Review Article

Metacognition and Pedagogy in the Era of Artificial Intelligence

Gabriel Goldmeier[1], Ronaldo Mota[1]

1. Chair of Artificial Intelligence, Universidade Federal do Rio de Janeiro, Brazil

This article examines the educational consequences arising from the emergence of technologies based on Artificial Intelligence (AI), with an emphasis on advancements resulting from the use of artificial neural networks and pattern recognition. Traditional educational methods, grounded in predictability, content transmission, and memorization of procedures and techniques, prove insufficient in a social context where information is increasingly accessible instantaneously and free of charge. In this scenario, pedagogy centred on metacognitive predicates, highlighting student self-awareness and emphasizing the importance of learning to learn, appears to be the most appropriate approach to address the social transformations driven by AI. This paper presents and discusses general guidelines for implementing metacognitive practices in the classroom, fostering reflection on their potential effects on the learning process.

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

I. About the Need for a New Pedagogy

In the past century, we successfully trained professionals grounded in content, procedures, and techniques, equipping them to face the challenges of a relatively predictable labour market. In the twenty-first century, the digital revolution has significantly complicated this task. Expectations regarding future professional demands are in constant flux, influenced by an accelerated and unpredictable dynamic. The very volatility of the present renders everyday life marked by rapid changes, hindering the development of a clear vision of future requirements for professional practice as well as for citizenship itself.

Furthermore, we know that routine tasks based on recipes and standardized procedures are being rapidly replaced by robots and AI systems capable of autonomous learning and adaptation [1]. This reality imposes the necessity of developing pedagogical strategies that are radically different from traditional methods. Currently, it is imperative to first equip learners with the ability to cope with abrupt changes through conscious adaptation processes, cultivating emotional balance and rationality in the face of unprecedented situations. To achieve this, it is essential for both students and citizens to be aware of the historical moment they are experiencing, understand the dynamics of the contemporary world, and be able to locate themselves geographically.

However, these tasks are not straightforward. The rapid pace of global transformation complicates the ability of educational managers and teachers to perceive this new reality immediately and to construct an educational framework tailored to these new challenges and demands. Fortunately, technology can contribute to this complex process. Through learning analytics, it is possible to identify effective strategies to develop specific skills and competencies for each student [2]. These tools enable both students and educators to engage in ongoing reflection on how to learn and teach, promoting lifelong learning. This is the core of metacognitive approaches, which focus on increasing awareness of one's own learning process and optimizing it across various contexts and for different purposes.

For example, data analysis generated by virtual learning environments provides systematic insights into the student's educational journey. This allows for a deeper understanding of their learning characteristics and preferences. By monitoring digital footprints, we can identify which media facilitate learning, in which times or contexts students perform best, and determine which methodologies are most effective for each profile. With this data, personalized educational pathways can be designed, tailored to the specific needs, contexts, and educational goals of each learner.

In this scenario, especially within the professional sphere, as a collaborator or leader, the development of metacognitive skills must be strongly fostered. Consequently, the formation of contemporary professionals requires cultivating independent and autonomous learning [3], logical reasoning (deductive, inductive, formal, informal), textual interpretation, and analytical reasoning, as well as understanding distinctions between causality and correlation, and concepts of probability and statistics, among other foundational elements essential to rationality and scientific thinking [4].

Therefore, contemporary educational projects, both formal and informal, must not only promote citizenship in the AI era but also prepare professionals for a rapidly changing labour market. Beyond technical skills, it will be necessary to develop abilities and competencies that, until recently, were often neglected [5]. The accelerated changes in the job market compel professionals to upskill continuously and frequently change their areas of expertise. This entire context underscores the urgent need for a new pedagogy, one capable of adapting to these challenges, ensuring training that is both appropriate and sufficiently agile to keep pace with these transformations.

II. Metacognition: The Art of Learning to Learn

An alternative way to understand metacognition is to consider it as a set of approaches that transcend basic and direct cognition. It represents mastery of internal skills that enable the learner to reflect on, monitor, and regulate their own learning processes consciously. Therefore, metacognitive skills can be divided into two main dimensions: (i) metacognitive knowledge, which includes understanding the factors that influence learning performance, mastery of various cognitive strategies, and the ability to adapt these strategies to specific situations—also involving recognition of one's strengths and difficulties, as well as understanding when and how to apply certain techniques to optimize learning; and (ii) metacognitive regulation, referring to actions such as planning, goal-setting, continuous monitoring, control of employed strategies, and assessment of results, thereby fostering more autonomous and effective learning, as it allows the student to adjust their actions and strategies according to progress and encountered obstacles [6].

Additionally, broadly speaking, encouraging students to reflect on their own learning processes involves promoting self-reflection and developing self-awareness skills [7]. This practice also enhances the ability to work collaboratively, fostering understanding of others, empathy, and cooperation—competencies that are essential for collaborative learning. These skills, in turn, constitute fundamental elements in the development of individuals capable of lifelong autonomous learning, especially in a context increasingly characterized by rapid and constant change.

In the process of student development, beyond acquiring traditional technical knowledge, space opens for the cultivation of socio-emotional skills. Many educators see this integration as a recovery of humanistic elements essential for a more holistic education, balancing the exclusive focus on technological and cognitive aspects. Consequently, the formulation of metacognitive strategies functions not only as a tool for "knowing how to learn" but also as a reassertion of humanistic values that prioritize understanding, empathy, creativity, and resilience.

Finally, this dynamic underscores the importance of an educational approach that values not only technical content but also the development of metacognitive and socio-emotional skills. This fosters the formation of individuals who are more independent, creative, critical, and adaptable to ever-changing scenarios—traits essential for successful navigation in today's and future society.

III. Current Landscape of AI

In 1997, IBM's Deep Blue, a digital chess-playing computer, achieved a historic milestone by defeating then-World Chess Champion Garry Kasparov. This victory signified a paradigmatic shift in our understanding of human cognitive abilities versus machine capacities [8]. A few decades later, virtually any smartphone processor has become sufficiently powerful to beat a world chess champion, exemplifying the exponential growth in computational power.

Nonetheless, the challenge appeared even greater when it came to the game of Go [9], due to its vastly larger number of possible variations and moves, which made programming machines capable of defeating top players more difficult. Many experts believed that attaining such a capability was still far off. That was until 2016, when AlphaGo, a neural network-based computer program developed by DeepMind—acquired by Google—employed an innovative reinforcement learning model to defeat Lee Sedol, widely regarded at the time as the best Go player. The following year, AlphaZero, an even more advanced version, defeated Stockfish 8, the strongest chess engine at that time, a direct successor to and vastly more powerful than Deep Blue, which relied on traditional evaluation methods and decision trees. This victory intensified the debate surrounding new frontiers in AI.

The groundbreaking innovation of AlphaZero lies in its starting point: it begins from scratch, meaning it does not rely on pre-established heuristics, databases, or fixed rules to guide its moves. Unlike Stockfish 8, which depends on predetermined rules and extensive opening databases, AlphaZero learns by playing against itself, applying principles of machine self-learning. Remarkably, within just four hours, AlphaZero transformed from a novice into one of the world's best players, all without direct human intervention or external data, relying solely on trial-and-error learning.

To understand this rivalry between Stockfish 8 and AlphaZero, it is essential to grasp that computer programming—by nature, based on rational logic—does not necessarily need to be grounded in the insertion of preset logical rules. While rooted in formal logic, current AI models—built on machine learning and pattern recognition—have gained prominence in solving complex problems. These models, which utilize statistical calculations, artificial neural networks, and deep learning techniques, have

doi.org/10.32388/T36PI8

demonstrated superior efficiency and flexibility in understanding complex processes. AlphaZero's victory in chess exemplifies this new approach but represents only one of many possible applications [10].

Practically speaking, these advances suggest a new perspective: instead of manipulating symbols through rigid rules, we can capture properties of complex objects via pattern recognition, simulating the operation of neural connections. Each property of an object or phenomenon is assigned a numerical value—a weight—that reflects its relevance to diagnosis or classification tasks. As a result, the system does not follow a fixed set of rules but constructs its understanding based on statistical distributions that determine the importance of each feature—this is the core of machine learning [11].

A classic example of this approach is facial recognition: upon capturing an image, the system processes and converts it into a standardized format, performing a visual encoding. It identifies facial landmarks—such as the distance between the eyes, the facial contour, mouth, nose, and scars—and compares these points to a previously stored database. The combination of these values enables the system to recognize or not recognize an individual with high accuracy.

These operations depend on artificial neural networks, whose basic structure involves processing layers: an input layer composed of measurable features, and an output layer representing a database of known faces. Each connection or synapse between neurons is weighted to reflect its relevance to identification. The key question is: who or what determines these weights? Surprisingly, the answer can be no one, or more precisely, these weights can even be initialized randomly. The network training process involves adjusting these weights through trial, error, and success until the system learns to correctly identify faces. The more the network "errs" and corrects itself, the more it improves, a process known as iterative feedback adjustment.

For more complex tasks, such as advanced facial recognition, it is common to incorporate multiple hidden intermediary layers, along with various other dimensions or categories representing different aspects or partial similarities. In other words, recognition challenges can be subdivided into stages, with each set of layers addressing part of an intermediate goal. Deep learning systems typically consist of networks with many such hidden layers, enabling the network to perform effectively after multiple training iterations. When results do not meet expectations, the system "reverts" to the beginning and readjusts initial weights, resembling a restart, until an acceptable accuracy level is achieved.

Understanding all the nuances of this process is undoubtedly a complex task requiring significant expertise. However, the most important takeaway is that pattern recognition, learning and identifying statistical patterns, outperforms traditional AI, which is based on logical deductions and manual programming. In other words, deep learning systems, unlike conventional approaches, in principle, do not require rigorous pre-established concepts or explicit logical inferences, relying instead on their capacity to detect patterns and self-adjust throughout training.

IV. Humans versus AI and Metacognition as the Final Frontier

Above, various considerations were made regarding the recent digital revolution, driven by the exponential growth in the participation of new AIs in scientific advancement and public debate, as well as a broad perspective on the ways in which different educational objectives can enhance benefits and mitigate risks associated with this transformation.

This discussion is connected to an evolving understanding of the “Humans versus AI” debate [12]. In this section, we will explore emerging research fields, particularly those related to the potentialities and limitations of AI, with an emphasis on the metacognitive limitations of these machines, especially at moments when so-called technological singularities are anticipated.

Homo sapiens possesses essential attributes such as physical strength, cognition, and, notably, metacognition, among others. In earlier times, society was content to abandon the idea of competing with machines in terms of physical strength. Recently, however, the greatest challenge has been recognizing that, in certain aspects of simple cognition, we are gradually being surpassed by machines capable of learning and adapting. Humans' remaining hope is to maintain competitiveness in the domain of metacognition.

Therefore, developing metacognition, as outlined above, is fundamental to improving learning, problem-solving, and decision-making. Additionally, it enhances individual self-confidence and autonomy within their educational process. As a reflexive activity about one's own reflection, metacognition enables the monitoring and regulation of thoughts, emotions, and behaviours, as well as the assessment of one’s own performance.

Over the past decade, models based on neural networks, reinforcement learning, and pattern recognition have, through a radical paradigm shift in programming, surpassed traditional methods primarily grounded in logical inference. Given this landscape, a crucial question concerns the extent to which machines are also developing, or not, metacognitive capabilities. In other words, we seek to understand whether AI systems can reflect on their own learning processes and identify weaknesses in their ability to learn how to learn—a skill considered fundamental for autonomous and adaptive operation.

By engaging in this analysis, the aim is to stimulate a debate on how, through fostering education aimed at developing metacognition, we can valorise the last frontier of human competitiveness in the ongoing contest with machines, a domain in which we can still maintain a significant and perhaps the most vital advantage: the capacity to reflect, learn to learn, and evolve continuously.

V. The Metacognition of Machines: The Case of DeepSeek

The relationship between metacognition and DeepSeek models has recently garnered increasing interest [13]. If we consider metacognition as also encompassing a system's capacity to monitor, understand, and adjust its own cognitive processes, including the active regulation of its actions, preliminary signals indicate that AI systems are progressing in this direction. For example, in the DeepSeek-R1 and DeepSeek-R1-Zero models, it is evident that the interactions between monitoring and control processes are crucial for achieving levels of reasoning that are coherent and often surprising, characteristics that define these models. In particular, the so-called "aha" moment is frequently cited as an example of metacognitive behaviour within the scope of DeepSeek.

In this context, as shown in a previous paper [13], DeepSeek may be pioneering a new pathway for reasoning in AI, favouring reinforcement learning (RL) over the more traditional approach of supervised fine-tuning (SFT). This study aims to analyse the implications of these innovations on machines' ability to simulate behaviour based on self-reflection and to act autonomously. It remains uncertain, however, to what extent these elements can definitively be associated with metacognition, a trait previously regarded as exclusive to humans. Nonetheless, current indications suggest a move in this direction.

The emergence of metacognitive control through RL points to a fundamental condition: it is not merely an auxiliary feature but an essential component for coherent information processing. A machine must possess awareness of its cognitive processes and the capacity to regulate them to maintain consistency in complex reasoning tasks. An even more impressive example of this ability was observed in the advanced DeepSeek-R1-Zero version. During intermediate training phases, this model demonstrated an enhanced capacity to allocate time for problem reflection dynamically, thereby optimizing its responses in real time. Instead of following a rigid, rule-based training regime, the system learned to adjust its problem-solving strategies autonomously, guided by appropriate incentives. This means that, rather than being explicitly programmed to recognize specific solution types, it developed sophisticated reasoning strategies through learning.

DeepSeek represents a significant advancement in the ability of machines to simulate behaviours related to introspection about their own cognitive processes and to act based on them, features traditionally associated with metacognitive predicates believed to be exclusive to humans. Although the question of whether machines can fully surpass humans in these capacities remains open, with few definitive elements available so far, it is undeniable that if machines ever achieve full metacognition, the DeepSeek models will have played a crucial role in this historical trajectory [13]. This phase, whose timeline remains unpredictable, marks the beginning of a new era in the relationship between humans and AI.

VI. Metacognitive Practices in the Classroom

As previously mentioned, within a metacognitive approach, it is essential to actively encourage learners to develop an awareness of their own historical, social, and geographical positioning. These elements are crucial for progressively expanding their understanding of how they learn within a specific context and for a specific educational purpose. It is important to emphasize that this process does not occur spontaneously; rather, it requires appropriate guidance from the educator. Additionally, interaction with peers forms an integral part of deepening awareness regarding the media employed and the contexts in which learning is amplified. Therefore, this constitutes a collective-cooperative process that simultaneously aims to promote student autonomy.

Several teaching practices that support metacognitive development are already becoming established; however, their implementation depends on the specific circumstances of each educational situation and the objectives of the project. Here, we will explore some examples to illustrate this approach, demystifying the idea that a metacognitive pedagogy is distant, overly costly, or excessively complex.

One example pertains to a simple assessment procedure [14]. Typically, the teacher responsible for a subject divides the content into topics, and partial assessments focus on measuring students' mastery of these contents, techniques, and procedures. After each assessment, the most common practice, especially in traditional teaching, is for the teacher to introduce a new topic in the subsequent class. Subsequently, the graded exams are often returned with minimal comments, only with the assigned grade or score. A metacognitive approach, in contrast to traditional pedagogical methods, would involve, for instance, fostering a reflective dialogue with students immediately after the assessment, preferably in the class following the exam. This activity, therefore, becomes a central and inseparable part of the evaluation process, during which students analyse their own performance, identify difficulties, successes, and strategies employed. Thus, assessment extends beyond merely assigning grades; it includes the student's capacity for self-reflection—an activity that enhances their autonomy to learn how to learn.

There are numerous other practices that support a metacognitive approach in the classroom, such as: (i) self-questioning: students ask themselves about what they are learning, with questions like "Did I really understand this?" or "What strategies can I use to solve this problem?"; (ii) self-assessment: students reflect on their progress, using checklists or rubrics to evaluate their skills and knowledge; (iii) think-aloud: students verbalize their reasoning while solving problems or reading, helping to make their cognitive processes explicit; (iv) planning and goal-setting: before beginning a task, students establish clear objectives and plan the strategies needed to achieve them; (v) comprehension monitoring: during learning, students pause periodically to verify their understanding of the content, identify confusing parts, and seek clarification; (vi) collaborative reflection: students share their thought processes and strategies with peers, promoting the exchange of different perspectives and enriching their understanding; (vii) reflection journals: students regularly record their learning experiences, reflecting on what they have learned, which strategies were effective, and what areas still need improvement; and (viii) use of graphic organizers: visual tools such as mind maps and concept maps assist students in structuring and reflecting on their knowledge, facilitating understanding and retention of content.

The implementation of these practices, along with other similar resources, enhances students' awareness of their own learning processes. This perception contributes to a deeper understanding of the content, improves academic performance, and strengthens autonomy in knowledge construction.

VII. Conclusions

Contemporary educational projects, beyond fostering citizenship in the era of AI, must focus on the development of professionals and citizens equipped with competencies aligned with the demands of this new period. This context highlights the urgent need for a novel pedagogical approach capable of adapting to constant challenges, ensuring agile, relevant, and socially and technologically responsive training. Consequently, it is crucial to cultivate metacognitive skills that enable individuals to become more autonomous, creative, critical, and adaptable, traits essential for success today.

AI systems based on neural networks and deep learning, unlike traditional approaches, dispense with predefined concepts or explicit logical inferences, instead relying on pattern detection and autonomous adaptation during training. In this landscape, where human physical and cognitive capabilities are progressively surpassed by machines, the ultimate battleground appears to focus on metacognitive competencies. Therefore, promoting the development of these skills becomes a strategic priority for humans to maintain a distinctive advantage, particularly in the capacity to reflect on their own reflection, learn how to learn, and continuously evolve throughout life.

In summary, a metacognitive approach in the classroom is fundamental for fostering students' self-awareness, combined with conscious reflection on their own learning processes, thereby supporting the development of autonomy and critical thinking. Such an approach can be effectively implemented through simple and accessible practices, including post-assessment reflections, self-questioning, verbalization of reasoning, self-evaluations, task planning, and the use of visual supports. These strategies encourage students to better understand their difficulties, identify effective strategies, and utilize available resources—strengthening their comprehension of content and their academic performance. Ultimately, metacognitive practice contributes to a more collaborative and critical pedagogy, promoting student autonomy in constructing their own knowledge and preparing them to face the key challenges of the contemporary world.

About the Authors

Gabriel Goldmeier, Research Associate (Postdoctoral) at the Chair of Artificial Intelligence at the Colégio Brasileiro de Altos Estudos of the Universidade Federal do Rio de Janeiro (CBAE/UFRJ) and PhD in Education from the Institute of Education/University College London (UCL).

Ronaldo Mota, Chair of Artificial Intelligence at the Colégio Brasileiro de Altos Estudos of the Universidade Federal do Rio de Janeiro (CBAE/UFRJ) and Retired Full Professor of Physics from the Universidade Federal de Santa Maria.

References

1. ^*Lee K-F, QiuFan C (2021). AI 2041: Ten Visions for Our Future. Crown Currency; Russell S, Norvig P (2022). A rtificial Intelligence – A Modern Approach. GEN LTC.*
2. ^*Ferguson R (2012). "Learning Analytics: Drivers, Development, and Challenges." IJTEL.* **4**(*5/6):304–317. doi: 10.1504/ijtel.2012.051816.*

3. ^Mota R, Scott D (2013). Education for Innovation and Independent Learning. Elsevier; Mota R (2013). Educando para Inovação e Aprendizagem Independente [Educating for Innovation and Independent Learning]. Editora Campus.
4. ^Pinker S (2022). Racionalidade [Rationality]. Intrínseca.
5. ^Goldmeier G, Mota R (2023). "Rationality and Scientific Thinking as Foundations for Leadership in the World of Work." Qeios. doi:10.32388/bkhxow.
6. ^Mota R (2019). "Learning How to Learn is More Than Learning." The Phys Educat. **1**(01):1950001–6. doi:10.1142/s2661339519500021.
7. ^Mota R, Goldmeier G (2024). "Metacognition: Strategy for Remote Learning." In Educação Não Presencial: Polêmicas e Controvérsias [Non-Presential Education: Polemics and Controversies] (pp. 135-146). Ed. Dos Autores.
8. ^Chess.com. Kasparov vs Deep Blue: The confrontation that changed history. https://www.chess.com/pt/article/view/kasparov-vs-deep-blue-o-confronto-que-mudou-a-historia; ChessGames.com. Chess match. https://www.chessgames.com/perl/chess.pl?tid=91944.
9. ^Boardgame Arena. Go: An ancient board game played in Asia. https://pt.boardgamearena.com/gamepanel?game=classicgo.
10. ^Harari Y (2018). 21 Lessons for the 21st Century. Companhia das Letras.
11. ^Kelleher JD (2019). Deep Learning. MIT Press.
12. ^Eysenck MW, Eysenck C (2023). Artificial Intelligence vs Humans. Artmed.
13. a, b, c Mota R (2025). "Is DeepSeek a Metacognitive AI?" Qeios. doi:10.32388/pj3pom.2.
14. ^Mota R (2014). "Classroom Practice and the Digital Vortex." ABMES Blog. https://abmes.org.br/blog/detalhe/18865/pratica-de-sala-de-aula-avaliacao-e-o-vortex-digital.

Declarations

Funding: This work was carried out with the support of the Rio de Janeiro State Research Support Foundation (Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ), through the “Visiting Researcher Emeritus Program” for Ronaldo Mota ,associated to the “Artificial Intelligence and Metacognition” Chair at the “Brazilian School of Advanced Studies” of the Federal University of Rio de Janeiro (CBAE/UFRJ).

Potential competing interests: No potential competing interests to declare.