Cognitive Entanglements: A Postphenomenological Approach to Generative AI in Education

Lorenzo De Stefano(✉)

University of Naples Federico II, Naples, Italy
lorenzo.destefano@unina.it

Abstract. This paper reframes debates on artificial intelligence in education by shifting focus from the question "Do machines think?" to the pedagogically salient issue of how large language models (LLMs) redistribute cognitive agency in learning environments. Drawing on postphenomenology and the extended mind thesis, we argue that LLMs are not neutral tools but mediators that co-constitute human–technology–world relations. Using Ihde's modalities (embodiment, hermeneutic, alterity, and background), layered with Fasoli's taxonomy of cognitive artifacts (constitutive, complementary, substitutive), we propose a framework for analyzing and designing the use of LLMs in education. The framework clarifies when LLMs scaffold interpretation and reasoning and when they displace it, linking these patterns to epistemic virtues and risks of technologically induced cognitive diminishment. We translate this analysis into pedagogical patterns that make AI participation visible and accountable. An illustrative teaching scenario in philosophy shows how LLMs can serve as interlocutors and co-interpreters while assessment shifts from product to process via prompt logs, revision trails, and oral defenses. We contend that technological neutrality is untenable: institutional policies, curricula, and assessment regimes actively configure cognitive ecologies. Consequently, education should stabilize constitutive and complementary uses, preserve non-substitutable spaces for independent reasoning, and embed transparency and AI literacy across programs. Education's task is not to reject or surrender to AI, but to design entanglements that expand, rather than erode, human agency.

Keywords: Postphenomenology · LLM · Education

1 The Problem of AI in Education: Moving Beyond the "Thinking Machine" Debate

Public and scholarly debates about Generative Artificial Intelligence (GenAI) have, for the most part, been organized around an old ontological axis: Do these machines think? Ever since the Turing Test was proposed as a heuristic for machine intelligence [74], the discourse around AI has returned, again and again, to the question of whether computational systems exhibit or simulate what we call "thought." This question has its place in

M. Adamoli et al. (Eds.): HELMeTO 2025, CCIS 2959, pp. 337–360, 2027.
https://doi.org/10.1007/978-3-032-31853-4_22

the philosophy of mind and has, in recent years, become one of the most urgent questions for educators, students, policymakers, and scholars in the field of artificial intelligence in education (AIEd), who are grappling with the day-to-day transformations these tools bring into the classroom. A different, more pedagogically and phenomenologically acute question should guide our inquiry: "How do these technologies reorganize cognitive agency, especially when they enter educational practice?" This formulation connects the ontological status of the machine with the structure of human–technology relations. It invites us to look at the mediated character of cognition rather than "machine thinking". When a student writes with the assistance of a large language model (LLM) or searches with a generative search interface, the task structure itself changes. The agentive locus of interpretation, composition, and reasoning is redistributed across human and non-human actors. In this sense, asking whether an LLM "thinks" is less useful than examining how its integration reorganizes who does what in the epistemic economy of the classroom. This question is particularly urgent in higher education, where LLMs challenge established practices of academic writing, assessment, and university students' epistemic development.

The field of Artificial Intelligence is experiencing remarkable growth, primarily fueled by advances in LLMs and GenAI platforms including ChatGPT, Claude, DeepSeek, Gemini, Mistral, and Llama, which can now produce highly convincing text and multimedia outputs automatically. For the first time in history, we are in a situation where it is almost impossible to determine, beyond reasonable doubt, whether a text was written by a human or generated by a machine. This has profound implications for research and assessment in higher education [33, 81]. These GenAI tools consist of deep neural networks containing billions or potentially trillions of parameters. Their "training" process—the determination of these parameter values—is carried out through gradient-based optimization techniques applied to enormous datasets collected from internet sources, encompassing website text, published books, and programming code. The fundamental training objective is to predict the next token (either a complete word or a word fragment) in any given sequence with high probability. Following this initial pretraining stage, these models undergo additional refinement using reinforcement learning techniques to enhance their behavioral characteristics and ensure better alignment with human values. Because each generated token conditions the next, the model produces sequences whose coherence can convincingly mimic the flow of human thought. Users can guide LLM outputs through prompts ranging from simple to complex, which may improve the relevance and precision of generated content. This practice is spreading rapidly in both education and academic work [49]. The remarkable ability of LLMs to comprehend human instructions and respond by generating well-organized, coherent, and convincing responses spanning virtually all areas of human knowledge has generated significant enthusiasm regarding their revolutionary potential and a certain overoptimism about their imminent achievement of Artificial General Intelligence (AGI). Nevertheless, the methodologies used to evaluate the accuracy, trustworthiness, and inherent biases within the training data that shape LLM outputs remain largely unclear [50]. Questions persist regarding whether data collection and LLM training procedures adequately respect copyright protections and privacy requirements [51]. Despite these concerns, LLM technologies have experienced extraordinary adoption rates in the two years following GPT-3's introduction. There is growing agreement that LLMs will be swiftly incorporated into countless human endeavors and automate many tasks with substantial impacts on the character of work, productivity levels, and decision-making processes [60]. These tools also possess the potential to serve as powerful aids to conventional scholarly practices, providing assistance with exploring and responding to diverse questions, composing text, consolidating and summarizing information, and proposing solutions to problems across diverse knowledge domains. The extraordinary capacity of AI to perform human cognitive tasks has naturally raised the question of our own educability [75]. However, LLMs neither provide systematic reasoning nor disclose information regarding the sources utilized in their training. Their probabilistic architecture and dependence on unverified internet-scale datasets make synchronizing their outputs with moral values, factual accuracy, and scientific principles an ongoing challenge [11]. The apparent originality and inventiveness they exhibit when generating responses to human prompts and producing new concepts are frequently challenged on the grounds that their outputs merely reproduce previously encountered information, lacking genuine authenticity and true creativity [8, 82]. Consequently, as LLM-based services evolve and multiply, they may produce unexpected ramifications, disrupting established frameworks for learning and scholarly endeavor. This transformation will affect education at all levels, potentially impeding the cultivation of social and emotional intelligence among young individuals, and may even alter fundamental human characteristics including empathy, curiosity, critical thinking, and creativity.

Given the lack of a clear understanding regarding how LLMs acquire knowledge [61] or the intermediate processes by which they arrive at their outputs, relying on them without independently and critically assessing their output against trustworthy sources can be hazardous.. Nevertheless, LLM tools are being rapidly integrated into educational settings, owing to their impressive capabilities, speed, and convenience at minimal cost. To analyze their inevitable influence on higher education, we must consider it in relation to the university's fundamental values and mission, specifically: to assist students in i) mastering the theoretical foundations of their discipline, ii) comprehending the scientific method, iii) developing sophisticated mental frameworks for the reasoned integration of facts, scientific principles, and moral values; iv) fostering the capacity for self-reflection, critical thinking and autonomous learning, and v) strengthening their ability to articulate complex reasoning clearly and creatively through written, oral and visual expression. To achieve these objectives, students have traditionally participated in scholarly work requiring personal dedication, thorough exploration, experimentation, and time for profound contemplation and reflection. Yet, with LLMs rapidly becoming the default intermediary between humans and our information ecosystem, in some cases replacing direct access to information sources, students may soon experience a reconfiguration of their learning environment and cognitive organization [31, 43]. The risk is serious: students could lose opportunities to cultivate core cognitive and intellectual capabilities, along with the agency fundamental to learning and scientific investigation. Consequently, they may become progressively reliant on LLM tools and susceptible to their biases, hallucinations, and concealed inaccuracies [7]. Conversely, preliminary research indicates that when LLMs are employed prudently and within controlled parameters, they can facilitate the learning process [26, 52, 68], for instance by extending personalized mentoring opportunities, which are frequently constrained by limited human resources, or by providing assistance with improving writing style, expediting routine tasks, brainstorming, conducting preliminary literature research, and similar activities. Additionally, across numerous fields of science and technology, including coding, biotechnology, and engineering, AI methodologies show promise in augmenting standard practices and workflows and even appear capable of generating novel concepts and discoveries [67, 69, 73]. As a result, LLM techniques are being swiftly incorporated into professional systems, streamlining problem-solving processes and reducing cognitive load by delegating tedious and routine tasks, ultimately enhancing human productivity. On that basis, exposing students to both the capabilities and dangers of such techniques may often be essential, as this will better equip them with the skills required in life and in the future workplace. However, intelligent machines can neither replicate the diversity and evolving nature of human intellect nor embody the depth of human agency [57, 65]; this implies that societies and educational systems must not surrender the cultivation of human thought and values to automated, algorithmic convenience. The challenge ahead is not whether LLMs should be utilized, but whether it is possible to guide their adoption in a manner that genuinely supports human development, maintains academic integrity, and aligns with the core humanistic values and missions of universities and schools. Even before considering how to address LLM adoption, societies need to consider whether they still uphold these principles. This is not a new problem. Each major media transformation in human history has been accompanied by intense cultural negotiation over the meaning of cognitive labor. In the *Phaedrus*, Plato famously depicts Socrates warning that writing would lead to forgetfulness, because people would rely on external inscriptions rather than cultivating internal memory. Skeptics of the printing press in early modern Europe feared the destabilization of oral traditions and the loss of scholarly authority [25, 30]. Critics of calculators in the 1970s warned of the decline of mental arithmetic; critics of digital ICT worried about the erosion of deep reading [80]. These fears were often exaggerated but not entirely unfounded. In a recent preprint titled *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task*, Kosmyna et al. [43] examine the cognitive and neural effects of using an LLM for essay writing. In a longitudinal experiment with 54 students assigned to LLM, search-engine, or brain-only conditions, the authors combined EEG measures, NLP-based text analysis, essay scoring, and interviews. They report that LLM use was associated with reduced neural connectivity, weaker memory recall (e.g., impaired quoting ability), and lower perceived ownership compared to the other groups, suggesting that sustained reliance on AI assistance may generate “cognitive debt” despite short-term efficiency gains. This suggests that media reconfigure cognitive capacities—sometimes weakening skills once essential, while enabling others that were previously unimaginable. This dynamic has been highlighted by various scholars from different perspectives and traditions [1, 16, 53, 54, 59]. The introduction of LLMs marks a turning point because these systems simulate not only data retrieval or calculation, but also linguistic and argumentative performances once thought to be uniquely human. This makes the redistribution of agency more profound: not just recall or calculation, but writing, interpreting, and reasoning are partially externalized in artificial beings that seem to possess a degree of autonomy and agency. That is the real philosophical and pedagogical problem. To analyze this question adequately, we require a framework that accounts for (a) the modalities of technological mediation, (b) the plasticity of human cognition, and (c) the normative and pedagogical stakes of educational practices. Postphenomenology and the extended mind thesis (EMT) together offer such a framework, but to appreciate their significance we must first understand the long genealogy of cognitive artifacts.[1] This framework—theoretical in scope, with empirical elaboration reserved for subsequent work—may inform higher education practices navigating the integration of LLMs into teaching, assessment, and institutional governance. To lay the foundations for the theoretical framework, it is necessary to begin with an anthropological–philosophical premise. Educational institutions and universities face unique challenges: maintaining academic standards while adapting pedagogy, preserving epistemic rigor in research training, and developing institution-wide policies that balance innovation with educational integrity.

2 A Genealogy of Cognitive Artifacts

The history of human evolution is also the history of its artifacts and technologies. André Leroi-Gourhan emphasized that, in human beings, biological evolution has given way to technological evolution [48]. It follows that the evolution of cognition is, at the same time, a history of the exteriorization of mental functions (e.g., memory) into tools, media, and shared environments. From the earliest marked stones and trails to contemporary AI systems, humans have co-evolved with their artifacts [48]. However, as Günther Anders, among others, noted [2], every technological novelty introduces a diachrony, a gap between the potentiality of the new technology and our imagination and representation of its effects and outcomes.[2] This means that learning and cognitive development are always environmental processes involving the interaction of the individual with others, with society, but also with the system of technologies in which they are immersed. Two of the most prominent figures in the pedagogy of the last century — John Dewey and Lev Vygotsky — emphasized that learning is a broad and social phenomenon. Dewey argued that learning takes place most effectively when students are engaged in practical and meaningful activities, rather than passively receiving information [23, 24]. *Learning by doing* involves both other people and the tools used in the activity; learning is always a contextual dynamic that engages both human and non-human actors. Vygotsky [78] further stressed that learning is essentially social and cultural and develops through relationships with other intentional subjects (teachers, peers, more experienced adults). The concept of the *Zone of Proximal Development* (ZPD) refers to the difference between what a child can do alone and what they can do with the help of an adult or a more knowledgeable peer. It is a dynamic space where potential learning takes place, made possible

[1] With Norman, we define a cognitive artifact as any artificial device, physical or digital, that maintains, displays, or operates on information to serve a representational function, thereby reorganizing the user's task and distributing cognition across the person, artifact, and environment [59].

[2] In his masterpiece *Die Antiquiertheit des Menschen* Günther Anders calls the discrepancy between what we can produce and what we can imagine the Promethean gap (*Prometheische Gefälle*).

through support, a process known as *scaffolding*. As we will see, scaffolding today can also be mediated by non-human actors such as AIs and LLMs, without diminishing the inherently social and mediated nature of learning. According to Vygotsky's theoretical framework, educators themselves act as mediators and supports for autonomy; equally central is the mediating activity of the most powerful communicative technology ever invented: language. We will see how these theories—particularly those of Vygotsky—remain extraordinarily relevant in learning mediated by ICT, which continues to be based on language and information technologies. Digital ICTs thus seem, on the one hand, to extend our cognitive faculties such as memory, language, calculation, and even artistic creativity [8], and on the other hand, to have a feedback effect, acting as new intentional agents. Postphenomenology has emphasized that technologies are not only products of human creative activity but also shape us in turn, as Verbeek [77] and Ihde and Malafouris [39] have noted; this reciprocal relation inaugurates a new kind of "artificial intentionality" [57, 58, 65]. As we will see, this theory of technological intentionality will provide a valid hermeneutic criterion for understanding how LLMs operate and for developing a hermeneutics of ICT in education. This insight is older than either framework. It runs through nineteenth- and twentieth-century philosophical anthropology and technological thought.

2.1 Organ Projection and Exteriorization

Ernst Kapp's 1877 *Grundlinien einer Philosophie der Technik* introduced the concept of organ-projection (*Organprojektion*), arguing that tools and technologies are externalizations of human organs and faculties. The hammer is the projected fist; the telescope, the extended eye; and infrastructures and telegraph cables externalize the nervous system [42]. For Kapp, technology is not merely an addition to human capacities but an exteriorization of those capacities in the world, which then feeds back into human existence. Exteriorization is simultaneously an extension and a transformation: once an organ is externalized, the human–world relation changes irreversibly. This early insight can be seen as a proto-theory of extended cognition *avant la lettre*. By projecting organs outward, humans create technological environments that alter their own behavior, perception, and cognition. In the early twentieth century, Paul Alsberg challenged Kapp's theory in *Das Menschheitsrätsel* (later published in English in a reduced version as In Quest of Man [1]) and added a critical nuance with the concept of *Körperausschaltung*—the deactivation or "body-liberation" of physical and cognitive functions once they are externalized in technologies. If a tool performs a function, the corresponding organ or faculty may atrophy from disuse and be replaced. However, this should not be confused with what Arnold Gehlen later called *Entlastung* (relief) [32]. Whereas Gehlen indeed sees the human being as a *Mängelwesen* (deficient being), Alsberg does not consider the technical nature of humans as originating from a structural lack, but rather from an active plastic faculty. Tools relieve humans of certain tasks, but this relief can produce both freedom and dependency. A hunter who relies on a spear can strike at a greater distance but may lose skills of hand-to-hand combat. A society that relies on writing need not memorize epics, but its mnemonic ecology changes fundamentally. Overreliance on GPS affects spatial orientation abilities, and the compulsive use of the Internet and search engines has been studied and shown to weaken our memory capacity and information processing [12, 19]. Summing up the main insights of early twentieth-century philosophical anthropology, we can identify what can be referred to as the threefold operation of technology: (i) *exteriorization* of a function or a capacity; (ii) *enhancement* of that function through externalization; (iii) *deactivation/liberation*, or refunctionalization of the original embodied skill. This triadic structure, implicit in technological history, remains underappreciated in contemporary debates on AI.

2.2 Writing, Print, and Digital Media

The shift from orality to writing is perhaps the most striking early instance of this threefold dynamic. Writing exteriorized memory, enhanced cultural storage, and deactivated oral mnemonic techniques [25, 48]. But the story does not end in loss: writing also enabled new forms of reasoning—linear argumentation, propositional logic, formal proof structures—that oral traditions alone did not sustain [21, 35, 53]. The printing press extended this dynamic by enhancing distribution, externalizing scholarly authority and eventually reshaping epistemic communities. In The Gutenberg Galaxy [54], Marshall McLuhan framed print as a medium that reorganized human perception and social structures. The linear, visual, and uniform logic of print fostered individualism, standardized language, and encouraged rational, abstract thought, contrasting sharply with the communal and oral character of pre-print culture. As a “hot” medium, it shaped modern subjectivity by privileging detached, analytical modes of engagement—a cognitive and cultural shift later unsettled by the re-tribalizing effects of electronic media [53]. Digital media in the late 20th and early 21st centuries followed a similar arc. Personal computing exteriorized calculation, enhanced processing power, and deactivated some mental arithmetic. Matteo Pasquinelli highlights how computers and AI were conceived in their early stages by Babbage as ways to automate calculation and cognitive labor [62]; Alan Turing, by equating thought with calculation, believed that, in principle, a universal Turing machine could reproduce and automate all human reasoning [74]. Turing outlined a horizon of research that, although pursued with technologies very different from those he used, such as machine learning and artificial neural networks [9], has ultimately led to contemporary AI. AI continues along these lines: the complete externalization of human thought and cognition. With LLMs, however, we reach a new layer: exteriorization of discursive and interpretive labor itself. What writing did to memory, AI may do to reasoning, argumentation, and composition. From this perspective, we can grasp the epochal challenge that AI poses to philosophy, pedagogy, and education: it is a technology capable of refunctionalizing cognitive and learning processes [21].

3 Postphenomenology and Technological Mediation

3.1 Ihde’s Framework: Mediation, Embodiment, and Perception

Don Ihde’s postphenomenology offers in this regard a powerful conceptual vocabulary and a taxonomy of human–technology relations for understanding how technologies shape human experience. Unlike classical phenomenology, which often emphasizes the subject-object relation, postphenomenology inserts technologies as mediators in this relation. Moreover, it conceptualizes intentionality not just as a "mental" feature, but as an embodied tension integrating Merleau-Ponty's views [55] on intentionality. When I look at the world through a microscope, I do not encounter the world directly; my perception is technologically mediated. Don Ihde [37] identifies four main ways in which technology mediates human experience. In *embodiment* relations, technology merges with the body, becoming an invisible extension that shapes perception, as with glasses or a blind man's cane. In *hermeneutic* relations, it provides representations—like thermometer readings or MRI scans—that require interpretation rather than direct experience. *Alterity* relations involve technologies that act as quasi-others, such as conversational agents, robots, or LLMs, with which users interact dynamically. *Background* relations, by contrast, operate silently in the periphery, structuring experience without conscious attention, as in climate control systems or pacemakers. These modalities illustrate how technologies do not merely mediate but actively configure human intentionality and the way the world is experienced (Table 1).

Table 1. Postphenomenological Relations Table

Type of relationship	Schematic representation
Embodiment relations	(Human–Technology) → World
Hermeneutic relations	Human → (Technology–World)
Alterity relations	Human → Technology (–World)
Background relations	Human → (Technology/World)

These relations structure perception and action in different ways, but in all cases, the human–world relation is technologically co-constituted. Technologies are not external tools but integral to lived experience. Ihde advocates a material hermeneutics [40] that extends its object from texts to tools and technologies. Grasping the hermeneutic process involves, in this case, the interpretation of the material elements that make up the figurative text; in other words, it means understanding the physical medium through which the figurative text is expressed [34]. In later work, Ihde [38] elaborated the idea of re-embodiment: technologies become incorporated into our bodily schema rather than simply being used. Glasses are not objects we constantly perceive but part of our perceptual apparatus. Similarly, digital devices—especially smartphones—often function as perceptual and cognitive prostheses: they are not merely "at hand" in a Heideggerian sense, but integrated into the body schema in everyday comportment. This insight is crucial for AI: for many students, interaction with LLMs is not a discrete "use" but a continuous, ambient mode of thought. The system sits beside or in front of the learner, ready to answer, rewrite, summarize, or generate ideas. It becomes, in Ihde's sense, part of the intentional arc through which the learner relates to the world of knowledge. Peter-Paul Verbeek extends Ihde's framework by distinguishing between mediated, hybrid, and composite forms of intentionality [77]. Mediated intentionality involves perceiving or acting through a technology; hybrid intentionality emerges when human and technological agencies merge (e.g., cochlear implants, neural interfaces); composite intentionality arises when human intentionality interacts with the quasi-intentional structures of technologies themselves, such as algorithmic decision systems. LLMs such as Claude and ChatGPT are paradigmatic examples of composite intentionality. They are not neutral interfaces but models whose embedded statistical priors, discursive scripts, and alignment mechanisms shape how questions are asked and answered. LLMs can act as genuine co-authors, as research assistants, and even as doubles that stimulate both students and teachers in their practices. The practice of the digital double—an AI trained on content we have produced or small language models trained on specific tasks—represents the latest frontier in which analog authorship will be increasingly hybridized with digital systems [21]. In this context, we will no longer be solely authors, but also curators of content generated in part by AI. Their affordances structure the user's epistemic orientation—guiding, scaffolding, nudging, sometimes misleading. To teach with or alongside such systems is to teach within a technologically configured epistemic space.

3.2 Integrating Postphenomenology with EMT and 4E Cognition

Andy Clark and David Chalmers [17] famously proposed the extended mind thesis (EMT): cognitive processes are not confined to the skull but extend into the environment through the coupling of brain, body, and world. A notebook used reliably and transparently to store information is, under the right conditions, part of the cognitive system. This claim, once controversial, is now widely discussed in philosophy of mind, philosophy of technology, cognitive science, and human–computer interaction. EMT offers a conceptual justification for treating technologies as integral components of cognitive activity, rather than as external aids. A student who routinely co-writes with an LLM is, in EMT terms, thinking with and through the system. According to Clark [14], extending the mind, memory, and other cognitive faculties into artifacts represents a strategic and intelligent evolutionary response rather than a merely reactive one. What has made humans remarkably intelligent is not simply the brain itself, but our unique capacity to create, manipulate, and integrate external supports for thought [14]. In light of the arguments presented thus far, such a statement could have been readily endorsed by Leroi-Gourhan, Clark, and McLuhan. In *Natural-Born Cyborgs* [14], Clark develops this idea further, arguing that humans are distinctive not for having disembodied minds but for being profoundly prosthetic creatures. Our biological endowment is technological by design: we are technology-ready animals, evolved to couple with artifacts to extend cognition. The stone tool, the stylus, the printing press, and the smartphone are all cognitive partners, not merely supports. This framing reorients the ethical debate. If externalizing cognition is a species-typical behavior, then the question is not whether we should externalize, but how. The problem is not technology per se, but the form of entanglement it produces. The 4E framework pushes this argument further (Embodied, Embedded, Enactive, Extended). Cognition is not only extended into artifacts; it is embodied in biological structures, embedded in environments, and enacted through sensorimotor engagements. This holistic view resists abstraction: cognition is situated, contingent, and deeply intertwined with material practices. If integrated wisely, AI and LLMs can serve as cognitive partners embedded in rich educational ecologies. At this juncture, one might ask: Why rely on postphenomenology and EMT? The answer lies in their focus on relationality. Postphenomenology highlights the mediating role of technologies in shaping perception and action, without obliterating the difference between subject and object. It allows us to understand how AI co-constitutes the learning environment and the modalities of this relation. EMT and 4E cognition, despite their different emphases, situate human intelligence not only in the subjective pole but across hybrid networks of brain, body, artifacts, and environment. Together, they shift our perspective from individual deficits or benefits to the structure of cognitive entanglement. For educational research and practice, this is a profound shift. It means that LLMs are neither neutral tools nor autonomous agents, but participants in distributed cognitive systems whose configuration determines whether learning is enhanced or diminished. This is critically important for educational environments. Reworking the EMT, Marco Fasoli [28] offers a taxonomy of cognitive artifacts that sheds light on the different ways in which cognitive artifacts operate within human–technology relations, identifying three main roles: *constitutive*, *complementary*, and *substitutive*. When functioning in a constitutive role, the artifact becomes an active part of the cognitive process itself—for example, when diagrams are used to structure and support reasoning, they do not merely assist thought but participate in it. In a complementary role, the artifact provides scaffolding that aids cognition while leaving the core of cognitive agency with the user, as in the case of outlining software that helps writers organize their ideas without taking over the creative process. In a substitutive role, by contrast, the artifact performs the task on behalf of the user, such as when an AI system automatically generates an entire text. Crucially, these roles are not fixed properties of technologies but depend on specific relations and contexts. The same LLM, for instance, can play a constitutive role when it is prompted to generate counterarguments in a debate, a complementary role when it is used to clarify complex concepts, or a substitutive role when it is tasked with producing a full essay. Moreover, these roles can shift over time: what begins as a complementary tool can gradually slide into a substitutive function as habits and expectations evolve. Recognizing this dynamic is essential for designing educational practices and institutional safeguards that encourage the constitutive and complementary use of technological tools, while limiting forms of unreflective substitution that may undermine active cognitive engagement.

4 The State of the Art on LLMs in Education

The current state of the art on LLMs in AIEd reveals a complex and rapidly evolving research landscape at the intersection of technological innovation, pedagogical transformation, and epistemological reflection. LLMs such as ChatGPT have emerged as highly accessible and versatile cognitive artifacts capable of producing coherent, human-like responses [22, 64]. Unlike earlier educational technologies, their accessibility, personalization, conversational structure, and low cost have led to their deep integration into everyday educational practices [20, 41]. Their introduction into schools has been accompanied by both enthusiasm and concern. Early debates centered on whether to restrict or regulate their use in classrooms, as researchers and educators began to document both the opportunities and the risks associated with their integration. On the one hand, numerous studies identify significant benefits of using LLMs in schools and universities: they can support educators through time-saving automation in assessment and content creation

[5, 83, 84], enhance pedagogical practices by generating lesson plans, quizzes, and differentiated materials, provide personalized tutoring and adaptive feedback [56], scaffold brainstorming and outlining processes [41], and assist in creative ideation. On the other hand, concerns about academic dishonesty, overreliance, and cognitive deskilling have also been widely reported, with fears that students may outsource essential intellectual operations, weakening skills such as critical thinking, problem-solving, and independent inquiry [18, 26, 41, 43, 70]. Recent scholarship converges on this non-neutrality thesis [6, 13], providing empirical grounding for the postphenomenological framework developed below. This has led to a growing emphasis on developing AI literacy and epistemic agency, ensuring students learn to critically evaluate AI-generated content rather than passively accepting it [10, 16, 29]. Parallel research has focused on the development of rigorous benchmarks and evaluation protocols to ensure pedagogical soundness and technical reliability. Equity and access considerations are also central to this research frontier. LLMs offer the potential to democratize access to high-quality, individualized learning experiences; however, they may also exacerbate educational inequalities if infrastructural disparities, inequitable assessment practices, and model biases remain unaddressed [46]. At the same time, ethical and regulatory concerns—including questions of privacy, academic integrity, authorship, and accountability—have become central policy issues as institutions grapple with their large-scale adoption. Philosophically, these developments resonate with the theoretical framework outlined above of technological mediation [37, 38, 77] and EMT [14, 15, 17], and cognitive exteriorization [48, 54], which frame LLMs not as external tools but as components of distributed cognitive systems. What is certain is that LLMs represent transformative infrastructures in education: they extend, augment, and sometimes substitute cognitive functions, enhancing personalization, scalability, and creativity, while introducing significant epistemic, pedagogical, and ethical challenges and concerns. Their educational value ultimately depends both on their intrinsic capabilities and on the relational, institutional, and cognitive ecologies within which they are embedded.

5 From Conceptual Framework to Pedagogical Design

The theoretical architecture developed thus far, integrating multiple layers (anthropology of technology, postphenomenological theory of mediation and Fasoli's taxonomy) must now be applied to concrete educational situations, if it is to gain real pedagogical and policy traction. Schools, universities and educational institutions are not abstract laboratories: they are structured, time-constrained, norm-laden environments where technological adoption interacts with assessment regimes, curricula, and diverse epistemic expectations. This section proposes an application of the threefold action model (exteriorization, enhancement, deactivation) to the specific case of LLMs in contemporary educational contexts. Rather than focusing on speculative futures, it examines existing and emerging practices through a postphenomenological lens, offering a structured way to design, analyze, and evaluate how these systems are integrated into classrooms.

The cognitive role that AI plays in education is determined not by technological intentionality alone but by the design of learning tasks and environments, as well as by institutional norms. We can thus steer AI use toward constitutive and complementary roles—those that enhance learning without displacing epistemic agency—and away from substitutive roles that accelerate cognitive deactivation. Postphenomenology teaches us that technologies are not external add-ons to human activity but active mediators of perception, cognition, and action. In the classroom, technologies do not merely support learning; they constitute the conditions under which learning takes place. The notions of composite and hybrid intentionality described by Verbeek [77] help us to focus on those dynamics. For example, a lecture hall with chalkboards affords different cognitive rhythms than a laptop-filled classroom. A seminar room oriented toward discussion stimulates different attentional structures than a computer lab. Once LLMs are introduced, the cognitive ecology shifts again. For example, students can externalize aspects of interpretation, summarization, and argumentation at will; responses are instantaneous, altering the temporal structure of problem-solving; opacity and fluency can generate epistemic over-trust in machine outputs; teachers and other educators may be pressured to adapt assessment structures rapidly.

The challenge is not simply to "allow" or "ban" LLMs but to design the learning ecology so that the technology functions as a productive actor rather than a mere substitutive one [28]. Traditional educational technology adoption often treats tools as invisible infrastructure—textbooks, projectors, search engines. LLMs, however, are high-agency technologies: they produce text, suggest ideas, formulate arguments, and simulate reasoning. Their presence is visible and agentive; LLMs enable what Ihde calls alterity relations, acting as a quasi-other. By updating the postphenomenological framework, we can observe how LLMs operate through hermeneutic, alterity, and background relations. A hermeneutic relation arises when an LLM assists students in clarifying or interpreting a text, identifying new connections within texts or other intellectual works, and offering new interpretations or co-interpretations. In this case, LLMs themselves operate as hermeneutic machines [36]. This mode of engagement is particularly valuable in the teaching of literary and philosophical subjects, where interpretation plays a central role. An alterity relation arises when LLMs can act as intellectual doubles, for example in philosophical debate activities [66], by serving as sparring partners that embody the perspective of an author or philosopher. They can also prepare students for oral examinations by simulating the teacher's perspective or by acting as peer tutors, thereby activating and personalizing cooperative learning dynamics and scaffolding processes, as well as enhancing oral and rhetorical skills [27, 47], without diminishing the role of educators and while emphasizing the need for a balanced integration of AI and human tutoring methods. Finally, a background relation occurs when the learning environment is technologically mediated from the outset, as in video-essay production, augmented reality classrooms or distance learning settings. These three modalities, drawn from postphenomenology, provide insight into the structure of human–technology relations. Upon this phenomenological taxonomy, which describes the '*what*' of the relationship, we can layer Fasoli's framework, which specifies *how* this relationship operates.

When used *constitutively*, an LLM becomes part of the student's cognitive process without replacing it. A classic example is using diagrams in reasoning: the diagram is part of thinking, not an external object. Similarly, when students prompt an LLM to generate counterarguments and then engage in reasoning with those outputs—testing, refuting, integrating—they are thinking *with* the system, not deferring to it. Examples include: (i) Socratic debate simulation where the model plays an interlocutor; (ii) collaborative ideation sessions where prompts structure the argumentative space but do not fix conclusions; (iii) using LLM outputs as material for analysis rather than final answers. Constitutive uses amplify cognitive effort rather than replacing it. They are especially suited for higher-order cognitive skills—critical thinking, synthesis, evaluation—rather than rote reproduction.

Complementary roles occur when the artifact scaffolds cognition—removing extraneous obstacles—while leaving the core epistemic act to the learner. LLMs excel at: (i) rephrasing instructions in simpler language; (ii) generating glossaries or conceptual maps; (iii) translating material for multilingual learners; (iv) providing structured feedback on drafts. Here, AI supports epistemic agency by lowering friction, not by supplying solutions. Complementary use is particularly valuable for accessibility, differentiated instruction, inclusive pedagogy and formative assessment.

Substitutive use occurs when the artifact performs the cognitive task instead of the student, leading to minimal engagement; this can be useful to liberate time from the most schematic and compilatory tasks, but must not substitute core intellectual work. Examples include (i) asking the model to write an essay, then submitting it uncritically; (ii) requesting summaries of essays and papers instead of reading them; (iii) auto-generating solutions without understanding the underlying principles; and (iv) using the model to perform the more mechanical and schematic functions of academic production, such as assisting with copy-editing, organizing bibliographies, standardizing formatting according to the requirements of different journals, arranging tables, and producing slides. Substitutive use maximizes convenience but minimizes epistemic agency. If habitual, it accelerates what Fasoli et al. call technologically-induced cognitive diminishment (TICD) [29]. A crucial point is that these roles are enabled by the new forms of agency afforded by these technologies, but they remain use-dependent and context-sensitive. A task that begins as complementary (e.g., using AI for translation) can drift into substitution if students outsource interpretation as well. This drift mirrors both historical patterns involving technologies such as calculators and search engines and the human tendency to conserve energy by externalizing and delegating physical and cognitive tasks to tools and machines. Educational practices must, therefore, stabilize constitutive and complementary roles through task structure, assessment criteria, and classroom culture.

One of the most promising areas of innovation in LLM integration concerns the design of pedagogical patterns that structure how these systems are used in learning environments [10]. A first pattern can be described as feedback amplification. Here, students begin by producing an initial draft of their work, e.g., an essay. The LLM is then prompted to map claims to supporting evidence, identify gaps in reasoning, or suggest counterarguments and relevant literature. Students subsequently evaluate this feedback, deciding what to accept or reject, and revise their text accordingly. Assessment focuses not only on the final product but also on the revision chain and the transparency of reasoning. This design encourages metacognition and active engagement, positioning the LLM not as a surrogate author but as a critical interlocutor. At the same time, it strengthens students' ability to detect when the model is wrong, cultivating epistemic virtues that are essential in an era of synthetic fluency. A second pattern concerns accessibility without displacement. For multilingual or neurodivergent learners, LLMs can serve as assistive technologies by translating technical language into more accessible forms, rephrasing complex questions, or producing structured outlines. Importantly, this support does not replace the core cognitive operations—such as analysis, synthesis, and argumentation—which remain firmly with the student. When structured correctly, this form of support helps level the epistemic playing field without eroding essential intellectual skills. Another key pattern is dialogic rehearsal. In disciplines such as philosophy, history, or law, LLMs can simulate Socratic dialogue or adversarial debate. Students define roles and constraints for the model, for instance instructing it to argue from the standpoint of a specific philosopher or tradition. The model responds, and the student is required to counter. Many apps, such as Debate AI or ArguMate, have been specifically developed to fulfil these functions [66]. Evaluation should then focus on the quality of the student's responses rather than the model's outputs, the ability to write effective and critical prompts, and the overall process [10]. This approach mirrors classical dialectical training while taking advantage of contemporary tools. It also helps to re-embed reasoning in dialogic rhythms, counteracting the passivity associated with substitutive uses of technology. Finally, triangulation as a pedagogical habit involves systematic comparison between multiple outputs. Students compare outputs from several models, prompt variations, or repeated queries, identify contradictions, and investigate their sources. Hallucinations and biases thus become objects of inquiry rather than obstacles to understanding. This process fosters epistemic vigilance and trains students to approach AI-generated content critically, regarding it as a set of claims to be examined rather than truths to be accepted. Over time, this practice contributes to the cultivation of intellectual virtues such as curiosity, rigor, and reflectiveness [3, 4].

6 Cognitive Diminishment and Its Thresholds

Cognitive offloading has long been a characteristic feature of human intelligence [1], shaping the ways in which individuals distribute mental tasks across external supports. However, the educational consequences of this practice depend greatly on its form, frequency, and function [29]. Temporary offloading, such as taking notes to externalize and later reinternalize information, often enhances learning by reducing cognitive load while maintaining active engagement. In contrast, chronic substitutive offloading, in which a task is consistently delegated to a machine—e.g., calculating with a calculator—tends to erode underlying skills and diminish cognitive autonomy. Between these poles lies hybrid offloading, in which learners alternate between externalization and internalization. This balanced approach can optimize performance while preserving and even strengthening cognitive abilities, as it encourages interaction between internal processing and external support. A crucial factor in understanding these dynamics lies in the existence of threshold effects. There are points at which offloading ceases to enhance learning and begins to deactivate core skills. These thresholds are highly task-dependent and difficult to detect. For example, occasional use of AI tools to help outline a paper may facilitate deeper learning by supporting structure and coherence, whereas habitual reliance on AI to generate entire texts can weaken argumentation and critical thinking skills. Similarly, extensive dependence on AI-driven search functions can undermine retention and synthesis by externalizing too much of the cognitive process [71]. For this reason, educators play a key role in identifying and managing these thresholds through thoughtful curriculum design, ensuring that technological supports reinforce learning rather than replace it.

7 Teaching Philosophy with LLMs

Philosophy represents a particularly rich testbed for the integration of AI into educational practice because it relies heavily on linguistic reasoning, demands the cultivation of epistemic virtues such as reflectiveness, open-mindedness, and curiosity [4], and is grounded in a long tradition of dialogic pedagogy, from the Socratic method to medieval disputation. These characteristics make philosophical education especially sensitive to the ways in which cognitive tasks are distributed between students and technological systems. Philosophical tasks such as argument reconstruction, critical commentary, and essay writing are particularly amenable to partial externalization through LLMs. These models can summarize primary texts, propose possible objections, or generate example arguments that students can engage with critically. When integrated thoughtfully, such uses do not replace the intellectual labor of the student but provide scaffolding and provocation for reasoning. A well-structured assignment can harness this potential by encouraging constitutive and complementary uses of technology. Students might, for example, generate counterarguments with an LLM, select and refine the most compelling ones, compose a critical response, and document their interaction with the system by providing a transparent record of prompts and outputs. In this configuration, the model functions as part of the reasoning process without displacing the student's agency or interpretive responsibility. By contrast, unstructured or substitutive use—such as asking the model to produce a full essay without engagement—risks eroding essential discursive and interpretive skills over time. If students are not required to reflect on, interrogate, or transform the outputs they receive, their ability to reason, argue, and interpret independently can weaken. This contrast underlines the pedagogical importance of designing activities that leverage AI's constitutive and complementary capacities while limiting its substitutive tendencies. If LLMs are to enhance rather than erode students' reasoning and interpretive abilities, educators must design activities that foreground process over mere output. Three complementary interventions support this shift: requiring revision trails and prompt logs makes the iterative student–model interaction visible and encourages metacognitive reflection; reorienting assessment from final product to reasoning process foregrounds epistemic agency rather than output; and oral defenses verify genuine comprehension [47], ensuring students can justify and elaborate the arguments they submit. These interventions realign educational priorities with epistemic responsibility, emphasizing not only *what* is produced but *how* it is produced. In technologically entangled learning environments, such shifts are essential to preserving the depth, reflexivity, and integrity of intellectual practice. In higher education contexts, these characteristics become especially pronounced as students develop scholarly argumentation skills and engage with complex disciplinary debates.

8 A Shift in the Evaluation Process

Cultivating epistemic virtues [3, 4, 63] as soft skills is central to structuring meaningful human–AI entanglements in education. The assessment of epistemic virtues should be incorporated into the overall process of competency evaluation. *Curiosity* enables students to transform prompting from passive consumption into active inquiry, encouraging them to probe uncertainty, test the limits of AI-generated answers, and seek out alternative perspectives. *Open-mindedness* and *attentiveness* enable students and educators to experiment with technologies in ways that minimize bias. Intellectual *humility* develops through recognizing that LLMs are prone to error, hallucination, and bias, fostering a critical stance and a willingness to revise beliefs based on evidence. *Intellectual courage*, *rigor*, *honesty*, and *perseverance* are cultivated when students learn to question, challenge, and trace machine reasoning through structured debates, triangulation tasks, and transparency requirements. These virtues anchor epistemic agency in hybrid cognitive systems. *Creativity* and *inquisitiveness* can be trained through complementary and constitutive relations with technologies. Moreover, integrating digital technologies into teaching helps to foster what Shannon Vallor calls *technomoral wisdom*: practical judgment, analogous to Aristotle's phronesis, concerning which virtues to apply, when, and how in novel technological situations [76]. Building on this foundation, the analysis moves from description to prescription: educational institutions must intentionally design policies and pedagogical structures that cultivate epistemic agency in technologically entangled environments. Neither banning AI outright nor allowing unrestricted adoption provides an adequate response. What is required is a structured, value-oriented integration that balances technological affordances with the cultivation of critical reasoning, reflective engagement, and intellectual responsibility. In higher education specifically, this might include redesigning capstone projects and thesis requirements to include mandatory documentation of AI interactions, implementing oral defenses for undergraduate theses and graduate dissertations to verify genuine comprehension, and developing rubrics that explicitly evaluate students' critical engagement with AI-generated content rather than merely assessing final outputs. Here is a comprehensive table that summarizes the postphenomenological relations, the roles of technologies, and the virtues that may be cultivated (Table 2):

Table 2. Postphenomenological and EMT frameworks in the teaching of philosophy[a]

Relation	Role	Example	Virtue stimulated
Hermeneutic	Constitutive	Close reading with counter-readings	Critical reflection
Hermeneutic	Complementary	Glossaries, simplified instructions	Curiosity
Hermeneutic	Substitutive	Automatic summaries	Passivity (risk)
Alterity	Constitutive	Socratic debate simulation	Courage
Alterity	Complementary	Feedback on theses	Dialogic openness

(*continued*)

Table 2. (*continued*)

Relation	Role	Example	Virtue stimulated
Alterity	Substitutive	Model-written essays	Dependency (risk)
Background	Complementary	Inline suggestions in LMS	Autonomy
Background	Substitutive	Auto assessments	Disengagement (risk)
Background	Constitutive/complementary	Writing copilot with revisions	Carefulness

[a] This table was developed independently of the findings reported by Laaksoharju et al. [45].

9 On Technological Neutrality

Given our analysis, it is now clear that education cannot be technologically neutral because technologies are enduring mediators that structure perception, shape action, and orient cognition. This non-neutrality can be articulated across three analytically distinguishable yet conceptually intertwined levels: (i) design-level non-neutrality; (ii) relational non-neutrality; and (iii) institutional non-neutrality. Design-level non-neutrality concerns the historically and culturally situated character of the datasets on which LLMs are trained. These systems inevitably sediment the epistemic hierarchies, normative orientations, and value-laden assumptions embedded within the socio-technical and economic milieus that produce them—what might be described, in a broader anthropological sense, as the "spirit" of a particular cultural formation.

Relational non-neutrality emerges from a postphenomenological understanding of technological mediation: technologies do not merely convey content but actively shape the horizon within which perception, interpretation, and action unfold. Their effects are recursive. By mediating practice, they gradually reorganize the very cognitive and normative structures that govern subsequent engagement [79].

Institutional—or, more precisely, politico-institutional—non-neutrality arises from assessment regimes, curricular architectures, and governance frameworks that distribute epistemic labor in specific ways. Yet it also reflects the strategic interests of states, private corporations, and broader economic configurations—often described under the rubric of platform capitalism or surveillance capitalism [85]—which operate as structurally entrenched forces and frequently resist ethical regulation. Although analytically distinct, these three levels are dynamically interwoven. It follows that technologies must be understood systemically and ecologically, as elements within complex socio-technical constellations rather than as isolated instruments whose effects can be confined to a single domain. Universities operate within asymmetrical relationships vis-à-vis platform providers, often lacking transparency regarding training data, model updates, or embedded optimization logics. Moreover, automation promises cost reductions and productivity gains, creating institutional pressures that may subtly incentivize substitutive deployments. Without acknowledging these structural conditions, calls for epistemic stewardship risk placing responsibility solely on individual educators and students. This implies that governance must address not only classroom design but infrastructural dependency and institutional bargaining power.

Policy, then, is non-neutral: assessment regimes, curricula, and institutional rules actively configure epistemic practices and, in AI-mediated environments, either counteract or accelerate drift toward substitution. In this sense, policy is a form of world-making that distributes thinking across human and machine actors. A normative response follows: require students to critically assess and take responsibility for what they offload; structure adoption to favor constitutive and complementary uses while guarding against substitutive drift; cultivate the epistemic virtues detailed in the preceding section; embed AI in rich cognitive ecologies that protect the slow temporalities of reflection; integrate transparency practices into assessment design and governance.

These principles should translate into governance instruments: assessment reform that evaluates processes and epistemic moves; curriculum-level integration that scaffolds AI literacy, alternates bare-brain and hybrid tasks, and teaches epistemic virtues explicitly. Clear red lines are also required for non-negotiable cognitive spaces—personal reflection, methodological design, original argumentation, and capstone projects—where AI may provide support but never replace intellectual work. Accessibility policies should leverage LLMs to reduce the digital divide and other barriers for multilingual and neurodivergent learners without collapsing into dependency, pairing complementary allowances with monitoring for drift and alternative skill-building pathways, while inclusive pedagogy ensures AI literacy does not become a new epistemic privilege. Because LLMs are opaque partners, institutions must build trust through transparency practices—prompt/output logging, triangulation across sources, and uncertainty tagging—that enable vigilance and auditing. These strategies should be situated within diverse global regulatory contexts (e.g., EU risk-based regimes versus fast-adoption models), aligned with institutional missions and academic freedom, and anchored in a shared ethos of cognitive stewardship: AI in education should prioritize human cognitive flourishing over the economic interests of private companies and service providers. Beyond restriction, a design-ethics approach proactively shapes cognitive environments—making processes visible, preserving material and social anchors, rewarding virtue, and stabilizing constitutive/complementary roles—while managing, rather than denying, the tensions between enhancement and deactivation. Ultimately, education bears ethical responsibility for stewarding cognitive capacities in an era when core intellectual functions can be rapidly externalized and substituted; abdicating this design task to platforms invites dependency, whereas acting with knowledge, wisdom, institutional courage, and a commitment to epistemic flourishing enables schools and universities to guide technological change. Therefore, AI guidance programs and related teaching practices need to be incorporated into both school and university curricula. At this juncture, one path leads to outsourcing and erosion of agency; the other, to designed entanglements in which technologies become partners in intellectual growth. By preserving human agency, privileging constitutive and complementary roles, cultivating virtues, embedding AI in rich medial ecologies, and securing transparency, accessibility, and ethical design, educational systems can shape how we think with and through machines and thus co-author the next chapter in the history of human culture. At the higher education level, this may involve developing faculty training initiatives on AI-mediated pedagogy and integrating AI literacy into degree requirements across disciplines.

10 Future Research Directions

While this paper provides a theoretical framework grounded in postphenomenology, EMT, and virtue epistemology, future empirical research should test these concepts across diverse educational contexts. A research approach combining mixed methods would integrate phenomenological depth with quantitative measures, employing phenomenological interviews to capture students' lived experience of AI-mediated learning, classroom observations documenting pedagogical patterns, analysis of student–AI interaction logs and revision trails, and validated instruments measuring epistemic virtues such as adapted versions of the Intellectual Humility Scale [44]. Longitudinal designs that repeatedly observe the same variables or individuals over an extended period (e.g., the development of undergraduate students across a three-year program, or tracking their entry-level characteristics across different academic years) may help track changes, trends, and developmental patterns. This would be particularly valuable for identifying threshold effects where AI use transitions from scaffolding to substitution, and for examining whether early formation of virtuous AI habits produces lasting epistemic benefits. Such research would also address critical questions: Do certain institutional policies (e.g., mandatory prompt logging, oral defense requirements) consistently stabilize complementary over substitutive uses? How do educator practices shape classroom norms? Finally, implementation research examining how European and national frameworks (EU AI Act, GDPR, education standards) translate into institutional practices would provide crucial guidance for policymakers. This empirical agenda would transform the philosophical framework offered here into actionable, evidence-based guidance for higher education institutions.

11 Conclusions

This paper proposes a framework for analyzing and informing the design of LLM use in education that integrates postphenomenology, the extended mind thesis, and Fasoli's taxonomy of cognitive artifacts. Its central claim is that the constitutive/complementary/substitutive distinction is not a fixed property of any technology but a relational outcome of how learning environments are designed. This reframes the pedagogical task: not managing a tool, but configuring a cognitive ecology. Without deliberate design, use tends to drift toward substitution, encouraging cognitive diminishment [72]. However, well-structured practices can stabilize more productive engagements. The paper outlined four pedagogical patterns: feedback amplification, accessibility without displacement, dialogic rehearsal, and triangulation, illustrated by a scenario of AI-mediated philosophy teaching. This reconfiguration also demands a rethinking of assessment. When AI participates in learning, what matters is not only the final product but the process: how students prompt, contest, and integrate its outputs, and what strategies they employ to construct and produce epistemically grounded knowledge. At the same time, education must actively cultivate the epistemic virtues that sustain critical engagement rather than passive consumption. Finally, technological neutrality is not an option. Policies, curricula, and assessments actively shape how cognition is distributed between humans and machines. Institutions must therefore set the conditions of integration: ensuring transparency, protecting spaces for independent reasoning, and promoting access without dependency. LLMs do not simply add another tool to the classroom; they reshape how intelligence itself is organized. The key issue is no longer whether machines can think, but how we will learn to think with and through them—and whether education can guide this entanglement toward the expansion rather than the erosion of human agency and knowledge.

Acknowledgments. The author received no specific funding for this research and has no additional individuals to acknowledge. The author used an LLM for copy-editing assistance in the preparation of this manuscript. All intellectual content, arguments, and conclusions are the author's own.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Alsberg, P.: In Quest of Man: A Biological Approach to the Problem of Man's Place in Nature. Pergamon, Oxford (1970)
2. Anders, G.: Die Antiquiertheit des Menschen, vol. 1. C.H. Beck, München (1956)
3. Baehr, J.: The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology. Oxford University Press, Oxford (2011)
4. Baehr, J.: Educating for intellectual virtues: from theory to practice. J. Philos. Educ. **47**(2), 248–262 (2013)
5. Baidoo-Anu, D., Owusu Ansah, E.: Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. J. AI. **7**(1), 52–62 (2023). https://doi.org/10.61969/jai.1337500
6. Beale, R.: The revolution has arrived: what the current state of large language models in education implies for the future. arXiv:2507.02180 (2025). https://doi.org/10.48550/arXiv.2507.02180
7. Bender, E.M., Gebru, T., et al.: On the Dangers of Stochastic Parrots: Can Language Models be Too Big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 610–623. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3442188.3445922
8. Boden, M.A.: Creativity in a nutshell. Think. **5**(15), 83–96 (2007)
9. Boden, M.A.: Artificial Intelligence: A Very Short Introduction. Oxford University Press, Oxford (2018)
10. Cassinadri, G.: ChatGPT and the technology–education tension: applying contextual virtue epistemology to a cognitive artifact. Philos. Technol. **37**(14) (2024). https://doi.org/10.1007/s13347-024-00701-7
11. Christian, B.: The Alignment Problem. Machine Learning and Human Values. W. W. Norton & Company, New York (2020)
12. Chrzanowski, J.: The changing memory: how internet search engines inhibit the creation of new ideas. J. Student Res. **11** (2022). https://doi.org/10.47611/jsr.v11i2.1628

13. Ciasullo, A.: Il futuro dell'apprendimento. Progettazione bioeducativa e intelligenze artificiali. Mondadori, Milano (2024)
14. Clark, A.: Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence. Oxford University Press, Oxford (2003)
15. Clark, A.: The Experience Machine: how our Minds Predict and Shape Reality. Pantheon Books, New York (2023)
16. Clark, A.: Extending minds with generative AI. Nat. Commun. **16**, 4627 (2025). https://doi.org/10.1038/s41467-025-59906-9
17. Clark, A., Chalmers, D.J.: The extended mind. Analysis **58**(1), 7–19 (1998)
18. Cotton, D.R.E., Cotton, P.A., Shipway, J.R.: Chatting and cheating: ensuring academic integrity in the era of ChatGPT. Innov. Educ. Teach. Int. **61**(2), 228–239 (2024). https://doi.org/10.1080/14703297.2023.2190148
19. Dahmani, L., Bohbot, V.D.: Habitual use of GPS negatively impacts spatial memory during self-guided navigation. Sci. Rep. **10**, 6310 (2020). https://doi.org/10.1038/s41598-020-62877-0
20. Dai, W. et al.: Can large language models provide feedback to students? A case study on ChatGPT. In: 2023 IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 323–325 (2023)
21. de Kerckhove, D.: L'uomo quantistico: Mente, società, democrazia: dove ci porterà la prossima rivoluzione digitale. RAI Libri, Roma (2025)
22. Deng, J., Lin, Y.: The benefits and challenges of ChatGPT: an overview. Front. Comput. Intell. Syst. **2**, 81–83 (2023). https://doi.org/10.54097/fcis.v2i2.4465
23. Dewey, J.: Democracy and Education. Macmillan, New York (1916)
24. Dewey, J.: Experience and Education. Macmillan, New York (1938)
25. Eisenstein, E.L.: The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe. Cambridge University Press, Cambridge (1979)
26. Extance, A.: ChatGPT has entered the classroom: how LLMs could transform education. Nature. **623**(7987), 474–477 (2023). https://doi.org/10.1038/d41586-023-03507-3
27. Fakour, H., Imani, M.: Socratic wisdom in the age of AI: A comparative study of ChatGPT and human tutors in enhancing critical thinking skills. Front. Educ. **10** (2025). https://doi.org/10.3389/feduc.2025.1528603
28. Fasoli, M.: Substitutive, complementary and constitutive cognitive artifacts: developing an interaction-centered approach. Rev. Philos. Psychol. **9**(3), 671–687 (2018)
29. Fasoli, M., Cassinadri, G., Ienca, M.: The dark side of cognitive enhancement: A framework for the technologically induced cognitive diminishment. J. Cogn. Enhanc. (2025). https://doi.org/10.1007/s41465-025-00331-7
30. Febvre, L., Martin, H.-J.: The Coming of the Book: The Impact of Printing, 1450-1800. Trans. D. Gerard. Verso, London (1976)
31. Gazzaley, A., Rosen, L.D.: The Distracted Mind. Ancient Brains in a High Tech World. MIT Press, Cambridge (2016)
32. Gehlen, A.: Man, His Nature and Place in the World. (Transl. C. McMillan, K. Pillemer). Columbia University Press, New York (1988)
33. Gibney, E.: What Are the Best AI Tools for Research? Nature's Guide, Nature, February 17 (2025). https://doi.org/10.1038/d41586-025-00437-0
34. Hasse, C.: Material hermeneutics as cultural learning: from relations to processes of relations. AI & Society **38**(5), 2037–2044 (2023). https://doi.org/10.1007/s00146-021-01171-7
35. Havelock, E.A.: The Literate Revolution in Greece and its Cultural Consequences. Princeton University Press, Princeton (1982)
36. Hongladarom, S., van der Vaeren, A.: ChatGPT, Postphenomenology, and the human-technology-reality relations. J. Human-Technol. Relat. **2** (2024). https://doi.org/10.59490/jhtr.2024.2.7386

37. Ihde, D.: Technology and the Lifeworld: from Garden to Earth. Indiana University Press, Bloomington (1990)
38. Ihde, D.: Experimental Phenomenology: Multistabilities, 2nd edn. SUNY Press, Albany (2012)
39. Ihde, D., Malafouris, L.: Homo faber revisited: Postphenomenology and material engagement theory. Philos. Technol. **32**, 195–214 (2019). https://doi.org/10.1007/s13347-018-0321-7
40. Ihde, D.: Material Hermeneutics: Reversing the Linguistic Turn. Routledge, London (2022)
41. Kasneci, E., et al.: ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ. Differ. **103**, 102274 (2023). https://doi.org/10.1016/j.lindif.2023.102274
42. Kapp, E.: Grundlinien einer Philosophie der Technik: Zur Entstehungsgeschichte der Kultur aus neuen Gesichtspunkten. G. Westermann, Braunschweig (1877)
43. Kosmyna, N., et al.: Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv preprint. https://doi.org/10.48550/arXiv.2506.08872 (2025)
44. Krumrei-Mancuso, E.J., Rouse, S.V.: The development and validation of the comprehensive intellectual humility scale. J. Pers. Assess. **98**(2), 209–221 (2016). https://doi.org/10.1080/00223891.2015.1068174
45. Laaksoharju, M., Lennerfors, T.T., Persson, A., Oestreicher, L.: What is the problem to which AI chatbots are the solution? AI ethics through Don Ihde's embodiment, hermeneutic, alterity, and background relationships. In: Lennerfors, T.T., Murata, K. (eds.) Ethics and Sustainability in Digital Cultures, pp. 31–48. Taylor & Francis, Milton Park (2023)
46. Labadze, L., Grigolia, M., Machaidze, L.: Role of AI chatbots in education: systematic literature review. Int. J. Educ. Technol. High. Educ. **20**, 56 (2023). https://doi.org/10.1186/s41239-023-00426-1
47. Lemasters, R., Hurshman, C.: A shift towards oration: teaching philosophy in the age of large language models. AI and Ethics **5**, 1203–1215 (2025)
48. Leroi-Gourhan, A.: Gesture and Speech. MIT Press, Cambridge, MA (1993)
49. Lin, Z.: How to write effective prompts for large language models. Nat. Hum. Behav. **8**, 611–615 (2024). https://doi.org/10.1038/s41562-024-01847-2
50. Longpre, S., Mahari, R., Chen, A., et al.: A large-scale audit of dataset licensing and attribution in AI. Nat. Mach. Intell. **6**, 975–987 (2024). https://doi.org/10.1038/s42256-024-00878-8
51. Lucchi, N.: ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. Eur. J. Risk Regul. **15**(3), 602–624 (2024)
52. Lyu, W., Wang, Y., Chung, T., Sun, Y., Zhang, Y.: Evaluating the Effectiveness of LLMs in Introductory Computer Science Education: A Semester-Long Field Study, arXiv:2404.13414v3 (2024)
53. McLuhan, M.: Understanding Media: the Extensions of Man. McGraw-Hill, New York (1964)
54. McLuhan, M.: The Gutenberg Galaxy: the Making of Typographic Man. University of Toronto Press, Toronto (1962)
55. Merleau-Ponty, M.: Phenomenology of Perception. Routledge, New York (1962)
56. Mollick, E.: The value of OpenAI and ChatGPT for the current learning environments and the potential future uses. SSRN Electron. J. (2023). https://doi.org/10.2139/ssrn.4439267
57. Mykhailov, D.: The phenomenological roots of technological intentionality: A postphenomenological perspective. Front. Philos. China. **15**(4), 612–635 (2020). https://doi.org/10.3868/s030-009-020-0035-6
58. Mykhailov, D., Liberati, N.: Back to the technologies themselves: phenomenological turn within postphenomenology. Phenomenol. Cogn. Sci. **24**, 1295–1314 (2025). https://doi.org/10.1007/s11097-023-09905-2
59. Norman, D.A.: Cognitive artifacts. In: Carroll, J.M. (ed.) Designing Interaction: Psychology at the Human-Computer Interface, pp. 17–38. Cambridge University Press, Cambridge (1991)

60. Noy, S., Zhang, W.: Experimental evidence on the productivity effects of generative artificial intelligence. Science. **381**, 187–192 (2023)
61. Ou, Y., et al.: How do LLMs acquire new knowledge? A knowledge circuits perspective on continual pre-training. In: Findings of the Association for Computational Linguistics: ACL 2025, Association for Computational Linguistics, Vienna, pp. 19889–19913 (2025). https://doi.org/10.18653/v1/2025.findings-acl.1021
62. Pasquinelli, M.: The Eye of the Master: A Social History of Artificial Intelligence. Verso, New York (2023)
63. Pritchard, D.: Intellectual virtue, extended cognition, and the epistemology of education. In: Baehr, J. (ed.) Intellectual Virtues and Education: Essays in Applied Virtue Epistemology, pp. 113–127. Routledge, New York (2015). https://doi.org/10.4324/9781315714127-7
64. Rahman, M.M., Watanobe, Y.: ChatGPT for education and research: opportunities, threats, and strategies. Appl. Sci. **13**, 5783 (2023). https://doi.org/10.3390/app13095783
65. Redaelli, R.: Intentionality gap and Preter-intentionality in generative artificial intelligence. AI & Society **40**, 2525–2532 (2025). https://doi.org/10.1007/s00146-024-02007-w
66. Rivoltella, P.C.: Disput IAmo. Sull'uso dei LLM per la didattica dell'argomentazione. In: Ferrari, M., Nardi, M. (eds.) Educare a Pensare, Carocci, Roma (2025)
67. Salathé, M.: Generative Biology: How Evo 2 Is Bringing Generative AI to the Code of Life, Substack blog post. https://engineeringprompts.substack.com/p/generative-biology (2025)
68. Scarlatos, A., et al.: Training LLM-based Tutors to Improve Student Learning Outcomes in Dialogues, arXiv:2503.06424v1 (2025)
69. Si, C., Yang, D., Hashimoto, T.: Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers, https://arxiv.org/abs/2409.04109 (2024)
70. Sok, S., Heng, K.: ChatGPT for Education and Research: A Review of Benefits and Risks. SSRN https://doi.org/10.2139/ssrn.4378735 (2023)
71. Sparrow, B., Liu, J., Wegner, D.M.: Google effects on memory: cognitive consequences of having information at our fingertips. Science. **333**(6043), 776–778 (2011). https://doi.org/10.1126/science.1207745
72. Stadler, M., Bannert, M., Sailer, M.: Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. Comput. Hum. Behav. **160**, 108386 (2024). https://doi.org/10.1016/j.chb.2024.108386
73. Trafton, A.: Artificial Intelligence Yields New Antibiotic: A Deep-Learning Model Identifies a Powerful New Drug That Can Kill Many Species of Antibiotic-Resistant Bacteria, MIT News (2020). https://news.mit.edu/2020/artificial-intelligence-identifies-new-antibiotic-0220
74. Turing, A.M.: Computing machinery and intelligence. Mind. **59**, 433–460 (1950)
75. Valiant, L.: The Importance of Being Educable: A New Theory of Human Uniqueness. Princeton University Press, Princeton (2024)
76. Vallor, S.: Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press, Oxford (2016)
77. Verbeek, P.P.: Cyborg intentionality: rethinking the phenomenology of human–technology relations. Phenomenol. Cogn. Sci. **7**(3), 387–395 (2008)
78. Vygotsky, L.S.: Mind in Society: the Development of Higher Mental Processes. Harvard University Press, Cambridge, MA (1978)
79. Wadden, J.J.: The postphenomenological impact of conversational artificial intelligence on autonomy and psychological integrity. Am. J. Bioeth. **23**(5), 37–40 (2023)
80. Wolf, M.: Reader, Come Home: the Reading Brain in a Digital World. HarperCollins, New York (2018)
81. Yan, L., Greiff, S., Teuber, Z., et al.: Promises and challenges of generative artificial intelligence for human learning. Nat. Hum. Behav. **8**, 1839–1850 (2024). https://doi.org/10.1038/s41562-024-02004-5

82. Yiu, E., Kosoy, E., Gopnik, A., et al.: Transmission versus truth, imitation versus innovation: what children can do that large language and language-and-vision models cannot (yet). Perspect. Psychol. Sci. **19**(5), 874–883 (2024). https://doi.org/10.1177/17456916231201401
83. Zhai, X.: ChatGPT and AI: The Game Changer for Education. ResearchGate Preprint https://doi.org/10.13140/RG.2.2.31107.37923 (2023)
84. Zhai, X.: ChatGPT user experience: implications for education. SSRN preprint (2022). https://doi.org/10.2139/ssrn.4312418
85. Zuboff, S.: The Age of Surveillance Capitalism. Profile Books, London (2019)