Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6 n. 1 (2026)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Rethinking Human Agency in AI-Mediated Education

Alessandro Ciasullo [1], Flavia Santoianni [1] and Alessio Manfredini [1, *1]

[1] University of Naples Federico II; alessandro.ciasullo@unina.it; flavia.santoianni@unina.it; alessio.manfredini2@unina.it
* Correspondence: alessandro.ciasullo@unina.it

Abstract: This paper examines the structural implications of Generative Artificial Intelligence (GAI) for contemporary educational systems, arguing for a shift from a tool-centred perspective to a relational paradigm grounded in co-agency. Rather than conceptualising AI as a neutral instrument, the study frames it as a cognitive co-agent that actively participates in the construction, negotiation, and regulation of knowledge. Drawing on literature on cognitive offloading, extended cognition, and shared metacognition, the analysis explores both the transformative potential and the risks of AI integration, including passive delegation, deskilling, and the weakening of higher-order cognitive processes. In response, the concept of educational architecture is advanced as a meta-design framework capable of intentionally structuring the distribution of agency across structural, epistemic, and ethical dimensions. Within this framework, teachers emerge as architects of inter-agency and co-agency, responsible for orchestrating hybrid learning environments that preserve cognitive sovereignty while fostering distributed metacognitive awareness. The paper ultimately advocates for the development of a critical AI curriculum aimed at sustaining human epistemic agency within increasingly AI-mediated educational ecologies.

Keywords: Co-Agency; Educational Architecture; Distributed Metacognition

[1] Authorship attribution: Alessandro Ciasullo conceived the study and developed the overall theoretical framework, contributing in particular to the sections on educational architecture, cognitive offloading, distributed metacognition, cognitive sovereignty, precision pedagogy, and the role of the teacher as architect of co-agency. Flavia Santoianni contributed to the pedagogical and epistemological framing of the work, providing theoretical supervision and scientific guidance, particularly with regard to the grounding of the educational architecture framework, the interpretation of implicit and explicit cognitive processes, and the broader positioning of the manuscript within contemporary pedagogical debate. Alessio Manfredini authored the section *"From AI-Agency to Co-Agency in Education"* and contributed to the conceptual development of the work at the intersection between psychology and pedagogy, particularly in relation to the relational and dialogic dimensions of AI as a cognitive co-agent.
All authors reviewed and approved the final manuscript.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

1. Introduction

The article focuses on the impact of Generative Artificial Intelligence (GAI) on educational systems, advocating for a shift from a tool-centred perspective to a relational and architectural paradigm of co-agency. The present study rethinks the role of AI as a cognitive co-agent that actively contributes to knowledge construction, thus moving beyond the reductive equation of AI as a neutral tool. The paper focuses on the potential and risks of AI integration, both in terms of transformative potential and risks, as discussed in recent literature on cognitive offloading, extended cognition, and shared metacognition. Empirical evidence has demonstrated that the integration of AI in educational settings frequently results in a decline in higher-order cognitive engagement, if it is not accompanied by deliberate pedagogical design. The integration of educational architecture as a meta-design framework has been demonstrated to serve as a deliberate structuring mechanism for the distribution of agency across a range of dimensions, including structural, epistemic, and ethical. This framework encompasses the dynamic interplay among students, teachers, and AI systems, thereby offering a systematic approach to the governance of these complex interactions. Within this paradigm, educators assume the role of co-agency architects, orchestrating the creation of hybrid learning environments that are designed to safeguard cognitive sovereignty and encourage distributed metacognition. Finally, the paper puts forward a crucial AI curriculum designed to encourage human epistemic agency within increasingly AI-mediated learning ecosystems.

2. The educational design for a co-agency

The theory is predicated on the concept of educational architecture (Santoianni, Ciasullo, 2024), which is defined as the intentional arrangement of spaces, rules, tools, roles, and practices that govern the distribution of agency during learning processes. Educational architecture is not a substitute for the straightforward design or classroom management of teaching activities; rather, it represents the meta-design level where the epistemic conditions of interaction between students, teachers and artificial intelligence are defined. It is a space for mediation and educational connection that can be both a shared junction between plural (but private) dimensions of teachers, students, educational management roles and policy makers, and shared spaces for relationships where the cognitive outcomes of AI use can be socially experimented with.

If AI is to act as a cognitive co-agent, it cannot be inserted into the educational setting in a neutral or additive way. Furthermore, it cannot be considered to be functional only in the educational setting throughout the entire learning process. Integration can be considered to encompass two distinct aspects: firstly, a social aspect, which is essential in the creation of new learning communities; and secondly, a private, subjective aspect, which is of greater significance.

The approach to the subject's knowledge is now three-dimensional, with social space, private space, and extended algorithmic space being replaced by two-dimensional approaches. The educational process, as it is currently conceptualised, originates from the isolated dimension of human learning and evolves into an extended space of expanded knowledge when it encounters the third dimension of the human/context/machine relationship. It is evident that the

presence of subjectivity throughout the process does not dissipate, contingent upon its ability to reaffirm its capacity for control, growth, and active management of its own knowledge. The emerging role of 'educational architects' will involve the management of the metalevel educational process, with the capacity to modify the cognitive ecosystem, redistribute epistemic load, and redefine the boundaries between autonomous production, cooperation, and delegation.
It is evident that the process of rewriting cognitive structures exerts a significant influence on the implicit processes of knowledge acquisition (Santoianni, 2026, 2014), thereby posing further challenges to the efficacy of explicit methods in educational settings.

In the absence of a specific design, educational architecture is frequently influenced by the logic of the technological tool and the three-dimensional space within which the subject may disintegrate. The notion of conscious pedagogy necessitates a paradigm shift within the educational process, wherein a collaborative effort is undertaken to co-design, co-act and govern the new palace of knowledge. This involves the delineation of when artificial intelligence should function as a metacognitive scaffold, when it should act as a dialogical interlocutor, and when, conversely, its involvement should be withheld, with the objective being to safeguard those phylogenetically matured neurosynaptic processes that have been instrumental in facilitating the development of species-specific skills.

It is evident that a minimum of three dimensions are implicated in the educational architecture of a co-agency.
1. The structural element pertains to the delineation of educational activities in which the term is employed, or alternatively, not employed.
2. The epistemic nature of the criteria for validating and evaluating knowledge produced through interaction with generative systems is of particular interest.
3. The ethical-normative is connected to the following:
- algorithmic transparency
- data regularity
- the regulation of responsible use

Educational architecture is defined as the tool that brings technological innovation back into a critical pedagogical framework. It is important to note that AI does not determine the form of learning; rather, the architecture designed by the teacher dictates how AI participates in the educational process.

3. Distributed Metacognition and Shared Metacognition

It is imperative to interrogate the distribution of agency and cognitive control of the learning subject in order to address the ever-evolving multiplicity of AI in the contemporary era. The issue pertains to the concept of cognitive sovereignty (Ostrom, 1984; Bygrave, 2020; Atkinson, 2025), which refers to the dimension of control and active management of the systems of organisation, supervision and structuring of one's own knowledge dynamics, from learning to its management in everyday actions. Bygrave (2020) emphasises that the concept of cognitive sovereignty is distinct from that of cognitive freedom, the former referring to a sophisticated array of systems that enable individuals to proactively and deliberately oversee their own knowledge management.

The accelerated pace of technological advancement necessitates a shift from a purely instrumental approach to precision education (Yang et al., 2021) to achieve a

genuine "precision pedagogy" (Ciasullo in Santoianni, Ciasullo, Silva, 2025), which can regulate human-machine interaction through data and theoretical contemplation. At this juncture, the theoretical objective becomes meta-theoretical by investigating the subjective cognitive dimension, the intersubjective dimension, and the additional dimension provided by the 'instrumental plus' associated with AI. This process involves the decomposition and broadening of these three dimensions until they become the primary focus of a new extended cognition (Clark and Chalmers, 1998) or even super-extended cognition (Gallagher, 2011). It appears imperative to undertake this endeavour within the domain of education.

The transformation of AI into a cognitive co-agent that actively participates in the construction of knowledge must be analysed.
It is imperative to assert the paramount significance of the metacognitive approach as the sole instrument for regulating interaction, thereby averting the encroachment of algorithmic opacity.

It is proposed that an intentional curricular structure be formulated, integrating AI not as an addendum, but as a normative and evaluative element.

The dynamic nature of AI, both additive and non-substitutive, has been shown to engender further transformations in knowledge management processes through the medium of metacognition. This metacognition, in turn, evolves from an individual process to a shared metacognition (Garrison, 2022; Devers et al., 2018; Loginov, 2025; Ciasullo et al., 2025). In this scenario, the subject not only contemplates their own thought processes but also meticulously observes the logic of AI, thereby identifying its biases, limitations, and inherent opacity to the greatest extent feasible.

Recent studies have indicated a direct correlation between the level of metacognitive awareness in the use of artificial tools and enhanced academic performance (Iqbal et al., 2025; Levin et al., 2025). The delegation of low-order or repetitive tasks to AI has been demonstrated to liberate cognitive resources, enabling the allocation of these to higher-order tasks, such as creativity, complex synthesis and ethical judgement. This is achieved through the conscious process of cognitive offloading.

The implementation of educational systems through a curriculum aimed at constructing new forms of metacognitive awareness through a critical approach has the potential to result in unprecedented processes of cognitive sovereignty in the field of precision pedagogy. In a pilot study, a group of teachers/researchers from various Italian universities, coordinated by the author, attempted to connect three elements.

Firstly, the construction of curricula integrating both humanities and STEM disciplines is imperative. In addition, such curricula must incorporate artificial intelligence and foster critical thinking.

Secondly, it is necessary to analyse metacognitive management processes through a targeted revision of the MAI (Metacognitive Awareness Inventory) (Schraw & Dennison, 1994; Young & Fry, 2008; Balcikanli, 2011).

Thirdly, the development of an MAI-C (where C stands for critical thinking) (Ciasullo et al., 2025, 2026, in press) is essential in order to assess students' awareness of their level of critical engagement when using artificial intelligence in their learning processes.

The development of these analytical tools is grounded in the need to make distributed thinking processes observable and to support student–AI metacognitive

self-monitoring, thereby enabling a more precise evaluation of the awareness fostered through the use of artificial intelligence.

4. From AI-Agency to Co-Agency in Education

Co-Agency is defined as an educational configuration in which students, teachers and artificial intelligence collectively engage in the exploration, generation, negotiation and revision of knowledge (Dwivedi et al., 2026; Katsenou et al., 2025). Within this framework, AI is neither a neutral tool nor an autonomous agent replacing human decision-making. Rather, it is a cognitive co-agent whose role is intentionally designed, pedagogically mediated by the teacher, and metacognitively governed by learners (Romero, 2025). The transition from AI-agency to Co-Agency necessitates a fundamental rethinking of the integration of generative AI into educational practices. While generative AI offers unprecedented opportunities for creative experimentation and personalised learning, these possibilities are realised only when interaction is deliberately structured as a dialogic, reflective, and co-constructive process, rather than as a functional delegation of cognitive tasks to the machine (Lee et al., 2025; Tsakeni et al., 2025). The educational value of AI does not primarily reside in its technical capabilities, but rather in the relational and situated ways in which students and teachers engage with it. The concept of agency does not represent a static attribute of isolated entities; rather, it is an emergent phenomenon that develops through interaction. In dialogic learning environments, co-agency materialises through ongoing intra-action, in which human reasoning and machine-generated outputs continuously shape one another, giving rise to negotiated and non-linear learning trajectories (Casebourne et al., 2025).

Within this theoretical framework, the notion of AI as a neutral assistant must be considered an oversimplification. Instead, it functions as a semi-autonomous interlocutor whose suggestions, ambiguities, limitations, and generative affordances actively influence the direction of learning. Adopting a structured dialogical stance, which involves posing follow-up questions, negotiating meaning, problematising AI responses, and framing the system as a collaborative partner, has been demonstrated to result in higher levels of conceptual clarity, metacognitive awareness, and critical engagement (Uricchio & Cizek, 2023). In such contexts, the concept of co-agency emerges as both a visible and a productive phenomenon, effecting a transformation of AI-mediated interaction into a genuine learning space. Conversely, when AI is approached primarily as a shortcut to rapid answers, learning experiences often become cognitively shallow, with reduced opportunities for reflection and critical thinking. This discrepancy underscores the pivotal question of whether the crux of the issue lies in the utilisation of AI, or rather, in the pedagogical and metacognitive design of its interaction (Gerlich, 2025).

Generative AI technologies, encompassing advanced natural language processing systems, content creation tools and interactive learning assistants, have the potential to personalise education, enhance student engagement and reshape the production of educational materials (Casebourne et al., 2025; Islam et al., 2026). Nevertheless, these developments also profoundly alter classroom dynamics and the relationships between teachers, students, and technologies, making intentional pedagogical redesign essential. In this context, the role of the teacher becomes pivotal. The transition towards Co-Agency necessitates not only the acceptance of AI, but also

a transformation of teachers' professional agency (Frøsig & Romero, 2024). It is incumbent upon educators to assume the role of architects of hybrid learning environments, meticulously designing conditions under which artificial intelligence can facilitate personalised learning, adaptive feedback, and co-creative educational experiences. Such a transformation can only be achieved if educators are equipped with the skills to critically assess AI systems, address issues of bias and ethics, and ensure that technological integration enhances rather than constrains the learning ecology (Zhai, 2025).

5. Reframing the Teacher as Architect of Inter-Agency and Co-Agency

As has been demonstrated, the concept of co-agency is understood as a relational configuration in which students, teachers and artificial intelligence share — albeit in different ways — the process of exploring, generating and revising knowledge. This does not imply ontological symmetry between humans and artificial intelligence; rather, it describes a dynamic distribution of cognitive action.

It is evident that this triadic rewriting of teacher, learner and AI, while offering expanded possibilities for exploration, cognitive dialogue and enhanced investigative skills, gives rise to further risks that are by no means remote: passive delegation, automation of cognitive processes, teacher deskilling and student dependency. The fundamental issue remains the pedagogical ability to conceptualise these tools within educational contexts.

It is evident that overcoming the 'calculator' metaphor and granting IAGs the status of 'intellectual partners' (Katsenou et al., 2025), which can produce complex probabilistic solutions but require constant human validation, is imperative.

The objective is to establish a critical curriculum for AI that is both intentional and normatively aware, engaging in a continuous dialogue with pertinent regulations, ethical principles, and data justice issues. This curriculum should be interdisciplinary in nature and meticulously designed to distinguish between AI-generated output and human learning processes. Education is not solely concerned with AI literacy; it is also imperative to utilise AI effectively and, above all, to retain the capacity to teach without AI when pedagogically necessary (Hau, 2025).

In this context, the teacher's role as an architect of co-agency becomes evident. They function as a mediator, designer of hybrid environments, facilitator of metacognitive reflections, and regulator of the distribution of agency between the student and the artificial system.

The field of educational design, which includes the deliberate use of writing prompts as a pedagogical strategy, has emerged as a significant domain of professional responsibility.

In addition to the dissemination of knowledge, educators may expand their professional repertoire by assuming the roles of 'director of agency distribution' and 'experimental designer of educational processes'. The quality of teaching is no longer solely based on content, but also on the ability to manage the interaction between humans and artificial intelligence.

In order to succeed in this new profession, it is necessary to have "intelligent TPACK (Technological, Pedagogical, and Content Knowledge)", which combines technological, pedagogical, and content knowledge with AI literacy (Koehler et al., 2013; Rosenberg & Kohler, 2015; Celik et al., 2025).

The educator serves as the intermediary who determines the appropriate moments for AI to provide assistance to the student, and when it should be disengaged to encourage autonomous cognitive effort. The World Economic Forum (2025) asserts that educational transformation should be predicated on human-AI collaboration, thereby precluding the substitution of professionals.

According to the perspective of the United Nations Educational, Social and Cultural Organization (UNESCO) (Casebourne & Wegerif, 2025), the ultimate objective is not individual performance but the enhancement of collective intelligence. The potential of AI to promote global collaboration is evident in its capacity to facilitate the synthesis of diverse perspectives within citizen science projects and authentic problem-solving scenarios.

The loss of cognitive diversity remains a significant risk that requires urgent attention. The pervasive utilisation of standard AI models has the potential to engender a diminution in interpretative variety and a homogenisation of thought. In order to maintain the plurality of human voices, pedagogy must design hybrid ecosystems that preserve the plurality of human voices, and ensure that AI stays a tool for democratic participation and not a generator of intellectual conformity.

6. Implications for Educational Research and Practice

This study suggests that using Generative AI in education needs a change from just using it to a more planned way of teaching. The idea of educational architecture provides a way to organise how teachers, students and AI systems work together. To practise this, teachers need to change their role. They must be able to control what happens between different organisations. They also need to encourage people to think about how they are learning. The educational value of AI depends on how it is used in learning processes.

References

Atenas, J. et al. (2025). Critical and creative pedagogies for AI and data literacy. Research in Learning Technology.

Atkinson, R. (2025). Cognitive sovereignty and neurocomputational harm in predictive digital platforms. *Ethics and Information Technology*, *27*(4), 66.

Balcikanli, C. (2011). Metacognitive awareness inventory for teachers (MAIT).

Bygrave, L. A. (2020). Machine learning, cognitive sovereignty and data protection rights with respect to automated decisions. *Version*, *2*.

Casebourne, I., & Wegerif, R. (2025). AI and Education for collective intelligence. UNESCO.

Casebourne, I., Shi, S., Hogan, M., Holmes, W., Hoel, T., Wegerif, R., & Yuan, L. (2025). Using AI to Support Education for Collective Intelligence. *International Journal of Artificial Intelligence in Education*, *35*(3), 1597–1629. https://doi.org/10.1007/s40593-024-00437-7

Celik, I. et al. (2025). Co-constructing adaptive lesson plans with GenAI. Computers & Education: Artificial Intelligence.

Ciasullo, A. (2025). Sostenibilità ambientale e Intelligenze Artificiali: Coerenza Ecologica e Innovazione Tecnologica nei Contesti Educativi. *SOCIETÀ ITALIANA DI PEDAGOGIA*, 447-452.

Ciasullo, A., Casalino, G., Miranda, A., Saliceto, L., Barpi, F., & Taibi, D. (2025). From Reflection to Regulation: A Pilot Study on Metacognitive Awareness Inventory and the Use of Generative AI Tools in Higher Education. In *Book of Abstracts* (p. 229).

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, *58*(1), 7-19.

Devers, C. J., Devers, E. E., & Oke, L. D. (2018). Encouraging metacognition in digital learning environments. In *Digital workplace learning: Bridging formal and informal learning with digital technologies* (pp. 9-22). Cham: Springer International Publishing.

Dwivedi, Y. K., Helal, M. Y. I., Elgendy, I. A., Alahmad, R., Walton, P., Suh, A., Singh, V., & Jeon, I. (2026). Agentic AI Systems: What It Is and Isn't. *Global Business and Organizational Excellence*, *45*(3), 253–263. https://doi.org/10.1002/joe.70018

Frøsig, T. B., & Romero, M. (2024). *Teacher agency in the age of generative AI: Towards a framework of hybrid intelligence for learning design* (arXiv:2407.06655). arXiv. https://doi.org/10.48550/arXiv.2407.06655

Gallagher, S. (2011). The overextended mind. *Versus: Quaderni di studi semiotici*, *113*, 57-68.

Garrison, D. R. (2022). Shared metacognition in a community of inquiry. *Online learning*, *26*(1), 6-18.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006

Hau, D. (2025). Beyond the loop: Reclaiming pedagogy in an AI age. UNESCO Idea Lab.

Iqbal, J., Hashmi, Z. F., Asghar, M. Z., & Abid, M. N. (2025). Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers. *Scientific Reports*, *15*(1), 16610.

Islam, M. A., Somu, S., & Aldaihani, F. M. F. (2026). The Rise of Agentic AI: Synthesis of Current Knowledge and Future Research Agenda. *Global Business and Organizational Excellence*, *45*(3), 402–416. https://doi.org/10.1002/joe.70019

Katsenou, R., Kotsidis, K., Papadopoulou, A., Anastasiadis, P., & Deliyannis, I. (2025). Beyond Assistance: Embracing AI as a Collaborative Co-Agent in Education. *Education Sciences*, *15*(8). https://doi.org/10.3390/educsci15081006

Lee, H.-P. (Hank), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence

Effects From a Survey of Knowledge Workers. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, 1–22. https://doi.org/10.1145/3706598.3713778

Loginov, A. (2025). Digital metacognition: Awareness, regulation, and analysis of thinking in digital environments. *Next Research*, 100770.

Maniadakis, M., Trahanias, P., & Tani, J. (2012). Self-organizing high-order cognitive functions in artificial agents: Implications for possible prefrontal cortex mechanisms. *Neural Networks, 33*, 76–87. https://doi.org/10.1016/j.neunet.2012.04.005

Ostrom, T. M. (1984). The sovereignty of social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition,* Vol. 1, pp. 1–38). Lawrence Erlbaum Associates Publishers.

Romero, M. (2025). From Consumption to Co-Creation: A Systematic Review of Six Levels of AI-Enhanced Creative Engagement in Education. *Multimodal Technologies and Interaction*, *9*(10). https://doi.org/10.3390/mti9100110

Rosenberg, J. M., & Koehler, M. J. (2015). Context and technological pedagogical content knowledge (TPACK): A systematic review. *Journal of research on technology in education*, *47*(3), 186-210.

Santoianni, F. (2014). *Modelli di studio. Apprendere con la teoria delle logiche elementari*. Edizioni Centro Studi Erickson, Trento.

Santoianni, F. (2026). Reti esplicite & implicite. La collaborazione funzionale nei sistemi cognitivi. *RESEARCH TRENDS IN HUMANITIES Education & Philosophy*, *13*(1), 17-24.

Santoianni, F., & Ciasullo, A. (2024, September). AI Mediated Learning Architectures. In International Conference on Higher Education Learning Methodologies and Technologies Online (pp. 226-233). Cham: Springer Nature Switzerland.

Santoianni, F., Ciasullo, A., & Silva, L. (2025). Exploring Students' Research Engagement on Digital Technologies and AI. The MENS Group and UNIMORE Case Studies. In *7th International Conference Higher Education Learning Methodologies and Technologies Online 23-25 September 2025, Naples, Italy Book of Abstracts* (pp. 117-119). Editore STUDIUM srl.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology*, *19*(4), 460-475.

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education.

Tsakeni, M., Nwafor, S. C., Mosia, M., & Egara, F. O. (2025). Mapping the Scaffolding of Metacognition and Learning by AI Tools in STEM Classrooms: A Bibliometric–Systematic Review Approach (2005–2025). *Journal of Intelligence*, *13*(11), 148. https://doi.org/10.3390/jintelligence13110148

Uricchio, W., & Cizek, K. (2023). Co-creating with AI. *the minnesota review*, *2023*(100), 118–131. https://doi.org/10.1215/00265667-10320940

World Economic Forum (2025). How AI and human teachers can collaborate to transform education.

Yang, C. C., Chen, I. Y., & Ogata, H. (2021). Toward precision education. *Educational Technology & Society*, *24*(1), 152-163.

Young, A., & Fry, J. D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning*, *8*(2), 1-10.

Zhai, X. (2025). Transforming Teachers' Roles and Agencies in the Era of Generative AI: Perceptions, Acceptance, Knowledge, and Practices. *Journal of Science Education and Technology*, *34*(6), 1323–1333. https://doi.org/10.1007/s10956-024-10174-0