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

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Beyond AI Teammates: Collective Vicarious Cognition as a Framework for Educational Multi-Agent Systems

Lydia Zampolini [1], Fabrizio Lo Presti [2], Giuseppina Rita Jose Mangione [1] and Manuel Gentile [2]

[1] INDIRE – Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa; l.zampolini@indire.it ; g.mangione@indire.it
[2] ITD-CNR – Istituto per le Tecnologie Didattiche – Consiglio Nazionale delle Ricerche (Palermo); fabrizio.lopresti@itd.cnr.it ; manuel.gentile@itd.cnr.it

Abstract: Artificial intelligence is assuming an increasingly prominent role in educational contexts, yet the notion of the "AI teammate" is often reduced to the generative capabilities of Large Language Models (LLMs). Such a view preserves a tool-centred logic in which AI supports individual productivity without substantially contributing to collaborative learning processes. This paper argues that meaningful forms of human–AI collaboration in education require socio-technical architectures capable of organizing cognitive plurality and making reasoning processes visible, interpretable and open to evaluation. Drawing on Social Learning Theory, the concept of vicariousness, and perspectives on collective agency, the paper introduces the concept of Collective Vicarious Cognition (CVC) as a pedagogical framework for understanding learning as engagement with multiple, interacting cognitive trajectories. Within this perspective, learning emerges not simply through exposure to information, but through the observation, comparison, interpretation, and regulation of distributed forms of reasoning. Multi-agent systems based on LLMs (LLM-MAS) are examined as a possible educational infrastructure capable of supporting these conditions by distributing cognitive roles across interacting agents and rendering epistemic plurality observable. The paper argues that the educational significance of AI teammates lies less in their capacity to generate responses than in their potential to support organized cognitive diversity, reflective judgment, collective agency, and collaborative knowledge construction within hybrid human–AI learning environments.

Keywords: Collective Vicarious Cognition; Educational Multi-Agent Systems; Social Learning Theory; Collective Agency; Vicarious Learning; Human–AI Collaboration

1. Introduction: The Misunderstood Idea of the AI Teammate

The notion of the "AI teammate" has gained increasing attention in educational and organizational research. In many discussions, however, the term is associated primarily with the generative performance of Large Language Models (LLMs), framing Artificial Intelligence (AI) as an advanced assistant rather than as a structurally integrated collaborator. This interpretation preserves a fundamentally tool-centric

[1] The present work is the result of an ongoing scientific and methodological collaboration between INDIRE and ITD-CNR (Palermo). The authors have jointly contributed to the conceptualisation and development of the study; however, authorship of the individual paragraphs is attributed as follows: paragraphs 1 and 4 are attributed to Lydia Zampolini; paragraphs 2 and 6 are attributed to Giuseppina Rita Jose Mangione; paragraph 3 is attributed to Fabrizio Lo Presti and paragraph 5 is attributed to Manuel Gentile.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

logic, where the system enhances productivity but does not fundamentally transform the architecture of collaboration.

Research on human–AI teaming suggests a more demanding perspective. Seeber et al. (2020) argue that treating machines as teammates requires rethinking collaboration at the level of coordination structures, role allocation, and shared intentionality. A teammate is not merely an executor of subtasks but a participant in interdependent processes of sense-making and decision-making. Similarly, McNeese et al. (2021) emphasize that the question "Who or what is my teammate?" is not trivial: team composition, perceived agency, and role differentiation fundamentally shape interaction dynamics in human–AI teams. The teammate metaphor, therefore, implies structural integration, not simply functional assistance.

Empirical research further complicates the picture, as studies demonstrate that trust calibration, perceived competence and teammate identity significantly influence cooperation with AI systems (Zhang et al., 2023). Moreover, varying levels of AI influence within a team measurably affect human reliance, decision patterns, and performance outcomes (Flathmann et al., 2024). These findings indicate that the status of AI as teammate cannot be inferred from its computational sophistication alone; it depends on how influence, responsibility, and epistemic authority are distributed within the team. When epistemic authority is opaque or unbalanced, collaboration risks becoming automation; when influence is transparent and role-differentiated, it can foster reflective coordination.

From a pedagogical perspective, this distinction is decisive, as education is not merely concerned with task completion but with the cultivation of agency, judgment and reflexivity. If educational discourse adopts the teammate metaphor without addressing these structural and relational dimensions, it risks conflating conversational fluency with collaborative agency. An AI system that merely generates responses remains within a proxy logic (Bandura, 2001), supporting individual action without participating in collective cognitive processes. Proxy agency may increase efficiency, but it does not necessarily support the development of self-efficacy or collective agency. A genuinely collaborative AI, by contrast, must be embedded in an architecture that organizes differentiated roles, renders reasoning processes observable and supports shared regulation of cognitive work. The shift from assistance to participation, therefore, entails a transformation of the learning environment itself: from individual-tool interaction to distributed cognitive engagement. From a pedagogical standpoint, this transformation is significant because learning depends not only on access to answers but also on opportunities to observe, compare, and evaluate alternative ways of reasoning. Learning environments become educationally meaningful when cognitive processes are made visible and discussable, thereby supporting the development of metacognitive awareness, judgment and learner agency.

This shift resonates with broader educational debates emphasizing plurality as an epistemic condition of learning in complex environments (Morin, 1999; Biesta, 2010). Contemporary learning contexts require the capacity to confront divergent perspectives, articulate criteria, and negotiate meaning in the face of uncertainty. However, plurality becomes pedagogically meaningful only when intentionally structured. Without organization, multiplicity may lead to fragmentation rather than understanding. The educational problem is not the presence of multiple perspectives, but the absence of architectures capable of rendering them comparable, discussable, and regulable. In educational terms, plurality becomes valuable only when learners can examine differences, articulate criteria for evaluation, and justify their judgments. The pedagogical challenge is therefore not simply to expose learners to multiple perspectives, but to transform cognitive diversity into a resource for reflection, comparison and meaning-making. The question, then, is how learning environments can be designed to make such organized plurality observable and pedagogically productive. Rather than asking whether AI can replace human participation, the educational issue becomes whether AI systems can support forms of modeling, comparison, deliberation and shared regulation that enhance learners' understanding of complex cognitive processes.

LLM-based Multi-Agent Systems (LLM-MAS) offer a possible architectural direction for addressing this challenge. Instead of concentrating interaction around a single generative model, LLM-MAS configurations organize multiple agents operating under differentiated roles, thereby introducing a structural form of plurality within human–AI interaction. Such configurations resonate with perspectives on distributed cognition (Hutchins, 1995), in which intelligent activity emerges from the organization of interactions rather than from isolated entities. The relevance of this approach, for education, lies not simply in technological sophistication but in its potential to render cognitive processes observable and comparable within the learning environment. Under these conditions, AI systems may begin to participate in structured processes of modelling and deliberation rather than merely generating responses.

Therefore, the transition from "AI as tool" to "AI as teammate" does not hinge on generative capacity, but on the structural organization of plurality and agency within the learning environment. What matters educationally is not the production of answers itself, but the possibility for learners to engage with reasoning as an observable, comparable, and regulable process. The central thesis of this paper can therefore be stated explicitly: an AI system qualifies as a teammate when it participates in structured modeling processes that render an organized plurality of cognitive strategies observable, comparable, and pedagogically regulable within a learning ecology. Under these conditions, distributed modeling becomes not a by-product of technological sophistication, but a deliberate educational resource.

2. Vicariousness, Agency, and Collective Vicarious Cognition in Learning

Reinterpreting the role of AI in education requires more than a technological reframing; it demands a re-articulation of the pedagogical categories through which learning itself is understood. If AI is to be conceived as a participant in learning environments, then the concepts of modeling, plurality and agency must be reconsidered within a coherent theoretical horizon. Social Learning Theory (SLT) (Bandura, 2001; Bandura, 1997), together with the concept of vicariousness articulated by Berthoz (2013) and the biological principle of degeneracy (Edelman & Gally, 2001), provides a fertile foundation for this reconfiguration. The aim of this section is not merely to review these traditions, but to show how they converge toward a pedagogical framework for understanding learning as engagement with visible, plural, and regulable cognitive processes. In this sense, vicariousness is not treated only as imitation or indirect experience, but as an educational condition through which learners encounter alternative ways of acting, reasoning, and judging.

Bandura's original formulation of SLT displaced strictly behaviorist accounts by emphasizing that learning is mediated by cognitive processes and occurs significantly through observation (Bandura, 1997). Observational learning entails attention, retention, symbolic processing, reproduction and motivational regulation. What is pedagogically decisive is that learners do not passively absorb behaviors, they interpret them; they compare observed strategies with their own competencies, anticipate possible consequences and regulate future action accordingly. Modeling is therefore inherently relational and evaluative. This point is crucial for education: the pedagogical value of modeling lies not in the mere presentation of an exemplar, but in the learner's active interpretation of the model's choices, constraints, and consequences.

Observing a model means entering into a comparative relation with a possible course of action. Vicarious learning is therefore already a form of mediated judgment.

The development of self-efficacy beliefs, understood as individuals' perceived capability to organize and execute courses of action (Bandura, 1997), is central to this process. Self-efficacy is shaped not only by direct mastery experiences but also by vicarious experiences. Observing others' successes and failures provides information about what is possible and about the effort required. In this sense, visibility of process is not ancillary to learning; rather, it is constitutive of agency formation. Contemporary research confirms that self-efficacy is closely linked to metacognitive regulation, persistence, and adaptive learning behaviors (Schunk & DiBenedetto, 2020). When learners encounter transparent reasoning trajectories, they are better positioned to calibrate their own judgments and to sustain effort in the face of difficulty. For this reason, visibility should be understood as a pedagogical condition rather than as a technical feature. When learners can follow the development of reasoning, they are not only informed about a result; they are invited to regulate their own understanding, compare strategies, and reposition themselves as capable participants in the learning process.

However, observational learning does not occur in isolation. Bandura's later elaboration of Social Cognitive Theory (SCT) situates learning within a system of reciprocal determinism between person, behavior and environment (Bandura, 2001). The distinction among personal, proxy, and collective agency becomes particularly relevant in techno-logically mediated contexts. Proxy agency—acting through others—can be pedagogically ambivalent: on the one hand, it allows learners to access resources beyond their immediate competence, on the other, it may reduce opportunities for self-regulation if delegation replaces engagement. The educational integration of AI must therefore be evaluated not only in terms of performance outcomes but in terms of how it redistributes and configures agency. In educational settings, this redistribution is decisive because AI can either displace learner agency or scaffold it. When learners delegate cognitive work without access to the process, proxy agency may weaken self-regulation. When, however, delegated processes remain visible, contestable, and interpretable, they can become resources for reflection and agency formation.

Collective agency introduces an additional layer of complexity. It refers to the shared belief in the conjoint capability to organize and execute actions required for collective goals (Bandura, 2001). Research in collaborative learning and socially shared regulation demonstrates that learning deepens when participants engage in joint planning, monitoring and evaluation (Volet et al., 2009, Järvelä & Hadwin, 2013). The pedagogical challenge is to design environments where distributed contributions do not fragment responsibility but instead foster coordinated regulation. Any technological system positioned as a "teammate" must therefore be examined considering its impact on collective agency: does it centralize epistemic authority, or does it scaffold shared regulation? This question prepares the ground for Collective Vicarious Cognition. If collective agency concerns the shared capacity to act toward common goals, CVC concerns the shared possibility of learning from the visible interplay of differentiated cognitive contributions. The emphasis shifts from simply acting together to observing, interpreting, and regulating how different forms of reasoning interact.

At this juncture, the concept of vicariousness offers a structural complement to the social-cognitive account. Berthoz (2013) defines vicariousness as the capacity of living systems to achieve the same objective through different functional configurations. Intelligence, from this perspective, is not a matter of linear optimization but of flexible substitution. The related concept of degeneracy (Edelman & Gally, 2001) further clarifies that robustness in complex systems derives from the coexistence of

structurally different elements capable of performing equivalent functions. Variability is not noise, it is a condition of adaptability. This means that learning should not be organized around a single privileged pathway whenever complex problems admit multiple legitimate strategies. The educational value of vicariousness lies in making alternative pathways
available for comparison. Learners do not simply learn that a solution exists; they learn that different configurations of reasoning may lead toward comparable goals.

Pedagogically interpreted, vicariousness suggests that learning environments should cultivate access to differentiated strategies oriented toward similar goals. However, this plurality must be carefully distinguished from relativism. The existence of multiple pathways does not imply the suspension of evaluative criteria. Research in epistemic cognition indicates that advanced learners are characterized not by indiscriminate acceptance of multiple perspectives, but by the capacity to evaluate competing claims according to disciplinary standards (Muis et al., 2015, Greene et al., 2018). Plurality becomes educationally generative only when differences are made visible and subjected to reflective judgment. Without such structuring, multiplicity risks producing epistemic confusion rather than adaptive competence. Therefore, the pedagogical problem is not plurality itself, but the regulation of plurality. CVC emerges precisely at this point: it names a learning condition in which multiple cognitive trajectories are not merely presented side by side, but organized so that learners can compare their assumptions, evaluate their limits, and understand their possible complementarity.

Within this framework, it is useful to differentiate between simple and systemic vicariousness. This distinction allows vicariousness to be understood at two pedagogical levels. Simple vicariousness occurs when learners observe a single model whose performance functions as an exemplar. Systemic vicariousness, by contrast, emerges when learners observe interacting cognitive processes in which differentiated strategies confront, critique, and refine one another. In systemic vicariousness, plurality is architecturally organized rather than incidental. The learner is exposed not merely to an answer but to a dynamic interplay of alternatives, each revealing underlying assumptions, criteria, and constraints. This exposure supports what Schwartz, et. al, (2005) describe as adaptive expertise: the capacity to balance efficiency with innovation and to transfer knowledge flexibly across contexts. Systemic vicariousness is therefore especially relevant for educational environments concerned with complex problem solving, creativity, and collaborative inquiry. In these contexts, learners benefit not only from seeing what works, but from understanding how alternatives are generated, challenged, revised, and integrated.

Systemic vicariousness therefore expands the pedagogical significance of modeling. Instead of internalizing a singular trajectory, learners engage in comparative analysis of interacting strategies. They observe not only success but disagreement, revision, and synthesis. Such environments can strengthen metacognitive awareness, as learners must interpret contrasts and articulate reasons for preferring one configuration over another. In doing so, they exercise judgment rather than mere reproduction.

On this theoretical ground, CVC can be defined as a pedagogical condition in which learners engage with plural, differentiated, and interacting cognitive trajectories made visible within a shared learning environment. It integrates the structural substitutability described by vicariousness (Berthoz, 2013), the observational and agentic dynamics of SCT (Bandura, 1997; 2001), and the coordinated regulation processes identified in collaborative learning research (Järvelä & Hadwin, 2013). In such a configuration, cognition is neither isolated nor delegated; it is distributed and made visible. Learning emerges from interpreting the relationships among strategies rather than merely receiving outputs and the pedagogical horizon shifts from automation to organized plurality, and from output optimization to the cultivation of reflective, distributed agency.

Within this perspective, CVC also functions as a conceptual hinge linking theories of observational learning with the design of contemporary learning environments. If learning emerges through engagement with organized plurality, the pedagogical question cannot be reduced to whether an AI system produces accurate outputs. Rather, it concerns whether the learning environment makes multiple cognitive trajectories observable, comparable and open to reflective evaluation. In this sense, the unit of analysis shifts from isolated human–tool interaction to the broader ecology of interactions through which modeling processes become distributed and visible, a move consistent with perspectives on distributed cognition (Hutchins, 1995).

Such a shift raises the question of what kinds of socio-technical architectures might sustain these conditions of organized plurality. From this standpoint, LLM-MAS become theoretically relevant, not primarily as technological innovations but as potential infrastructures capable of organizing differentiated cognitive processes within hybrid human–AI learning environments. The following section therefore examines LLM-MAS not as intelligent tools in themselves, but as possible pedagogical architectures for sustaining CVC.

3. Multi-Agent Systems as Architectures of Organized Plurality

If learning can emerge from observing organised pluralities of cognitive strategies, it is essential to understand through which configurations GenAI can support collective vicarious learning. Historically, the first examples of AI agency can be observed in rule-based agents, which base their behaviour on static decision trees and are therefore unable to learn or manage unexpected scenarios. The advent of LLMs has enabled a real paradigm shift and change in the structure of AI-based agents, transforming them from static and rigid entities into architectures capable of learning and dynamically adapting to changing contexts.

Currently, user interaction with AI-based agents occurs primarily through the mediation of conversational agents such as ChatGPT or Claude (Anthropic, 2024; OpenAI, 2026). Unlike traditional rule-based systems, conversational agents are distinguished by the integration of an LLM as their central cognitive engine. This architecture ensures exceptional natural language processing (NLP) capabilities, freeing interaction from pre-programmed paths and enabling dynamic, contextualised and semantically rich dialogue with the user. Such systems, while falling within the broader definition of artificial entities capable of perceiving, reasoning and acting in their environment (Russell & Norvig, 2003), essentially operate as advanced personal assistants, answering questions, clarifying concepts and generating material on demand.

Interaction is therefore confined to linear prompt sequences: the user enters a command, the tool generates an output and waits for the next instruction. Consequently, although these tools are effective in supporting the user in relatively simple and linear contexts (Lo Presti et al., 2025), they maintain a tool-centric logic, i.e. they are considered passive tools whose usefulness is limited to the reactive execution of a task imposed from outside. To overcome this limitation, these interfaces are evolving into truly autonomous intelligent agents. In this paradigm, the system does not merely respond to isolated queries but proactively organises a work plan to achieve a set goal. To do this, the LLM ceases to perform the simple task of generating text and becomes the logical core of a more complex architecture consisting of three pillars:

- A system profile that defines its expertise, objective and operational constraints;
- A memory module useful for long-term preservation of the scenario context;
- Access to external tools, allowing the agent to query external sources and exe-cute code.

Despite the potential offered by this configuration, the structural limitation of a single-agent approach lies in the fact that the human user will be confronted with a monolithic perspective, lacking in plurality. To use an analogy, asking a single agent to support the management of complex educational scenarios is equivalent to expecting a single individual to take care of the entire value chain for the creation of a product, from market analysis to design, production and marketing. As in real professional contexts, complex scenarios that require problem setting and problem solving skills need a team of people who bring their expertise to the table, contributing to the definition of the entire action plan rather than just providing an isolated output.

It is at this juncture that the need for a further architectural leap becomes apparent. Multi-Agent Systems (MAS) are configured as infrastructures of autonomous agents that interact and coordinate within a shared context, generating collective behaviours that are functional to the management of complex and distributed systems (Luzolo et al., 2024). AI-enhanced MAS can structurally organise plurality by building an ecosystem that features:

- The assignment of differentiated roles to the agents that make up the system, to optimise performance through domain specialisation and reduce the model's hallucination rate;
- The introduction, in addition to user-system dialogue, of inter-agent dialogue, in order to enable a dynamic exchange of contextual information and enable self-correction, making the system capable of moving forward without the need for con-tinuous external prompts;
- The explicit reasoning traces of individual agents, in order to ensure the trans-parency of internal decision-making processes;
- The coordination of work, carried out by a specialised agent (supervisor). This ensures that interactions remain focused on the objective and the shared resolution of the problem.

From an educational perspective, this type of configuration introduces significant differences in the way in which knowledge construction processes can be made visible and observable.

4. Multi-Agent Systems as Architectures for Structured Plurality in Learning

Building on this perspective, LLM-MAS can be interpreted not only as a technological solution for managing complex systems, but also as a possible architecture capable of implementing the cognitive plurality assumed by the theoretical frame outlined. Rather than concentrating interaction around a single generative model, LLM-MAS distribute cognitive functions among differentiated agents operating according to distinct roles (Stone & Veloso, 2000; Lo Presti et al., 2025; Zampolini et al., 2025a), with a view to distributed cognition (Hutchins, 1995), according to which intelligent performance emerges from the organisation of interactions between multiple entities rather than from the action of a single isolated agent. Within this framework, CVC can be understood as the pedagogical process through which learners engage with multiple visible cognitive trajectories unfolding within a shared environment. Rather than learning from a single model or solution pathway, learners observe how different perspectives generate hypotheses, evaluate alternatives, challenge assumptions and progressively coordinate their contributions. Learning therefore emerges not only from exposure to knowledge, but from the interpretation of relationships among competing and complementary forms of reasoning.

Furthermore, the focus of learning also shifts: it is no longer the single interaction between user and tool that forms the heart of the educational experience, but rather the ecology of interactions through which different modelling processes become simultaneously observable, comparable and adjustable. n pedagogical terms, the object of learning expands from knowledge products to knowledge-building processes. Learners are encouraged to examine how ideas emerge, evolve, and are evaluated rather than focusing exclusively on final answers.

The pedagogically decisive point is that this architecture makes the differentiation of cognitive processes explicit: activities such as hypothesis generation, critical evaluation, information verification, or the synthesis of alternatives can be organised as distinct processes that interact with one another, rather than being compressed into a single output produced by a single model.

In this type of setup, reasoning paths become observable parts of the learning environment. Students are no longer positioned solely as recipients of AI-generated content, but as participants in an environment in which multiple strategies are articulated, compared and potentially reorganised. Viewed in this light, the educational value of AI shifts from simply producing answers to organising a structured plurality of cognitive processes. This condition is central to CVC because it transforms cognitive diversity into a pedagogical resource. The visibility of alternative reasoning trajectories enables learners to compare approaches, identify assumptions, and exercise judgment regarding the adequacy of different strategies. This organised plurality comes together through specific architectural features typical of LLM-MAS, which are especially important when looked at from an educational point of view

First, LLM-MAS allow different roles to be assigned to the agents that make up the system (Lo Presti et al., 2025; Zampolini et al., 2025a). Each agent can specialise in a specific cognitive function, for example, generating solutions, critical evaluation, verifying information or summarising alternatives. While from a technical standpoint, this specialisation improves the overall performance of the system, its pedagogical significance is even more relevant: the differentiation of roles makes explicit the plurality of cognitive strategies through which the same problem can be addressed. Students no longer observe a single answer produced by a model, but a set of reasoning trajectories that reflect different perspectives. This comparative exposure supports forms of vicarious learning that extend beyond imitation. Learners can analyse why different strategies emerge, under which conditions they are effective and how distinct cognitive roles contribute to the construction of shared understanding. In this sense, plurality is not a side effect of the system, but a structurally organised property of the learning environment.

The transition from a single LLM-based agent to an LLM-MAS ecosystem introduces a substantial difference in the way knowledge construction processes can be observed and interpreted. The assignment of differentiated roles to the agents of the system, based on pedagogical principles (Zampolini et al., 2025b), can be interpreted as a form of epistemic plurality structurally incorporated into the architecture. Through the specialisation of agents, the system is able to replicate the diversity of perspectives that characterises many real professional contexts, allowing complex problems to be addressed from different angles.

A second important feature concerns the introduction of inter-agent dialogue. In LLM-MAS systems, interaction is not limited to the user-system relationship but includes exchanges between the agents themselves. This dialogue allows agents to share contextual information, compare hypotheses and correct each other's inferences.

From a computational perspective, these mechanisms allow the system to progress in solving the problem without requiring continuous external intervention. From a pedagogical view, however, they make deliberation processes observable that normally remain implicit: the comparison between agents exposes students to dynamics of agreement, disagreement and revision of hypotheses, transforming, for example, the problem-solving process into an explicitly dialogical phenomenon. Thus, the negotiation of meanings takes the form of a cognitive conflict that is made explicitly visible to the student. The user is no longer the recipient of authoritative output, but rather an observer of a dynamic in which ideas are proposed, criticised, defended and progressively refined through the interaction of different perspectives. Such processes are pedagogically valuable because they externalize forms of cognitive conflict that are often hidden in individual reasoning. Learners gain access not only to conclusions but also to the argumentative dynamics through which knowledge claims are negotiated and revised.

A third aspect concerns the possibility of making individual agents' reasoning traces explicit. In many LLM-MAS architectures, the intermediate stages of reasoning can be made accessible and analysable. This transparency allows us to observe both the final outcome of the decision-making process and the cognitive operations that produced it. In terms of SLT, the ability to observe and interpret these processes is crucial: observational learning depends on the subject's ability to attribute meaning to the strategies adopted by the observed models and to evaluate their controllability and effectiveness (Bandura, 1997). LLM-MAS configurations allow these intermediate stages of reasoning to be externalised, making visible how different strategies evolve through criticism, revision and negotiation. The educational significance of this visibility lies in its capacity to support metacognitive regulation. By observing how alternative reasoning pathways develop and change, learners can reflect on their own cognitive strategies and evaluate possible courses of action.

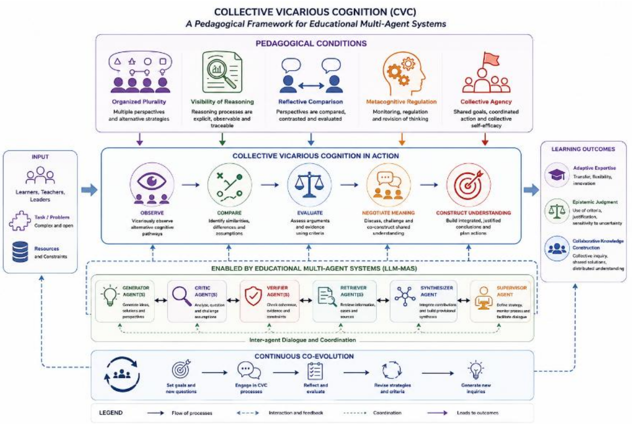
By explaining the intermediate logical steps, the alternatives considered and discarded and the collective decision-making processes, the cognitive path of the system is thus transformed into a real, observable educational artefact. From this perspective, learning takes place through observation of the strategies that led to its construction. This condition creates the conditions for vicarious learning, in which students internalise complex cognitive strategies through the interpretation of the processes observed. When the reasoning trajectories and cognitive roles of agents are made explicit and interpretable, students can observe, evaluate and compare different strategies, activating forms of reflective regulation that research on collaborative learning identifies as central to deep learning processes (Volet et al., 2009; Järvelä & Hadwin, 2013).

Moreover, LLM-MAS introduce coordination mechanisms that regulate interaction between agents. This role is often assigned to a supervisory agent tasked with orchestrating the system's work and keeping interactions focused on achieving the common goal. From a pedagogical perspective, these coordination mechanisms recall the shared regulation dynamics studied in the literature on collaborative learning. The presence of a level of orchestration does not eliminate the plurality of perspectives, but organises their interaction, allowing different strategies to contribute to the construction of a shared solution. In CVC terms, coordination is not intended to suppress diversity but to make diversity educationally productive. The pedagogical goal is not consensus for its own sake, but the creation of conditions under which differences can be examined, compared, and integrated through reflective judgment. For this plurality to become pedagogically generative, however, it cannot be limited to a simple proliferation of perspectives: differences must be made salient, comparable and assessable according to explicit criteria. Research on epistemic cognition suggests that

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

advanced epistemic competence does not consist in the indiscriminate acceptance of multiple perspectives, but in the ability to critically compare them in light of disciplinary and argumentative criteria (Muis et al., 2015; Greene et al., 2018).

Figure 1. Collective Vicarious Cognition (CVC) as a pedagogical framework for educational multi-agent systems. The figure illustrates how organized plurality, visibility of reasoning, reflective comparison, metacognitive regulation, and collective agency create the conditions through which learners engage with multiple cognitive trajectories and develop adaptive expertise, epistemic judgment, and collaborative knowledge construction.



Emerging evidence suggests that such configurations may indeed transform the role of AI in educational contexts. The LLM-MAS architecture proposed by Hou et al. (2025), for example, introduces a conversational assessment system based on agents with distinct pedagogical roles coordinated by a supervisory agent. In this context, the assessment process emerges from the negotiation between different perspectives. Similarly, the work of Gajewska et al. (2025) shows how LLM-MAS can be used to generate complex problem-setting scenarios in educational contexts, allowing teachers to dynamically explore decision-making alternatives and analyse the consequences of their choices in simulated environments. Other studies highlight further educational applications of MAS based on multiple perspectives. Mushtaq et al. (2025), for example, explore the use of LLM-MAS in engineering problem solving, simulating a multidisciplinary team of agents collaborating on problem definition and resolution. In different contexts, Harbar and Chudziak (2025) and Park and Seo (2025) apply LLM-MAS architectures to structured debate scenarios, showing how interaction between agents with divergent positions can stimulate and evaluate students' critical thinking.

Taken together, these contributions suggest that LLM-MAS can support forms of distributed reasoning in which multiple perspectives are made explicit and comparable. However, such work tends to focus primarily on the functional outcomes of LLM-MAS interactions, while their interpretation as pedagogical infrastructures for systemic vicariousness and the development of collective agency remains less explored. From this perspective, the educational value of LLM-MAS lies less in their capacity to automate cognitive work than in their capacity to make cognitive work observable. Their pedagogical significance resides in supporting environments where learners can engage with organized plurality, exercise reflective judgment, and develop forms of collective agency through participation in processes of CVC.

5. Risks and Boundaries: When Collective Vicarious Cognition Breaks Down

While LLM-MAS offer a possible infrastructure for organising forms of observable cognitive plurality, their integration into educational contexts also raises a series of risks and limitations that must be carefully considered. The transition from a tool-centred interaction model to human-AI collaboration configurations requires precise design conditions; in their absence, what appears to be collaboration can easily degenerate into sophisticated forms of automation. From the perspective of CVC, the central question is not simply whether AI systems perform effectively, but whether they sustain the pedagogical conditions that make learning through organized plurality possible. An LLM-MAS may be technically successful while pedagogically ineffective if learners are unable to interpret, compare, evaluate, or regulate the cognitive processes unfolding within the system. The boundaries of AI as teammate therefore coincide with the boundaries of meaningful participation in collective cognitive activity.

A first risk concerns the opacity of the deliberative processes within the system. For vicarious learning to take place, the strategies adopted by the observed models must be interpretable. If the dialogues between agents and the intermediate stages of reasoning remain invisible or difficult to understand, the system risks functioning as a deliberative black box. Under such conditions, the user observes only the final result of the decision-making process, losing access to the cognitive trajectories that constitute the true object of observational learning (Bandura, 1997). In such circumstances, CVC collapses because learners can no longer access the diversity of reasoning trajectories that should constitute the basis for comparison and reflective judgment. What remains is the observation of outcomes rather than engagement with cognitive processes.

A second risk concerns the possible amplification of forms of proxy agency. As Bandura (2001) points out, action mediated by other agents can facilitate the achievement of complex goals, but it can also reduce opportunities for personal agency if delegation replaces active participation. In LLM-MAS systems, this risk arises when deliberation between agents takes place entirely within the system, relegating the user to the role of mere observer or validator of the decisions produced. From a pedagogical perspective, this represents a shift from participation to cognitive outsourcing. Rather than engaging with alternative strategies, learners may become dependent on the system's deliberations, weakening opportunities for self-regulation, metacognitive monitoring and the development of adaptive expertise.

A further limitation concerns the management of epistemic plurality. The presence of multiple perspectives does not in itself guarantee deeper learning. If plurality is not accompanied by explicit criteria for comparison and evaluation, it can generate interpretative fragmentation rather than understanding. Research on epistemic cognition shows that advanced epistemic competence does not consist in the indiscriminate acceptance of different perspectives, but in the ability to critically compare them in light of disciplinary and argumentative criteria (Muis et al., 2015, Greene et al., 2018). Without such criteria, learners may confuse epistemic diversity with epistemic equivalence, treating all perspectives as equally valid regardless of their evidential or argumentative quality.

A related but distinct risk concerns the possibility of superficial plurality. The mere presence of multiple agents does not necessarily generate meaningful cognitive diversity. Agents may reproduce similar assumptions, rely on shared training biases, or converge rapidly toward homogeneous solutions. In such cases, plurality becomes simulated rather than substantive. From the standpoint of CVC, educational value depends not on the number of voices involved, but on the visibility of genuinely differentiated perspectives capable of generating productive comparison, cognitive conflict, and reflective judgment.

Finally, there is a risk that the metaphor of the "teammate" masks a form of collaboration that is only apparent: a system can simulate dialogic dynamics and deliberative processes without the user actually participating in the regulation of the cognitive process. In such cases, the interaction remains essentially asymmetrical, with AI producing and organising solutions, while humans merely approving them. Under these conditions, the appearance of collaboration masks the concentration of epistemic authority within the system itself. Learners may be exposed to the outcomes of collective reasoning without participating in the interpretative and regulatory processes through which those outcomes are produced. Rather than fostering forms of collective agency, the system risks transforming learners into consumers of organised cognition. From the perspective of CVC, this represents a critical failure, since learning depends not only on the visibility of multiple cognitive trajectories but also on the learner's active engagement in interpreting, comparing, and evaluating them.

Rather than supporting forms of collective agency, such configurations risk consolidating cognitive dependence on the system. These limitations suggest that the pedagogical value of LLM-MAS does not depend solely on their technical sophistication, but on the extent to which their architecture supports meaningful participation in collective cognitive activity. From the perspective of CVC, educationally valuable systems are not those that simply produce better answers, but those that make cognitive diversity visible, interpretable, and open to reflective judgment. Only when reasoning trajectories remain accessible, contestable, and subject to evaluation can learners engage with organized plurality as a resource for agency development, metacognitive regulation and adaptive expertise. Under these conditions, LLM-MAS may function as infrastructures for collective vicarious learning; otherwise, organized plurality risks becoming little more than a sophisticated form of automation.

6. Pedagogical Perspectives and Future Directions

The analysis developed in this paper suggests that the notion of the "AI teammate" cannot be reduced to the growing capabilities of generative models. What is at stake is a transformation in how educational environments organize cognitive activity and render reasoning processes visible. When AI systems participate in distributed processes of modeling, the locus of engagement shifts beyond the individual interaction with a tool toward hybrid configurations in which reasoning unfolds through the interaction of multiple agents and supports both learning processes and collective reflection within educational communities.

The pedagogically decisive issue concerns the organization of plurality. Educational environments become particularly fertile when different strategies, interpretations, and reasoning trajectories can be observed and compared. The educational relevance of AI therefore, lies less in the generation of answers than in its capacity to make the dynamics of reasoning visible and interpretable, enabling forms of vicarious observation through which alternative cognitive pathways can be examined and appropriated.

LLM-MAS offer a promising architectural configuration in this regard: by distributing cognitive roles across differentiated agents and enabling interaction among them, LLM-MAS architectures can render visible the processes through which hypotheses are proposed, challenged and progressively refined. Knowledge appears not as a finished output but as a trajectory shaped through the interaction of multiple perspectives, making the principle of vicariousness operational within socio-technical learning environments. More fundamentally, the framework proposed here suggests a shift in the pedagogical unit of analysis. Rather than focusing primarily on the relationship between an individual learner and a source of information, CVC directs attention toward the organization of visible cognitive trajectories within a shared learning environment. Learning becomes a process of engaging with, interpreting, and evaluating distributed forms of reasoning rather than simply acquiring knowledge from an authoritative source. In this perspective, agency is exercised through participation in the interpretation and regulation of organized cognitive plurality.

Several educational scenarios emerge from such configurations. LLM-MAS environments may support simulated interdisciplinary studies, where agents embody distinct epistemic perspectives and participants observe how different disciplinary approaches frame and interpret shared problems. LLM-MAS architectures can also sustain policy simulation environments, allowing educational communities to explore complex organizational or societal issues through structured encounters between competing viewpoints.

Another direction concerns the development of hybrid innovation laboratories, where teachers, school leaders, and AI agents collaboratively explore design alternatives and problem-solving strategies connected to educational practice and institutional development. In such contexts, reflection is not limited to the learning task itself but extends to the organizational and strategic dimensions of school life. These examples point toward a broader educational challenge. If hybrid human–AI teams become increasingly common in educational contexts, pedagogical research will need to investigate not only learning outcomes, but also the conditions under which cognitive diversity becomes educationally productive. Questions concerning judgment, meta-cognitive regulation, collective agency and epistemic responsibility become central in environments where multiple human and artificial actors participate in knowledge construction.

LLM-MAS environments may also function as spaces dedicated to cognitive conflict, where disagreement becomes a visible and productive component of reasoning processes. Such environments allow participants to experience how alternative interpretations confront one another and contribute to the progressive clarification of problems. Future studies may therefore explore how different configurations of organized plurality influence learning processes. Particular attention should be devoted to understanding how learners navigate competing perspectives, how evaluative criteria are developed, and how participation in distributed reasoning environments contributes to the development of adaptive expertise and collective self-efficacy.

Through sustained engagement with plural reasoning processes, groups may develop forms of collective self-efficacy grounded in the experience of coordinated cognitive activity across multiple perspectives. In this sense, collective agency emerges not through the elimination of differences, but through the capacity to interpret, coordinate, and work productively with cognitive diversity. The educational significance of organized plurality therefore lies not only in improving problem solving, but in cultivating dispositions for collaborative inquiry and reflective participation in complex socio-technical environments. These environments can also contribute to preparing educational communities for complex socio-technical contexts, where decision-making increasingly emerges from interactions among networks of human and artificial agents. LLM-MAS can therefore be interpreted not simply as advanced educational technologies, but as emerging pedagogical architectures for organizing cognitive plurality.

The concept of Collective Vicarious Cognition proposed in this paper provides a framework for understanding how learning may emerge through engagement with visible, differentiated, and interacting forms of reasoning. In this sense, vicariousness is no longer confined to observing a single model, but becomes a collective process through which learners engage with multiple cognitive pathways, compare alternative strategies, and develop shared capacities for judgment and action. As educational environments become increasingly populated by artificial agents, the central pedagogical challenge may no longer be access to information itself, but the design of conditions under which cognitive diversity becomes a shared resource for judgment, agency, collective inquiry, and collaborative knowledge construction.

References

Anthropic. (2024). Claude [Large language model]. Anthropic.

Bandura, A. (1997). Self-efficacy: The exercise of control. W.H. Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. Annual Review of Psychology, 52(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Berthoz, A. (2013). La vicariance: Le cerveau créateur de mondes. Odile Jacob.

Biesta, G. J. J. (2010). Good education in an age of measurement: Ethics, politics, democracy. Routledge. https://doi.org/10.4324/9781315634319

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. Proceedings of the National Academy of Sciences, 98(24), 13763–13768. https://doi.org/10.1073/pnas.231499798

Flathmann, C., McNeese, N. J., Schelble, B., Knijnenburg, B., & Freeman, G. (2024). Understanding the impact and design of AI teammate etiquette. Human–Computer Interaction, 39(5–6), 444–471. https://doi.org/10.1080/07370024.2023.2189595

Gajewska, E., Wawer, M., Budzynska, K., & Chudziak, J. A. (2025). Leveraging a multi-agent LLM-based system to educate teachers in hate incidents management. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), Artificial Intelligence in Education: 26th International Conference, AIED 2025, Palermo, Italy, July 22–26, 2025, Proceedings, Part I (pp. 332–339). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-98462-4_42

Greene, J. A., Sandoval, W. A., & Bråten, I. (Eds.). (2018). Handbook of epistemic cognition. Routledge. https://doi.org/10.4324/9781315795225

Harbar, Y., & Chudziak, J. A. (2025). Simulating Oxford-style debates with LLM-based multi-agent systems. Lecture Notes in Computer Science, 286–300. https://doi.org/10.1007/978-981-96-6008-7_21

Hou, X., Forsyth, C., Andrews-Todd, J., & Graesser, A. C. (2025). An LLM-enhanced multi-agent architecture for conversation-based assessment. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), Artificial Intelligence in Education: 26th International Conference, AIED 2025, Palermo, Italy, July 22–26, 2025, Proceedings, Part III (pp. 119–134). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-98417-4_9

Hutchins, E. (1995). Cognition in the wild. MIT Press.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. Educational Psychologist, 48(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Lo Presti, F., Gentile, M., Mangione, G. R. J., & Zampolini, L. (2025). Reimagining AI agents for education: The evolution towards collaborative large language model multi-agent systems. In ICERI2025 Proceedings (pp. 6936–6944). https://doi.org/10.21125/iceri.2025.1903

Luzolo, P. H., Elrawashdeh, Z., Tchappi, I., Galland, S., & Outay, F. (2024). Combining multi-agent systems and artificial intelligence of things: Technical challenges and gains. Internet of Things, 28, 101364. https://doi.org/10.1016/j.iot.2024.101364

McNeese, N. J., Schelble, B. G., Canonico, L. B., & Demir, M. (2021). Who/what is my teammate? Team composition considerations in human–AI teaming. IEEE Transactions on Human-Machine Systems, 51(4), 288–299. https://doi.org/10.1109/THMS.2021.3071180

Morin, E. (1999). Les sept savoirs nécessaires à l'éducation du futur. UNESCO.

Muis, K. R., Psaradellis, C., Lajoie, S., Di Leo, I., & Chevrier, M. (2015). The role of epistemic emotions in mathematics problem solving. Contemporary Educational Psychology, 42, 172–185. https://doi.org/10.1016/j.cedpsych.2015.06.003

Mushtaq, A., Naeem, M. R., Ghaznavi, I., Taj, M. I., Hashmi, I., & Qadir, J. (2025). Harnessing multi-agent LLMs for complex engineering problem-solving: A framework for senior design projects. arXiv. https://doi.org/10.48550/arxiv.2501.01205

OpenAI. (2026). ChatGPT (GPT-5) [Large language model]. OpenAI. https://chat.openai.com/

Park, B., & Seo, K. (2025). Assessing critical thinking through a multi-agent LLM-based debate chatbot. https://doi.org/10.1145/3706599.3721207

Russell, S. J., & Norvig, P. (2003). Artificial intelligence: A modern approach. Prentice Hall.

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. Contemporary Educational Psychology, 60, 101832. https://doi.org/10.1016/j.cedpsych.2019.101832

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), Transfer of learning from a modern multidisciplinary perspective (pp. 1–51). Information Age Publishing.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G. J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as team-mates: A research agenda on AI in team collaboration. Information & Management, 57(2), 103174. https://doi.org/10.1016/j.im.2019.103174

Stone, P., & Veloso, M. (2000). Multiagent systems: A survey from a machine learning perspective. Autonomous Robots, 8(3), 345–383. https://doi.org/10.1023/A:1008942012299

Volet, S., Vauras, M., & Salonen, P. (2009). Self- and social regulation in learning contexts: An integrative perspective. Educational Psychologist, 44(4), 215–226. https://doi.org/10.1080/00461520903213584

Zampolini, L., Gentile, M., & Mangione, G. R. J. (2025). Empowering teachers' creativity: Agentification for design thinking and school innovation. In ICERI2025 Proceedings (pp. 7185–7193). https://doi.org/10.21125/iceri.2025.1970

Zampolini, L., Mangione, G. R. J., & Gentile, M. (2025). Structured roles in design thinking to foster school innovation: A pilot study in small Italian schools. In EDULEARN25 Proceedings (pp. 6491–6500). https://doi.org/10.21125/edulearn.2025.1597

Zhang, G., Chong, L., Kotovsky, K., & Cagan, J. (2023). Trust in an AI versus a human teammate: The effects of teammate identity and performance on human–AI cooperation. Computers in Human Behavior, 139, 107536. https://doi.org/10.1016/j.chb.2022.107536