Bakhtin, M. (2026). Epistemic Trust and Learner Agency in Generative AI-Mediated Education: A Philosophical-Pedagogical Model of Responsible Co-Agency. *The 3rd EIID International Conference "Pedagogy and Psychology of Trust and Learner Agency in the Age of Generative Systems". ESEJ* (pp. 46–66). Ostrava. April, 2026.

DOI: 10.47451/esej-ped-45

The paper is published in Web of Science, Crossref, ICI Copernicus, BASE, Zenodo, OpenAIRE, LORY, Academic Resource Index ResearchBib, J-Gate, ISI International Scientific Indexing, ADL, JournalsPedia, Scilit, EBSCO, Mendeley, and WebArchive databases.

***Maxim Bakhtin***, Doctor of Philosophical Sciences, Professor, Chief Director, International Professors' Business Club. Ragusa, Italy.
E-mail: maximbakhtin@gmail.com

# Epistemic Trust and Learner Agency in Generative AI-Mediated Education: A Philosophical-Pedagogical Model of Responsible Co-Agency

***Abstract:***
This article is devoted to the theoretical analysis of epistemic trust and learner agency in education mediated by generative artificial intelligence. The relevance of the study is determined by the rapid integration of generative systems into educational practice, where they are increasingly used for explanation, feedback, writing support, problem-solving, assessment assistance and the individualisation of learning. The scientific novelty of the study lies in interpreting trust in generative AI not as a purely technical or ethical issue, but as a philosophical-pedagogical condition for preserving learner agency. The object of the study is educational interaction in learning environments mediated by generative AI systems. The subject of the study is the philosophical and pedagogical mechanisms through which epistemic trust and learner agency are formed, transformed or weakened in the use of generative systems. The study aims to develop and theoretically substantiate a philosophical-pedagogical model of responsible co-agency in generative AI-mediated education. The study has a theoretical and analytical research design. Its methodology includes theoretical analysis, conceptual reconstruction, comparative interpretation, systematisation and philosophical-pedagogical modelling. The source base includes works on philosophy of education, epistemic trust, learner agency, AI in education, generative AI, responsible AI, digital pedagogy and human–AI interaction. The article establishes that epistemic trust in generative AI should not be understood as passive confidence in algorithmic output. It must be structured as critical, calibrated and verifiable trust based on comparison, verification, interpretation and human responsibility. The study identifies algorithmic delegation as a key risk of AI-mediated learning: learners may formally complete educational tasks while transferring interpretation, judgement, argumentation or authorship to the system. At the same time, generative AI may support learner agency when it is used as a dialogical and cognitive tool for questioning, reflection, comparison, revision and intellectual exploration. The author proposes the model of responsible co-agency, in which the learner remains the subject of judgement and responsibility, the teacher acts as the organiser of pedagogical and ethical conditions, and the generative system functions as a supportive cognitive tool rather than an autonomous epistemic authority. The study identifies pedagogical conditions necessary for preserving agency: epistemic AI literacy, agency-protective task design, explicit distribution of responsibility, dialogical AI use, teacher-mediated trust calibration and process-oriented assessment. The author concludes that the central challenge of generative AI-mediated education is not technological, but philosophical and pedagogical. Generative systems can become valuable educational tools only when they strengthen learners' capacity for understanding, judgement and responsibility rather than replace intellectual activity with ready-made algorithmic output.

***Keywords:*** epistemic trust, learner agency, generative AI, responsible co-agency, AI-mediated learning, philosophy of education, digital pedagogy, algorithmic delegation, epistemic AI literacy, critical thinking, human responsibility, educational autonomy.

## Introduction

In the context of the rapid development of generative artificial intelligence, education is undergoing a profound transformation that affects not only teaching methods and learning tools, but also the philosophical foundations of the educational process. Generative systems are increasingly involved in explanation, feedback, writing, translation, problem-solving, assessment support, creative modelling and the individualisation of learning trajectories. As a result, the learner no longer interacts only with the teacher, the textbook, the task and their own reasoning, but also with an algorithmic system capable of producing plausible, coherent and context-sensitive responses.

The relevance of the study is determined by the need to rethink the relationship between trust and learner agency in AI-mediated education. In traditional pedagogical contexts, trust is primarily associated with the teacher, the educational institution, the curriculum, scholarly knowledge and the learner's own intellectual effort. In generative AI-mediated learning environments, however, a new epistemic actor appears: a system that can generate explanations and solutions without possessing human understanding, responsibility or pedagogical intention. This creates a new educational situation in which learners may use generative AI as a cognitive and dialogical tool, but may also delegate judgement, interpretation and responsibility to it.

The problem of the study is associated with the ambiguity of epistemic trust in generative systems. On the one hand, trust in AI-generated support may expand learner agency by providing access to explanations, alternative formulations, examples, feedback and intellectual scaffolding. On the other hand, excessive or uncritical trust may weaken agency when learners accept generated outputs as authoritative knowledge, avoid independent reasoning, reduce verification practices or transfer responsibility for the final result to the system. Therefore, the key pedagogical problem is not whether generative AI should be trusted or distrusted, but how epistemic trust should be structured so that it supports rather than replaces learner agency.

The scientific novelty of the study lies in interpreting trust in generative AI not as a purely technical, psychological or ethical issue, but as a philosophical-pedagogical condition for preserving and developing learner agency. The article proposes the concept of responsible co-agency, within which the learner remains the subject of judgement, interpretation, choice and responsibility, while the generative system functions as a cognitive, dialogical and methodological tool. Within this approach, AI-mediated education is understood not as the replacement of human thinking by algorithmic output, but as a structured interaction in which agency must be consciously distributed and pedagogically regulated.

The object of the study is educational interaction in learning environments mediated by generative AI systems.

The subject of the study is the philosophical and pedagogical mechanisms through which epistemic trust and learner agency are formed, transformed or weakened in the use of generative systems.

The study aims to develop and theoretically substantiate a philosophical-pedagogical model of responsible co-agency that explains how epistemic trust in generative AI can support, rather than replace, learner agency.

To achieve this aim, the following research objectives have been defined:

- to analyse the philosophical meaning of epistemic trust in education mediated by generative AI;
- to clarify the concept of learner agency in the context of generative systems;
- to identify the risks of blind algorithmic trust and the delegation of judgement to AI-generated outputs;
- to distinguish between supportive AI-mediated learning and agency-reducing dependence on generative systems;
- to develop a model of responsible co-agency between learner, teacher and generative AI system;
- to substantiate pedagogical conditions for maintaining learner autonomy, critical thinking, verification practices and responsibility in AI-mediated education.

The theoretical significance of the study lies in the development of a philosophical-pedagogical interpretation of trust and agency in the age of generative systems. The article expands the conceptual apparatus of contemporary pedagogy by introducing responsible co-agency as a category that allows AI-mediated learning to be analysed not only in terms of technological efficiency, but also in terms of epistemic responsibility, autonomy, judgement and the preservation of the learner as an active subject of education. This approach contributes to the philosophy of education, digital pedagogy and the emerging field of AI-mediated learning.

The practical significance of the study consists in the possibility of applying its results to the design of educational strategies, teacher training programmes, AI literacy modules, academic integrity policies and learning activities involving generative AI. The proposed model may help educators distinguish between pedagogically productive uses of generative systems and practices that weaken learner autonomy. It may also support the development of tasks that require learners to compare, verify, criticise, revise and justify AI-generated outputs rather than passively consume them.

*Thus*, the study addresses the need for a philosophical and pedagogical framework that can explain how trust should function in generative AI-mediated education. The central argument of the article is that learner agency can be preserved and strengthened only when epistemic trust is combined with critical verification, reflective judgement and human responsibility. In this sense, the future of education in the age of generative systems depends not on replacing learners' intellectual activity with algorithmic assistance, but on forming responsible co-agency between learners, teachers and technological systems.

## Methods

The present study has a theoretical and analytical research design and is aimed at developing a philosophical-pedagogical model of responsible co-agency in generative AI-mediated education. In accordance with the nature of the research problem, the study does not include empirical measurement, experimental intervention or quantitative data collection. Instead, it is based on

theoretical analysis, conceptual reconstruction, comparative interpretation and philosophical-pedagogical modelling.

The choice of this design is determined by the need to analyse epistemic trust and learner agency not only as psychological or technological phenomena, but also as philosophical and pedagogical categories. Generative AI-mediated education changes the structure of educational interaction, since the learner receives explanations, suggestions, feedback and possible solutions from an algorithmic system that does not possess human intentionality, responsibility or understanding. Therefore, the methodological focus of the study is directed towards clarifying the conceptual conditions under which trust in generative systems can support, rather than weaken, learner agency.

The research material consists of scholarly works devoted to philosophy of education, epistemic trust, learner agency, educational autonomy, digital pedagogy, AI in education, generative AI, academic integrity, critical thinking, responsible AI and human–AI interaction. The source base includes theoretical and methodological studies that make it possible to analyse the transformation of learning under the influence of generative systems. Particular attention is paid to works that examine the relationship between trust, knowledge, responsibility, autonomy, human judgement and technological mediation in educational contexts.

The criteria for selecting sources were as follows: relevance to the concepts of epistemic trust, learner agency and AI-mediated learning; significance for philosophy of education and digital pedagogy; applicability to the analysis of generative AI as a learning tool; contribution to the discussion of autonomy, responsibility and verification in educational practice; and representativeness for contemporary interdisciplinary research on artificial intelligence in education. Sources were selected in order to provide a conceptual basis for interpreting generative AI not merely as a technological instrument, but as a factor that transforms the epistemic and pedagogical structure of learning.

The analytical procedure included several consecutive stages. At the first stage, theoretical approaches to trust in education, epistemic authority and learner agency were analysed. At the second stage, the specific features of generative AI-mediated learning were identified, including the production of plausible outputs, personalised explanation, dialogical interaction, feedback simulation and the risk of delegating judgement to algorithmic systems. At the third stage, the risks of blind algorithmic trust and agency-reducing dependence were systematised. At the fourth stage, the conditions under which generative systems can support learner autonomy, critical thinking and reflective judgement were determined. At the final stage, the obtained results were integrated into a philosophical-pedagogical model of responsible co-agency.

The methodology of the study includes general scientific methods and specialised theoretical methods. General scientific methods include analysis, synthesis, comparison, generalisation and classification. These methods were used to identify key categories, compare existing approaches and systematise the pedagogical risks and possibilities of generative AI-mediated education. Specialised methods include conceptual reconstruction, philosophical interpretation, comparative theoretical analysis and pedagogical modelling. Conceptual reconstruction was used to clarify the meaning of epistemic trust, learner agency and responsible co-agency. Philosophical interpretation made it possible to examine the normative and epistemic dimensions of trust, autonomy and responsibility. Comparative theoretical analysis was applied to distinguish between supportive AI-

mediated learning and dependence on AI-generated outputs. Pedagogical modelling was used to construct an integrated model of responsible co-agency.

Within the framework of the study, epistemic trust is understood as the learner's readiness to rely on a source of information, explanation or guidance while preserving the capacity for verification, interpretation and critical judgement. Learner agency is interpreted as the learner's ability to act as a subject of learning: to formulate questions, make decisions, evaluate information, revise understanding, justify conclusions and assume responsibility for the final intellectual result. Responsible co-agency is defined as a pedagogically regulated form of interaction in which the learner, the teacher and the generative system participate in the learning process, but human judgement and responsibility remain central.

The study proceeds from the assumption that generative AI has an ambivalent pedagogical status. On the one hand, it may support learner agency by providing explanations, alternative perspectives, examples, feedback, scaffolding and opportunities for self-directed learning. On the other hand, it may weaken learner agency when students use generated outputs as substitutes for reasoning, interpretation, verification or authorship. Therefore, the central methodological task of the study is to identify the conditions under which generative AI becomes a tool of agency development rather than a mechanism of cognitive delegation.

The proposed model was developed through the identification of several interrelated components of responsible co-agency: epistemic, reflective, critical, dialogical, ethical and pedagogical. The epistemic component concerns the learner's ability to distinguish between information, explanation, probability, argument and knowledge. The reflective component includes awareness of one's own learning goals, difficulties and decisions. The critical component involves verification, comparison of sources and evaluation of AI-generated outputs. The dialogical component reflects the use of generative AI as a partner in questioning, reformulation and exploration rather than as an authority. The ethical component concerns authorship, responsibility and academic integrity. The pedagogical component includes the role of the teacher in designing tasks, rules and learning environments that preserve learner agency.

The validity of the study is ensured by the logical consistency between the research problem, aim, methodological procedure and proposed model. It is also supported by the interdisciplinary comparison of approaches from philosophy of education, digital pedagogy, AI ethics and educational technology studies. The reliability of the study is achieved through the transparent description of the analytical stages, the consistent application of selected methods and the clear differentiation between conceptual analysis, normative interpretation and pedagogical modelling.

The limitations of the study are associated with its theoretical character. The proposed model has not yet been empirically tested in a specific educational environment. Therefore, the results should be regarded as a conceptual framework for further research and practical implementation. Future studies may include empirical investigation of student interaction with generative AI, qualitative interviews with learners and teachers, analysis of AI-supported assignments, comparative studies of learning outcomes, and the development of diagnostic tools for assessing epistemic trust and learner agency in AI-mediated education.

Thus, the chosen methodology corresponds to the aim of the study and makes it possible to analyse the transformation of trust and agency in the age of generative systems. The combination of theoretical analysis, conceptual reconstruction and philosophical-pedagogical modelling

provides a basis for interpreting generative AI-mediated education through the concept of responsible co-agency.

## Literature Review

The problem of epistemic trust and learner agency in generative AI-mediated education is situated at the intersection of several research traditions: philosophy of education, epistemology, psychology of agency, digital pedagogy, artificial intelligence in education, ethics of AI and human–AI interaction. The reviewed literature demonstrates that the integration of generative AI into educational environments cannot be adequately understood only as a technological innovation. It requires a deeper philosophical-pedagogical analysis of how knowledge, trust, autonomy, responsibility and human judgement are transformed when learners interact with systems capable of generating explanations, texts, solutions, arguments and feedback.

A fundamental theoretical basis for analysing learner agency is provided by Bandura's concept of human agency. Bandura (*2006*) argues that agency involves intentionality, forethought, self-reactiveness and self-reflectiveness. From this perspective, learners are not passive recipients of information, but active subjects capable of setting goals, regulating behaviour, evaluating outcomes and reflecting on their actions. This understanding is especially significant in the context of generative AI, because algorithmic systems can either support these dimensions of agency or weaken them. When generative AI is used as a tool for exploration, feedback and reflection, it may enhance the learner's capacity for self-directed learning. However, when it is used as a substitute for judgement, reasoning or authorship, it may reduce the learner's active role in the educational process.

The philosophical dimension of learner agency is further developed in Biesta's theory of education. Biesta (*2010*) criticises the reduction of education to measurable outcomes and emphasises the ethical, political and democratic dimensions of educational practice. His approach is important for the present study because it allows learner agency to be understood not merely as performance, productivity or individual choice, but as subjectification: the formation of the learner as a responsible subject capable of judgement and action. In generative AI-mediated education, this raises a crucial question: does the use of AI contribute to the learner's development as a subject, or does it transform learning into the efficient production of outputs? This question is central to the philosophical-pedagogical analysis of responsible co-agency.

The epistemological problem of trust is connected with the fact that learning always involves dependence on others. Hardwig (*1985*) demonstrates that epistemic dependence is an unavoidable feature of knowledge practices: individuals often rely on the testimony, expertise and judgement of others. This idea is directly relevant to AI-mediated education, since learners may increasingly rely on generative systems for explanations, information and intellectual support. However, epistemic dependence on AI differs from traditional dependence on teachers or experts. A generative system may produce plausible responses without possessing understanding, accountability or pedagogical intention. Therefore, epistemic trust in AI requires specific forms of verification, critical evaluation and pedagogical regulation.

The ethical dimension of epistemic trust is further clarified by Fricker's concept of epistemic injustice. Fricker (*2007*) shows that knowledge practices are shaped by power, credibility and the unequal distribution of epistemic authority. In the context of generative AI, this raises the problem

of algorithmic epistemic authority. Learners may attribute excessive credibility to AI-generated outputs because they appear fluent, structured and confident. At the same time, generative systems may reproduce biases, inaccuracies or culturally limited perspectives. Thus, trust in AI is not neutral. It must be analysed as part of the broader epistemic structure of education, where authority, credibility, interpretation and responsibility are distributed among learners, teachers, institutions and technological systems.

The broader philosophical context of digital transformation is provided by Floridi's analysis of the infosphere. Floridi (*2017*) argues that contemporary human life is increasingly shaped by informational environments in which the boundaries between online and offline, human and artificial, production and mediation of knowledge become more complex. This approach is important for understanding generative AI-mediated education because learning increasingly takes place within an infosphere where algorithmic systems participate in the creation, organisation and circulation of knowledge. In such conditions, education must develop not only digital skills, but also epistemic orientation: the ability to understand how information is generated, mediated, evaluated and responsibly used.

The ethical and anthropological implications of artificial intelligence are also central to Benanti's approach. Benanti (*2022*) emphasises the need to preserve human decision-making in interaction with artificial intelligence and develops the idea of the human-in-the-loop. This position is directly relevant to the concept of responsible co-agency proposed in the present study. In education, the learner and teacher must remain in the loop not only technically, but also epistemically and morally. Generative AI may support learning, but it should not become an autonomous authority that replaces human judgement, responsibility or pedagogical intentionality. Therefore, responsible co-agency requires a clear distinction between assistance and delegation.

A pedagogical-technological perspective is represented by Rivoltella and Rossi (*2019*), who analyse technologies for education as components of learning environments rather than as neutral instruments. This approach is important because it prevents the reduction of AI to a simple tool. Educational technologies shape practices, interactions, roles and expectations. In generative AI-mediated education, this means that the introduction of AI changes the structure of learning tasks, the role of the teacher, the learner's strategies, the forms of feedback and the criteria of academic integrity. Consequently, generative AI should be pedagogically designed into the educational process, rather than informally added as an uncontrolled external assistant.

The field of artificial intelligence in education provides an important empirical and conceptual background for the present study. Zawacki-Richter et al. (*2019*), in their systematic review of AI applications in higher education, show that much research has been technologically driven and that the role of educators has often been underrepresented. This finding is significant because it confirms a gap between technological development and pedagogical reflection. AI in education cannot be analysed only through automation, prediction or efficiency. It must also be examined in relation to teaching, learning, agency, responsibility and the educational purposes that guide the use of technology.

Ouyang and Jiao (*2021*) identify three paradigms of artificial intelligence in education: AI-directed, learner-as-recipient; AI-supported, learner-as-collaborator; and AI-empowered, learner-as-leader. This framework is especially useful for the present study because it provides a conceptual basis for distinguishing between different degrees of learner agency. In the first paradigm, AI may

dominate the educational process and reduce the learner to a passive recipient of algorithmically organised content. In the second, AI supports learning through collaboration and feedback. In the third, AI becomes a means of empowering learners to set goals, make decisions and take responsibility for learning. The model of responsible co-agency proposed in the present study is closest to the learner-as-leader paradigm, but it adds a philosophical emphasis on epistemic trust, judgement and responsibility.

Recent research on generative AI in education further clarifies the opportunities and risks of large language models. Kasneci et al. (*2023*) analyse the potential of ChatGPT and similar systems for education, highlighting opportunities for personalised learning, feedback, language support and accessibility, while also identifying challenges related to accuracy, bias, overreliance, assessment and academic integrity. This balanced approach is important because it avoids both technological optimism and technological rejection. Generative AI may be pedagogically valuable, but only if learners are trained to interact with it critically and responsibly. This supports the central thesis of the present study: the key issue is not the presence of AI in education, but the structure of trust and agency within AI-mediated learning.

UNESCO (*2023*) provides an institutional and normative framework for the use of generative AI in education and research. The guidance emphasises the need for human-centred approaches, regulation, inclusion, transparency, teacher support, protection of learners and the development of AI literacy. For the present study, UNESCO's position is significant because it confirms that generative AI must be integrated into education under conditions of human responsibility and pedagogical governance. AI literacy should include not only technical knowledge of how systems work, but also the ability to verify outputs, understand limitations, recognise risks and maintain human agency.

The most directly relevant contemporary review is provided by Roe and Perkins (*2024*), who examine generative AI and agency in education. Their critical scoping review demonstrates that generative AI may both support and undermine agency, depending on how it is used, how tasks are designed and whether learners remain active participants in the learning process. This source is particularly important for the present article because it confirms the need to analyse agency not as a fixed characteristic of the learner, but as a dynamic relation shaped by educational design, technological affordances and institutional expectations. The present study develops this line by introducing epistemic trust as a key mediating category between generative AI and learner agency.

The comparative analysis of the reviewed sources shows that each research tradition explains an important aspect of the problem. Bandura (*2006*) provides a psychological understanding of agency as intentional and reflective self-regulation. Biesta (*2010*) expands this understanding through the philosophy of education and the formation of the learner as a subject. Hardwig (*1985*) and Fricker (*2007*) reveal the epistemological and ethical dimensions of trust, dependence and credibility. Floridi (*2017*) and Benanti (*2022*) provide a philosophical and ethical framework for understanding human action in technologically mediated informational environments. Rivoltella and Rossi (*2019*) interpret educational technologies as structuring elements of learning environments. Zawacki-Richter et al. (*2019*), Ouyang and Jiao (*2021*), Kasneci et al. (*2023*), UNESCO (*2023*), and Roe and Perkins (*2024*) show that AI and generative AI create both new opportunities and significant risks for education.

At the same time, the literature reveals a clear research gap. Existing studies analyse learner agency, epistemic trust, AI ethics, digital pedagogy and generative AI in education, but these areas are often examined separately. Research on AI in education frequently focuses on applications, benefits, risks and technological implementation. Philosophical discussions of trust and agency often remain insufficiently connected to the concrete pedagogical conditions of generative AI-mediated learning. Conversely, practical discussions of generative AI in education do not always provide a sufficiently developed philosophical model of how trust should be structured so that learner agency is preserved.

This gap determines the need for the present study. A philosophical-pedagogical model is required that can explain how epistemic trust, learner agency and generative AI interaction should be organised within education. The concept of responsible co-agency proposed in this article addresses this need by interpreting AI-mediated learning as a structured interaction between learner, teacher and generative system. In this model, the learner remains the subject of judgement, interpretation and responsibility; the teacher organises the pedagogical conditions for critical and reflective AI use; and the generative system functions as a cognitive and dialogical tool rather than as an autonomous epistemic authority.

*Thus*, the reviewed literature confirms the relevance of analysing epistemic trust and learner agency in generative AI-mediated education. It also shows that the educational value of generative AI depends not on the mere availability of algorithmic assistance, but on the formation of pedagogical conditions under which learners can use AI critically, reflectively and responsibly. This conclusion provides the theoretical basis for developing a model of responsible co-agency in the age of generative systems.

## Results

*1. Epistemic Trust as a Regulated Pedagogical Condition rather than Passive Reliance on Generative AI*

The study established that epistemic trust in generative AI-mediated education should not be understood as the learner's simple confidence in the correctness of AI-generated outputs. In pedagogical terms, such an interpretation is insufficient and potentially dangerous, since generative systems are capable of producing fluent, coherent and persuasive responses without possessing human understanding, responsibility or intentionality. Therefore, trust in generative AI must be interpreted as a regulated pedagogical condition that enables the learner to use AI-generated content critically, reflectively and selectively.

The analysis showed that epistemic trust in education traditionally emerges from the relationship between the learner, the teacher, the curriculum, scholarly knowledge and institutional authority. In the case of generative AI, this structure becomes more complex, because an algorithmic system enters the learning process as an additional source of explanation, suggestion and apparent expertise. However, unlike a teacher or expert, generative AI cannot be treated as a subject of pedagogical responsibility. It does not understand the learner's educational development in the human sense and cannot assume responsibility for the learner's final judgement. This creates a new epistemic asymmetry: the system may appear authoritative, while the basis of its authority remains opaque to the learner.

The study identified three forms of epistemic trust in generative AI-mediated learning. The first form is instrumental trust, in which the learner uses AI as a tool for clarification, reformulation,

summarisation, comparison or idea generation. This form is pedagogically productive because it preserves the learner's active role. The second form is interpretative trust, in which the learner treats AI-generated content as a possible perspective that requires evaluation, contextualisation and comparison with other sources. This form may support deeper learning if the learner remains critically engaged. The third form is delegative trust, in which the learner accepts AI-generated output as a ready-made solution and transfers judgement, authorship and responsibility to the system. This form is pedagogically destructive because it reduces the learner's agency.

The results indicate that the key pedagogical problem is not the presence of trust itself, but the absence of its regulation. Complete distrust of generative AI would make it impossible to use its educational potential, while blind trust would undermine the learner's autonomy and critical capacity. Therefore, the study proposes the concept of calibrated epistemic trust. This means that the learner may rely on generative AI as a source of support, but only under the conditions of verification, comparison, interpretation and personal responsibility. Trust becomes educationally valid only when it is combined with critical distance.

This interpretation develops Hardwig's idea of epistemic dependence by showing that dependence on a source of knowledge is not necessarily negative if it is accompanied by awareness of limits, criteria of credibility and responsibility for judgement (*Hardwig, 1985*). However, in the case of generative AI, such dependence requires additional safeguards, because AI-generated outputs may imitate expertise without providing transparent grounds for knowledge. Fricker's concept of epistemic injustice also becomes relevant here, since learners may overestimate or underestimate sources of knowledge depending on credibility structures, power relations and perceived authority (*Fricker, 2007*). Generative AI may become an artificially inflated epistemic authority if learners are not trained to question its outputs.

*Thus*, the first result of the study is the substantiation of epistemic trust as a pedagogically regulated condition of AI-mediated learning. Trust in generative AI should not be passive reliance on algorithmic fluency, but a calibrated relation in which the learner uses AI support while preserving verification, critical judgement and intellectual responsibility.

This, epistemic trust in generative AI becomes pedagogically productive only when it is structured as critical, calibrated and verifiable trust. It should support the learner's thinking rather than replace it.

*2. Learner Agency under the Risk of Algorithmic Delegation*

The second result of the study concerns the transformation of learner agency in generative AI-mediated education. The analysis showed that generative AI does not simply add a new tool to the learning process; it changes the distribution of cognitive actions between the learner and the technological system. A learner may use AI to generate ideas, structure arguments, explain concepts, correct language, solve problems, prepare drafts or evaluate alternatives. Each of these actions may either strengthen or weaken agency depending on how the learner positions themselves in relation to the generated output.

Learner agency may be understood as the capacity of the learner to formulate goals, ask questions, make choices, evaluate information, regulate learning strategies, justify conclusions and take responsibility for the final intellectual result. This interpretation corresponds to Bandura's understanding of human agency as intentional, self-regulative and reflective action (*Bandura, 2006*).

In the context of generative AI, however, the learner's agency becomes unstable because many actions that previously required independent reasoning can now be outsourced to the system.

The study identified the phenomenon of algorithmic delegation as one of the central risks of generative AI-mediated education. Algorithmic delegation occurs when learners do not merely use AI for support, but transfer to it key functions of learning: problem formulation, interpretation, argument construction, verification, evaluation and final decision-making. In such cases, the learner may formally complete an educational task while internally withdrawing from the process of intellectual formation. The result is not genuine learning, but the production of an educational artefact with reduced personal cognitive participation.

This risk is especially important because generative AI often produces outputs that look complete. A coherent answer may create the illusion that the problem has been understood, even when the learner has not performed the necessary cognitive work. In this sense, generative AI may produce what can be called the appearance of competence. The learner receives a well-structured text, explanation or solution, but the internal structure of understanding remains underdeveloped. This is one of the most serious pedagogical risks of generative systems: they can separate the external product of learning from the internal process of learning.

At the same time, the study showed that generative AI may also strengthen learner agency if it is used as a tool of inquiry rather than as a substitute for thinking. AI can help learners compare explanations, identify gaps, generate counterarguments, test interpretations, reformulate unclear ideas and receive immediate feedback. In such cases, the learner remains the subject of the process, while the system functions as a dialogical and cognitive scaffold. This distinction corresponds to Ouyang and Jiao's differentiation between AI-directed, AI-supported and AI-empowered learning paradigms (*Ouyang & Jiao, 2021*). The pedagogically desirable model is not AI-directed learning, where the learner becomes a recipient, but AI-empowered learning, where technology expands the learner's capacity to act.

The study therefore proposes a distinction between agency-supporting AI use and agency-reducing AI use. Agency-supporting use involves questioning, verification, revision, comparison and reflective appropriation of AI-generated content. Agency-reducing use involves copying, passive acceptance, substitution of judgement and avoidance of cognitive effort. The same technological system may support both types of use; the difference lies in pedagogical design, task structure and the learner's epistemic habits.

This result also develops Biesta's philosophical understanding of education as subjectification, that is, the formation of the learner as a responsible subject rather than merely a performer of measurable tasks (*Biesta, 2010*). If generative AI is used only to optimise output production, education risks losing its subject-forming dimension. If, however, AI is used to provoke questioning, reflection and judgement, it may support the learner's becoming as an autonomous educational subject.

Thus, the second result of the study is the identification of algorithmic delegation as a key threat to learner agency and the substantiation of the conditions under which generative AI can support agency. The learner must remain the author of questions, the evaluator of answers and the responsible subject of the final result.

*Thus*, generative AI strengthens learner agency only when it expands the learner's capacity for questioning, interpretation and judgement. It weakens agency when it replaces the learner's cognitive participation with ready-made algorithmic output.

*3. Responsible Co-Agency as a Philosophical-Pedagogical Model of AI-Mediated Learning*

The third result of the study is the development of the concept of responsible co-agency as a philosophical-pedagogical model for generative AI-mediated education. The model is based on the idea that learning with generative AI should not be interpreted either as autonomous human learning with an external tool or as human dependence on an intelligent system. Instead, it should be understood as a structured interaction among three elements: the learner, the teacher and the generative system.

In this model, the learner remains the central subject of learning. The learner formulates questions, interprets information, evaluates AI-generated outputs, integrates knowledge and assumes responsibility for the final intellectual product. The teacher performs the role of pedagogical organiser, epistemic guide and ethical regulator. The teacher designs tasks, defines acceptable uses of AI, teaches verification strategies, supports reflection and protects the learner from uncritical dependence. The generative system functions as a cognitive and dialogical tool that can provide explanations, examples, alternatives and feedback, but cannot replace human judgement or responsibility.

The model of responsible co-agency differs from ordinary AI-assisted learning because it focuses not on assistance as such, but on the structure of responsibility within assistance. A learning process may be technologically advanced but pedagogically weak if responsibility is displaced from the learner to the system. Conversely, a learning process may be genuinely educational when AI participation is organised in such a way that the learner's judgement becomes more active, not less active.

The study identified six components of responsible co-agency. The epistemic component concerns the learner's ability to distinguish between information, explanation, interpretation, probability and knowledge. The reflective component concerns the learner's awareness of how and why AI is being used. The critical component includes verification, source comparison, error detection and argument evaluation. The dialogical component involves using AI for questioning, reformulation and exploration rather than final authority. The ethical component concerns authorship, academic integrity, responsibility and transparency of AI use. The pedagogical component concerns the teacher's role in designing learning environments that preserve human agency.

This model is consistent with Benanti's human-in-the-loop approach, which emphasises the need to preserve human decision-making in interaction with artificial intelligence (*Benanti, 2022*). However, the present study extends this idea pedagogically. In education, being "in the loop" should not mean merely approving or rejecting AI-generated outputs. It should mean remaining intellectually, ethically and reflectively present in the learning process. A learner is truly in the loop only when they understand the task, evaluate the generated response, revise it critically and can justify the final result.

The model also corresponds to UNESCO's human-centred approach to generative AI in education and research, which emphasises transparency, inclusion, teacher support, AI literacy and

the protection of learners (*UNESCO, 2023*). However, the present study adds a philosophical-pedagogical dimension: the central value to be protected is not only safety or fairness, but learner agency as the capacity for responsible judgement.

The study proposes that responsible co-agency should be operationalised through specific pedagogical practices. These may include requiring learners to explain how they used AI, compare AI outputs with scholarly sources, identify weaknesses in generated answers, revise prompts, justify final decisions and reflect on what they understood independently. Such tasks transform AI use from hidden substitution into visible learning activity. They also allow the teacher to evaluate not only the final product, but the learner's process of judgement.

Thus, the third result of the study is the substantiation of responsible co-agency as a model in which generative AI participates in learning without becoming an autonomous epistemic authority. The model preserves the learner's subject position while recognising the real cognitive potential of generative systems.

*Thus*, responsible co-agency means that AI may participate in the learning process, but human judgement remains central. The learner acts, the teacher regulates, and the generative system supports rather than replaces educational agency.

*4. Pedagogical Conditions for Preserving Trust and Agency in the Age of Generative Systems*

The fourth result of the study is the identification of pedagogical conditions necessary for preserving epistemic trust and learner agency in generative AI-mediated education. The analysis showed that the productive use of generative systems does not emerge automatically. It requires purposeful pedagogical design, explicit rules, reflective practices and assessment models that value reasoning rather than only final outputs.

The first condition is AI literacy understood not as technical familiarity with tools, but as epistemic literacy. Learners must understand that generative AI produces probabilistic outputs, not guaranteed knowledge. They should know that AI-generated responses may contain errors, omissions, biases, invented references and unsupported claims. Therefore, AI literacy should include the ability to ask: What kind of answer is this? What evidence supports it? What may be missing? What should be checked? How does this output relate to the task and to my own reasoning?

The second condition is task design that makes delegation unproductive. If a task can be fully completed by copying an AI-generated answer, then the task itself no longer protects learner agency. In the age of generative systems, educational tasks should require comparison, justification, contextualisation, personal reasoning, process documentation and reflective evaluation. The learner should not only present a final answer, but demonstrate the intellectual path through which the answer was examined and transformed.

The third condition is the explicit distribution of responsibility. Learners should understand that AI may assist in producing intermediate materials, but responsibility for the final content remains human. This includes responsibility for accuracy, argumentation, citation, interpretation, ethical use and academic integrity. The educational value of AI use depends on whether the learner can explain and defend the final result as their own intellectual position.

The fourth condition is dialogical use of AI. Generative systems should be used to stimulate thinking through questions, alternatives and counterarguments rather than to provide final

conclusions. For example, learners may ask AI to generate opposing views, identify weaknesses in an argument, suggest further questions or explain a concept at different levels of complexity. Such use supports learner agency because it expands the field of reflection rather than closing it.

The fifth condition is teacher-mediated trust calibration. Teachers should help learners distinguish between situations in which AI can be useful and situations in which reliance on AI is risky. For example, AI may be useful for brainstorming, reformulation or preliminary explanation, but risky for specialised factual claims, source attribution, ethical judgement or final assessment. Trust calibration should therefore become part of pedagogical guidance.

The sixth condition is assessment of process rather than only product. In AI-mediated education, assessment should include evidence of the learner's reasoning, verification and revision. Possible assessment formats include reflective commentaries, AI-use logs, comparison tables, annotated drafts, oral defence, source verification tasks and critical analysis of AI-generated responses. Such formats make the learner's agency visible.

The study also identified a broader institutional implication. Educational institutions should not respond to generative AI only through prohibition or uncontrolled acceptance. Prohibition may ignore the reality of technological change, while uncontrolled acceptance may normalise hidden delegation. A more productive strategy is the creation of pedagogical norms that distinguish acceptable, questionable and unacceptable AI use according to the educational purpose of the task.

This result is consistent with contemporary research showing that generative AI creates both opportunities and risks for learning (*Kasneci et al., 2023*; *Roe & Perkins, 2024*). However, the present study adds that the central pedagogical criterion should be agency preservation. AI use is educationally justified when it increases the learner's capacity for understanding, judgement and responsibility. It is pedagogically problematic when it reduces the learner to a manager of generated outputs.

Thus, the fourth result of the study is the identification of pedagogical conditions that transform generative AI from a potential source of dependency into a tool of responsible learning. These conditions include epistemic AI literacy, agency-protective task design, explicit responsibility, dialogical AI use, teacher-mediated trust calibration and process-oriented assessment.

*Thus*, learner agency in the age of generative systems is preserved not by rejecting AI, but by designing educational conditions in which AI use requires verification, reflection, justification and human responsibility. The pedagogical task is to teach learners not merely to use generative systems, but to remain subjects of learning while using them.

## Discussion

The results of the study demonstrate that the integration of generative AI into education should be interpreted not only as a technological innovation, but also as a transformation of the epistemic and pedagogical structure of learning. The central issue is no longer limited to whether generative systems can provide useful explanations, examples, feedback or support. A deeper question concerns how learners relate to AI-generated outputs, how they distribute trust and responsibility, and whether their agency is strengthened or weakened in the process. In this respect, the study confirms that generative AI-mediated education requires a philosophical-pedagogical

framework capable of explaining the relationship between epistemic trust, learner agency and responsibility.

The first major result of the study concerns the interpretation of epistemic trust as a regulated pedagogical condition. This finding is significant because trust in generative AI is often discussed either in technical terms, such as reliability and accuracy, or in ethical terms, such as transparency and safety. The present study suggests that, in education, trust must also be understood pedagogically. A learner does not simply use information; they learn through the process of selecting, interpreting, questioning and justifying information. Therefore, epistemic trust in AI becomes educationally meaningful only when it is connected with verification, comparison and reflective judgement.

This result develops Hardwig's concept of epistemic dependence, according to which reliance on others is an unavoidable part of knowledge practices (*Hardwig, 1985*). However, dependence on generative AI differs from dependence on teachers, scholars or experts. A human expert can be questioned as a responsible subject, whereas a generative system produces outputs without human understanding, intentionality or accountability. Therefore, the study shows that epistemic dependence on AI requires a special form of calibrated trust. Learners may rely on AI as a source of support, but they must not treat its outputs as self-sufficient knowledge. This distinction is crucial for preventing algorithmic authority from replacing educational judgement.

The findings also correspond to Fricker's analysis of epistemic injustice, where credibility, authority and knowledge are connected with power relations (*Fricker, 2007*). In generative AI-mediated learning, a new form of epistemic imbalance may emerge: the learner may attribute excessive credibility to algorithmic output because it appears fluent, structured and confident. This creates the risk of what may be called artificial epistemic authority. The system looks like a knowledgeable interlocutor, but its apparent authority is not equivalent to understanding. Consequently, pedagogical practice must teach learners to distinguish between fluency, plausibility, evidence and truth.

The second major result concerns learner agency under the risk of algorithmic delegation. The study shows that generative AI can both support and weaken agency depending on how it is used. This finding is consistent with Bandura's understanding of agency as intentional, self-reflective and self-regulative action (*Bandura, 2006*). If learners use AI to ask better questions, compare perspectives, identify gaps and revise their reasoning, generative systems may strengthen self-regulated learning. However, if learners use AI to replace interpretation, argumentation, verification or authorship, their agency is weakened.

The concept of algorithmic delegation is one of the key theoretical contributions of the study. It explains a specific risk of generative AI-mediated education: the learner may formally complete a task while internally withdrawing from the cognitive process. In such cases, the educational product remains visible, but the educational experience becomes hollow. The learner submits a text, solution or project, but the intellectual work that should have formed understanding, judgement and competence has been outsourced to the system. This distinction is especially important because traditional assessment often focuses on the final product, while generative AI makes it necessary to assess the process of thinking.

This result also develops Biesta's understanding of education as subjectification (*Biesta, 2010*). If education is understood merely as the production of measurable outputs, generative AI appears

highly efficient. It can produce essays, answers, plans and explanations quickly. However, if education is understood as the formation of the learner as a responsible subject, then the question changes. The issue is not whether AI helps to produce an answer, but whether the learner becomes more capable of judgement, interpretation and responsibility through the process. The study therefore confirms that the philosophical purpose of education cannot be reduced to output optimisation.

The third major result is the model of responsible co-agency. This model offers a way to avoid two extremes: technological rejection and uncritical technological acceptance. On the one hand, generative AI cannot simply be prohibited or ignored, since it has already become part of contemporary knowledge practices. On the other hand, it cannot be allowed to function as an uncontrolled substitute for learner reasoning. Responsible co-agency proposes a structured relationship among learner, teacher and generative system. The learner remains the subject of learning; the teacher organises and regulates the pedagogical environment; the generative system functions as a cognitive and dialogical tool.

This model is close to Benanti's human-in-the-loop approach, which emphasises the need to preserve human decision-making in interaction with artificial intelligence (*Benanti, 2022*). However, the present study extends this idea in a specifically pedagogical direction. In education, being "in the loop" should not mean merely approving AI-generated content. It should mean remaining cognitively, ethically and reflectively involved in the learning process. A learner is in the loop only when they understand the task, evaluate the generated output, compare it with other sources, revise it critically and can justify the final result.

The model also corresponds to UNESCO's human-centred approach to generative AI in education and research, which emphasises regulation, AI literacy, transparency, inclusion and protection of learners (*UNESCO, 2023*). However, the present study adds that the central pedagogical value to be protected is learner agency. The danger of generative AI is not only misinformation, bias or academic dishonesty, but also the gradual weakening of the learner's capacity to act as an author of thought. Therefore, AI literacy should be understood not only as knowledge about AI tools, but also as epistemic literacy: the ability to evaluate, question and responsibly use algorithmic outputs.

The fourth major result concerns pedagogical conditions for preserving trust and agency. The study identified several conditions: epistemic AI literacy, agency-protective task design, explicit distribution of responsibility, dialogical AI use, teacher-mediated trust calibration and process-oriented assessment. These conditions are important because they transform the discussion from abstract principles to educational design. Generative AI becomes pedagogically productive only when learning tasks require learners to think with AI rather than allow them to hide behind AI.

This conclusion is consistent with Ouyang and Jiao's distinction between AI-directed, AI-supported and AI-empowered learning paradigms (*Ouyang & Jiao, 2021*). The model proposed in the present study rejects AI-directed learning, in which the learner becomes a passive recipient of algorithmically structured content. It also goes beyond simple AI-supported learning, where technology provides assistance but the structure of responsibility remains underdeveloped. Responsible co-agency corresponds most closely to AI-empowered learning, but adds a stronger philosophical emphasis on trust, judgement, authorship and ethical responsibility.

The results also support the concerns raised by Zawacki-Richter et al. (*2019*), who note that research on AI in higher education has often been technologically driven and has insufficiently addressed the role of educators. The study confirms that teachers remain central in generative AI-mediated education. Their role is not diminished, but transformed. The teacher becomes not only a source of knowledge, but also a designer of epistemic conditions: they define how AI may be used, how outputs should be verified, how responsibility is distributed and how the learner's agency is made visible.

The findings are also consistent with Kasneci et al. (*2023*), who show that large language models create both opportunities and challenges for education, including personalised support, feedback, accessibility, bias, overreliance and academic integrity issues. The present study develops this balanced view by proposing that the decisive criterion is not whether AI is used, but whether its use contributes to learner agency. A practice is pedagogically justified when AI helps the learner understand more deeply, ask better questions, revise reasoning and assume responsibility. It is problematic when AI becomes a substitute for the learner's intellectual participation.

The theoretical contribution of the study consists in the development of the concept of responsible co-agency as a philosophical-pedagogical model of generative AI-mediated education. Existing research discusses AI literacy, academic integrity, learner agency, trust and AI ethics, but these issues are often analysed separately. The proposed model integrates them into one conceptual framework. It shows that epistemic trust, learner agency and AI use are not independent variables, but interconnected dimensions of educational interaction.

The study also contributes to the philosophy of education by clarifying the normative status of the learner in the age of generative systems. The learner should not be understood merely as a user of AI tools or as a producer of assessable outputs. The learner remains a subject of judgement, interpretation and responsibility. This position is especially important because generative systems can produce the appearance of competence. The article therefore argues that education must protect the difference between having an answer and understanding an answer.

The practical significance of the study lies in the possibility of applying the proposed model to educational design. First, the model may be used in the development of AI literacy courses. Such courses should include not only prompt engineering or technical familiarisation with tools, but also source verification, recognition of hallucinations, bias detection, ethical authorship and reflection on cognitive delegation. Secondly, the model may inform academic integrity policies by distinguishing between acceptable AI-supported learning and unacceptable substitution of learner work.

Thirdly, the proposed model may be used in teacher training. Teachers need methodological tools for designing assignments that make passive AI use ineffective and reflective AI use productive. For example, tasks may require students to compare AI-generated explanations, identify errors, explain revisions, document their use of AI, defend their final conclusions orally or submit reflective commentaries. These practices make learner agency visible and assessable.

Fourthly, the model has implications for assessment. In generative AI-mediated education, assessment should focus not only on the final product, but also on the process of reasoning. Process-oriented assessment may include AI-use logs, annotated drafts, comparison tables, verification reports, reflective essays and oral defence. Such formats allow educators to evaluate whether the learner has engaged critically with AI-generated material or merely delegated the task.

Despite its theoretical and practical value, the study has several limitations. The first limitation is its theoretical character. The model of responsible co-agency has been developed through conceptual analysis and philosophical-pedagogical modelling, but it has not yet been empirically tested in real educational environments. Therefore, future research should examine how the model functions in different educational contexts and disciplines.

The second limitation concerns the rapidly changing nature of generative AI. The capabilities, limitations and forms of interaction with generative systems are developing quickly. As a result, any theoretical model must remain flexible and open to revision. Future systems may have greater multimodal capacity, stronger personalisation and deeper integration into learning platforms, which may create new forms of agency support and new forms of dependency.

The third limitation concerns contextual variability. The relationship between trust, agency and AI may differ across educational levels, disciplines, cultures and institutional policies. For example, the use of generative AI in philosophy, engineering, medicine, language learning or creative writing may produce different risks and opportunities. Therefore, the proposed model should be adapted to concrete pedagogical contexts rather than applied mechanically.

Future research should develop in several directions. First, empirical studies are needed to investigate how students understand and practise epistemic trust in generative AI. Such studies may include interviews, classroom observations, analysis of student prompts, AI-use diaries and comparison of learning outcomes. Secondly, researchers should examine how different task designs influence learner agency. It would be useful to compare assignments that allow easy delegation with assignments that require verification, reflection and justification.

Thirdly, future research should develop diagnostic criteria for responsible co-agency. These criteria may include the learner's ability to formulate questions, evaluate AI-generated outputs, identify errors, compare sources, revise reasoning, disclose AI use and justify final conclusions. Fourthly, further studies should investigate the role of teachers as mediators of epistemic trust. Teacher training programmes should be analysed in terms of their capacity to prepare educators for regulating AI-mediated learning environments.

Finally, future research should examine the ethical and institutional implications of responsible co-agency. Universities and schools need policies that do not reduce the problem to prohibition or permission. Instead, they should define different levels of acceptable AI use according to the educational aim of each task. The same tool may be appropriate in one learning context and inappropriate in another. Therefore, educational governance of generative AI must be pedagogically differentiated.

*Overall*, the discussion confirms that the central challenge of generative AI-mediated education is not technological, but philosophical and pedagogical. Generative systems can support learning only when trust is calibrated, agency is preserved and responsibility remains human. The proposed model of responsible co-agency makes it possible to interpret AI-mediated education as a structured interaction in which the learner does not surrender judgement to the system, but uses the system to deepen understanding, strengthen autonomy and develop responsible intellectual action.

## Conclusion

The study conducted made it possible to establish that epistemic trust and learner agency are central philosophical and pedagogical categories for understanding education in the age of generative systems. Generative AI does not merely provide new technical instruments for explanation, writing, feedback or problem-solving. It changes the structure of educational interaction by introducing an algorithmic system into the relationship between learner, teacher, knowledge and responsibility. Therefore, the use of generative AI in education requires not only methodological regulation, but also a deeper philosophical understanding of how trust, judgement and agency should be organised.

The aim of the study, which consisted in developing and theoretically substantiating a philosophical-pedagogical model of responsible co-agency in generative AI-mediated education, was achieved. The analysis showed that learner agency can be preserved and strengthened only when epistemic trust in generative AI is structured as critical, calibrated and verifiable trust. Trust becomes pedagogically productive not when learners passively rely on AI-generated outputs, but when they use such outputs as objects of interpretation, comparison, verification and revision.

The research objectives were consistently fulfilled. The philosophical meaning of epistemic trust in AI-mediated education was analysed. The concept of learner agency was clarified in the context of generative systems. The risks of blind algorithmic trust and delegation of judgement to AI were identified. The distinction between agency-supporting and agency-reducing AI use was substantiated. A philosophical-pedagogical model of responsible co-agency between learner, teacher and generative system was developed. Pedagogical conditions for maintaining learner autonomy, critical thinking, verification practices and human responsibility were also determined.

The first major result of the study is the interpretation of epistemic trust as a regulated pedagogical condition. The study demonstrated that trust in generative AI should not be equated with confidence in the correctness of algorithmic output. Generative systems may produce fluent and persuasive responses without possessing human understanding, responsibility or pedagogical intention. Therefore, epistemic trust must be calibrated through verification, comparison of sources, critical judgement and awareness of the limitations of AI-generated content. In this sense, trust becomes not a passive attitude, but an active educational practice.

The second major result is the identification of algorithmic delegation as a key risk for learner agency. The study showed that learners may formally complete educational tasks while internally transferring problem formulation, interpretation, argumentation, verification or authorship to the generative system. This creates the appearance of competence without necessarily forming understanding. At the same time, generative AI may support learner agency when it is used as a tool for questioning, reflection, comparison, revision and intellectual exploration. Thus, the pedagogical value of AI depends not on the system itself, but on the structure of learner participation.

The third result is the development of the concept of responsible co-agency. Within this model, the learner remains the subject of judgement, interpretation, choice and responsibility; the teacher acts as the organiser of pedagogical, epistemic and ethical conditions; and the generative system functions as a cognitive and dialogical tool. Responsible co-agency does not imply equal responsibility between human and machine. Rather, it describes a regulated educational interaction in which AI may participate in the learning process, but human judgement remains central.

The fourth result is the identification of pedagogical conditions necessary for preserving trust and agency in the age of generative systems. These conditions include epistemic AI literacy, agency-protective task design, explicit distribution of responsibility, dialogical use of AI, teacher-mediated trust calibration and process-oriented assessment. The study showed that education should not respond to generative AI only through prohibition or uncontrolled acceptance. A more productive approach is the design of learning environments in which AI use requires explanation, verification, reflection, justification and transparent responsibility.

The theoretical significance of the study lies in the development of a philosophical-pedagogical interpretation of generative AI-mediated education. The article expands the conceptual apparatus of digital pedagogy by introducing responsible co-agency as a category that integrates epistemic trust, learner agency, teacher mediation and human responsibility. This makes it possible to analyse generative AI not merely as an educational technology, but as a factor transforming the epistemic structure of learning.

The practical significance of the study consists in the possibility of applying the proposed model in educational design, teacher training, AI literacy programmes, academic integrity policies and assessment practices. The model may help educators distinguish between pedagogically productive AI use and forms of algorithmic delegation that weaken learner agency. It may also serve as a basis for designing assignments that require learners to verify, criticise, revise and justify AI-generated outputs rather than passively reproduce them.

At the same time, the study has several limitations. Its results are theoretical and require empirical verification in real educational environments. Future research should examine how students and teachers understand epistemic trust in generative AI, how different task designs affect learner agency, and how responsible co-agency can be assessed in practice. It is also necessary to develop diagnostic criteria for identifying agency-supporting and agency-reducing forms of AI use across different educational levels and disciplines.

*In conclusion*, the article demonstrates that the main challenge of generative AI-mediated education is not technological, but philosophical and pedagogical. Generative systems may become valuable educational tools only when they are embedded in a model that preserves the learner as a responsible subject of learning. The future of education in the age of generative systems depends not on replacing human intellectual activity with algorithmic assistance, but on forming responsible co-agency in which trust is calibrated, agency is protected and responsibility remains human.

## Conflict of Interests

The author declares that there is no conflict of interests that could have influenced the objectivity of the study, the interpretation of the results or the presentation of the conclusions. The article was prepared independently, without external funding, institutional pressure or the involvement of organisations or individuals with a direct financial, personal or professional interest in the outcomes of the research.

The study is theoretical in nature and is based on the analysis of scholarly literature in the fields of philosophy of education, digital pedagogy, epistemic trust, learner agency, artificial intelligence in education, generative AI, AI ethics and human–AI interaction. The selection and interpretation of sources were carried out in accordance with the aim, objectives and methodological framework of the article.

The author confirms that there were no financial, institutional or personal circumstances that could be interpreted as influencing the research position, conceptual framework, methodological approach or conclusions of the study. All results presented in the article are based on independent theoretical analysis, conceptual reconstruction and philosophical-pedagogical modelling.

*Thus*, the present declaration confirms compliance with the principles of academic integrity, publication transparency and ethical standards of scholarly research.

## References:


Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*(2), 164–180. https://doi.org/10.1111/j.1745-6916.2006.00011.x

Benanti, P. (2022). *Human in the loop: Human decisions and artificial intelligences* [*Human in the loop: Decisioni umane e intelligenze artificiali*]. Mondadori Università. (In Ita.)

Biesta, G. J. J. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. Paradigm Publishers.

Floridi, L. (2017). *The fourth revolution: How the infosphere is transforming the world* [*La quarta rivoluzione: Come l'infosfera sta trasformando il mondo*]. Raffaello Cortina Editore. (In Ita.)

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy, 82*(7), 335–349. https://doi.org/10.2307/2026523

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, Article 100020. https://doi.org/10.1016/j.caeai.2021.100020

Rivoltella, P. C., & Rossi, P. G. (Eds.). (2019). *Technologies for education* [*Tecnologie per l'educazione*]. Pearson. (In Ita.)

Roe, J., & Perkins, M. (2024). *Generative AI and agency in education: A critical scoping review and thematic analysis*. arXiv. https://arxiv.org/abs/2411.00631

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0