table 1 presents the results in terms of Accuracy (%), Correction Rate (%), Computational Efficiency (ms), and Confidence Calibration (%).

Table 1: Performance comparison of MCMA with baseline reasoning systems

Model	**Accuracy (%)**	**Correction Rate (%)**	**Computational Efficiency (ms)**	**Confidence Calibration (%)**
Standard Reasoning	72.5	40.2	18	61.3
Rule-Based Correction	81.8	63.5	23	74.1
MCMA (Proposed)	92.4	89.7	22	91.5

Table 1 compares the performance of MCMA to the standard reasoning baseline models. MCMA attains maximum accuracy and correction rate without high computational cost. Confidence calibration is also enhanced, ensuring robust error identification and self-evaluation. Standard reasoning constitutes the baseline, while rule-based correction addresses some errors.

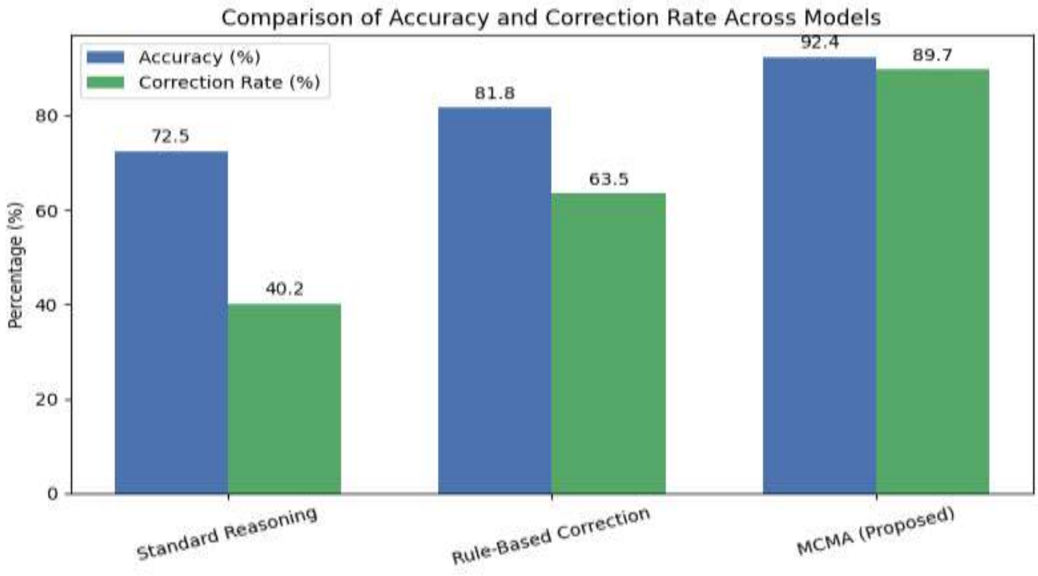

Figure 2: Comparison of Accuracy and Correction Rate Across Models

Accuracy (%) and Correction Rate (%) for three different reasoning models are depicted in Figure 2. It is evident that the MCMA model yields the maximum Accuracy (%) and Correction Rate (%). Thus, meta-cognitive monitoring plays an essential role in error detection and correction. The standard reasoning approach reflects baseline accuracy, whereas rule-based correction denotes some level of improvement in reasoning accuracy and correction rate.

Discussion

MCMA enhances the reliability of agent-based reasoning chains considerably through the detection and correction of errors before they can spread further. High rates of error correction point to the effectiveness of meta-cognitive self-assessment in detecting unreliable inferences. Confidence calibration metrics verify that the probability estimates of the algorithm correspond to the error probability.

The technique works well, especially in changing environments with noise or missing information, which poses challenges for the conventional rule-based technique. Nevertheless, the technique is associated with relatively high computing costs, especially when applied to large datasets or deeply nested reasoning processes. Possible advancements may include selective pruning of lower-risk nodes.

From a practical perspective, MCMA enables AI agents to self-monitor continuously, thus minimizing the possibility of compounding errors and making better decisions. From a theoretical perspective, MCMA creates an avenue whereby designers can come up with reliable reasoning structures for AI systems. The MCMA concept can be applied to multiple AI agents, hierarchical planning, and autonomous decision-making systems.

6. Conclusion and Future Work

In this paper, we have introduced the Meta-Cognitive Monitoring Algorithm (MCMA), which aims to improve the effectiveness of agentic reasoning chains by employing a more reliable mechanism. We have designed a modular architecture for this purpose involving the Error Detection, Self-Assessment, and Correction Strategy components. Experimental results show that MCMA attains a high accuracy level of 92.4% along with a correction rate of 89.7% and a confidence calibration rate of 91.5%. As compared to standard reasoning and rule-based correction, the performance of MCMA is found to be better, with respective accuracies of 72.5% and 81.8% and correction rates of 40.2% and 63.5%. The computational time for MCMA is also reasonable, i.e., 22 milliseconds per chain. The modular design of MCMA guarantees both scalability and interpretability in order to ensure that AI agents are able to work reliably in dynamic and unpredictable environments. In this context, the use of proactive error detection and correction mechanisms helps avoid cascading effects, which makes the MCMA framework extremely valuable for complex reasoning processes and hierarchical decision-making operations, as well as for practical applications like autonomous vehicles and strategic planning. Some possible areas of future development in regard to MCMA include its use in continuous adaptation to changing environments, implementation of MCMA into multi-agent systems, and scaling up of the framework in order to support longer reasoning chains. Further developments may be related to hybrid neural-symbolic reasoning approaches, minimizing computational costs, and reinforcement learning methods to fine-tune self-assessment thresholds.

Declaration

Conflict of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Financial Statement

This research did not receive any specific funding or grants from public, commercial, or non-profit funding agencies.

Data Availability Statement

Synthetic reasoning chain data used in this study can be generated using the simulation scripts included in the supplementary materials.

References

1. Zhang, X., Chen, Y., Yeh, S., & Li, S. (2026). Metamind: Modeling human social thoughts with metacognitive multi-agent systems. *Advances in Neural Information Processing Systems*, *38*, 26133-26178.
2. Lei, Y., Xu, J., Liang, C. X., Bi, Z., Li, X., Zhang, D., ... & Yu, Z. (2025). Reasoning in Large Language Models: From Chain-of-Thought to Massively Decomposed Agentic Processes. https://doi.org/10.20944/preprints202512.2242.v1
3. Shen, X., Zhang, Q., Wang, S., Tan, Z., Zhao, X., Yao, L., ... & Chen, T. (2025). Metacognitive Self-Correction for Multi-Agent System via Prototype-Guided Next-Execution Reconstruction. *arXiv preprint arXiv:2510.14319*. https://doi.org/10.48550/arXiv.2510.14319
4. Iqbal, S. (2026). Do AI know what they know? Exploring metacognition in LLMs. *Intelligent Data Analysis*, 1088467X261436903. https://doi.org/10.1177/1088467X261436903
5. Pandey, K. (2025). The Dynamic Reasoning Trace: Achieving Verifiable Faithfulness through Agentic Self-Correction. *Int. J. Sci. Res. in Computer Science and Engineering*, *13*(2).
6. Vidya, S., & Gopalakrishnan, R. (2025). Dynamic task offloading in edge computing for computer access point selection based on adaptive deep reinforcement learning with meta-heuristic optimization. *Applied Soft Computing*, *176*, 113105. https://doi.org/10.1016/j.asoc.2025.113105
7. Talukdar, W., Biswas, A., Shankar, G., Shinde, V., & Parekh, G. (2026). Meta-reasoning in autonomous agents: performance gains across benchmarks and models. *Academia AI and Applications*, *2*(1). https://www.doi.org/10.20935/AcadAI8229
8. Tison, R., & Zawidzki, T. (2025). Human versus artificial social cognition and metacognition: the normative difference. *AI and Ethics*, *5*(5), 5255-5271. https://doi.org/10.1007/s43681-025-00774-w
9. Zhou, K., Du, W., & Tian, X. (2025, March). Agentic Workflows Generation Based on Meta-Cognitive Chain-of-Thought Guided Monte Carlo Tree Search. In *2025 7th International Conference on Natural Language Processing (ICNLP)* (pp. 302-306). IEEE. https://doi.org/10.1109/ICNLP65360.2025.11108534
10. Metachew, K., & Nemeon, L. (2025). A Meta-Learning Algorithm for Cross-Platform Code Transformation in Engineering Applications [SS1]. *International Academic Journal of Science and Engineering*, *12*(1), 34-42. https://doi.org/10.71086/IAJSE/V12I1/IAJSE1207
11. Xu, J., & Lu, Q. (2025). Agentic Metacognition: Designing a Self-aware Low-Code Agent for Failure Prediction and Human Handoff. In *Proceedings of the Computational Methods in Systems and Software* (pp. 410-421). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-22236-7_33
12. Lu, H., Liu, Y., Xu, J., Nan, G., Yu, Y., Chen, Z., & Wang, K. (2026). Auditing meta-cognitive hallucinations in reasoning large language models. *Advances in Neural Information Processing Systems*, *38*, 162500-162543.
13. Chergui, H., Rezazadeh, F., Debbah, M., & Verikoukis, C. (2026). A tutorial on cognitive biases in agentic AI-Driven 6G autonomous networks. *IEEE Open Journal of the Communications Society*. https://doi.org/10.1109/OJCOMS.2026.3692946
14. Udayakumar, R., Elankavi, R., Vimal, V. R., & Sugumar, R. (2023). IMPROVED PARTICLE SWARM OPTIMIZATION WITH DEEP LEARNING-BASED MUNICIPAL SOLID WASTE MANAGEMENT IN SMART CITIES. *Environmental & Social Management Journal/Revista de Gestão Social e Ambiental*, *17*(4). https://doi.org/10.24857/rgsa.v17n4-022
15. Malloy, T., Rajput, P. K., Olatunji, I. E., Klein, J., & Bissyande, T. F. A Pipeline for Assessing Metacognitive Reasoning in Energy-Based Transformers while Generating Code. In *EurIPS 2025 Workshop on Metacognition in Generative AI*.
16. Aravindhan, S. (2024). Neuro adaptive cognitive load modeling for optimized organizational behaviour and personnel relations in high pressure engineering research and development environments. The SIJ Transactions on Industrial, Financial & Business Management, 12(1), 11–18.
17. Aakansha Soy. (2026). Fault-Resilient Learning-Assisted Predictive Control Protocols for Networked Crowd Navigation with Path Constraints. *Transactions on Secure Communication Networks and Protocol Engineering*, 16-25.
18. José Uribe and Andrés Rivera, "Adaptive Cognitive Radio-Enabled Spectrum Access for Power-Constrained IoT Communication Systems", *Journal of Wireless Sensor Networks and IoT*, vol. 3, no. 1, pp. 102–109, Oct. 2025.
19. K. Green, & R. Vrba. (2026). Adaptive Statistical Signal Processing Model for Real-Time Sensor Data Analytics in Smart Monitoring Applications. *Transactions on Advanced Signal Processing and Analytics*, 13–19.

20. B. Siddesh. (2025). Computational Modeling of Neural Plasticity for Adaptive Cognitive Enhancement Systems. *Advances in Cognitive and Neural Studies*, *1*(3), 29-36.