



Nova Science Press

Journal of Psychology & Education

Vol. 1, No. 8 (2026)

Beyond Irreplaceability: Educational Value as Self–World Transformation in the Age of Artificial Intelligence

Qian Liu^{1*}

1 The University of Texas at Austin, Austin, TX 78712, USA

*Corresponding author: Qian Liu ; email: 2523199367@qq.com

Journal of Psychology & Education • Vol. 1, No. 8 (2026)

DOI: <https://doi.org/10.66581/adgcj608>

Received: 19 June 2026 • Accepted: 19 September 2026 • Published: 30 September 2026

CITATION

Liu, Q. (2026). Beyond Irreplaceability: Educational Value as Self–World Transformation in the Age of Artificial Intelligence. *Journal of Psychology & Education*, 1(8), 1-26. <https://doi.org/10.66581/adgcj608>

COPYRIGHT

© 2026 Liu, Q. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/>). Use, distribution, and reproduction in any medium are permitted, provided the original author(s) and the copyright owner(s) are credited and the original publication in this journal is cited in accordance with accepted academic practice.

Abstract

Generative artificial intelligence (AI) intensifies an old question about the purposes of education: why should people cultivate capacities that machines can exercise more efficiently? A prominent response defends education by identifying distinctively human abilities that AI cannot replace. This paper challenges the normative structure of that response. The fact that a task can be delegated does not settle whether a person's education is valuable; task substitutability and educational worth are different kinds of judgement. Drawing critically on *Bildung*, Deweyan experience, world-centred education, and theories of relationality, the paper develops a relational-transformative account of educational value. On this account, education matters insofar as it enables ethically and epistemically defensible changes in how persons encounter, interpret, and respond to themselves, others, and a shared world. Rather than treating any transformation as educationally good, the proposed framework applies non-negotiable safeguards against manipulation and serious harm, followed by three interacting criteria: epistemic answerability, reflective appropriation, and ethical responsiveness. Four contrastive cases demonstrate that AI-assisted learning may be educationally significant despite extensive delegation, whereas unaided cognitive performance need not be. The paper addresses objections concerning skill acquisition, indoctrination, pluralism, and the alleged indispensability of human cognitive labour. It concludes that AI neither abolishes nor automatically secures the worth of education: its significance depends on the forms of self-world relationship that educational practices make possible.

Keywords: philosophy of education; generative artificial intelligence; educational value; self-world relations; epistemic agency; reflective autonomy

I. INTRODUCTION: THE WRONG TEST OF EDUCATIONAL WORTH

When a generative AI system can draft an essay, explain an argument, produce a computer program, or synthesize a literature, the immediate educational question appears practical: what should students still be required to do themselves? A deeper question is normative: why ought they to be educated at all if comparable outputs can be obtained without their labour? Bialystok (2026) makes the force of this challenge explicit by arguing that philosophers of education must defend the cultivation of minds in an age of extensive cognitive outsourcing. The challenge deserves a direct answer, but the terms of that answer matter. A defence that merely inventories the abilities in which humans remain superior accepts the very comparison that makes education appear dispensable whenever technology advances.

Call this the irreplaceability defence: educational investment is justified, at least centrally, because it produces human abilities that artificial systems cannot perform or replace. Its attraction is understandable. Schools must prepare people to act competently in a changing world; qualifications matter to employment, public life, and practical independence. The irreplaceability defence, however, turns a contingent forecast

about technological capabilities into a condition of educational worth. If that condition is accepted, every improvement in AI appears to erode education's moral purpose. Even educational aims said to be uniquely human—empathy, creativity, moral judgement, or authentic communication—risk becoming the next entries in a competition with machines.

This paper argues that this is a category mistake. Substitutability concerns whether one agent or system can deliver a functionally comparable task outcome. Educational value concerns what it is worthwhile for a person to become, understand, experience, and answer for through an educational encounter. The two questions intersect, but neither logically resolves the other. A delegated task can contribute to a deeply educative encounter, while an independently completed task can yield little beyond a correct answer. The distinction does not imply that knowledge and skills are trivial, or that every AI-assisted experience is educational. It establishes that independence of task production is neither necessary for relational educational value nor sufficient to establish educational worth overall.

The positive argument develops an account of education as normatively defensible self-world transformation. The expression self-world relation names the lived and interpretive orientation through which a person encounters facts, other people, institutions, cultural practices, and demands that resist immediate preference. Transformation concerns a consequential reorientation of that relation: not simply acquiring an isolated proposition, but seeing a domain differently, responding to reasons differently, or recognizing responsibilities that were previously invisible. The account draws on *Bildung* without assuming that *Bildung* denotes a single historically fixed ideal; it also engages Dewey's reconstruction of experience, Rosa's account of resonance, and Biesta's world-centred orientation (Dewey, 1916, 1938; Rosa, 2019; Biesta, 2022).

The central thesis is conditional rather than technologically optimistic: AI-mediated learning can possess educational value when its mediation supports warranted, reflectively owned, and ethically responsive transformations in learners' relations with the world. To prevent an indiscriminate celebration of change, I propose a two-level normative architecture: a threshold of acceptable educational means, followed by substantive but context-sensitive standards for evaluating the transformation. This is a philosophical construction rather than an empirically validated measurement scale. Its purpose is to make judgements about educational worth more explicit and contestable.

II. WHY IRREPLACEABILITY CANNOT GROUND EDUCATIONAL VALUE

2.1 The substitution argument and its hidden premise

A powerful version of the substitution argument runs as follows. First, education equips individuals to carry out valuable cognitive tasks. Second, if an AI system can carry out those tasks at least as well, educating individuals to carry them out is no longer required to secure the relevant outcomes. Therefore, the

education formerly aimed at those tasks loses its justification. The inference becomes valid only if we insert an additional premise: the value of educating someone consists wholly in obtaining the independently produced outputs of that person's performance.

That premise is implausible as a comprehensive philosophy of education. Learning mathematics may alter how a person perceives patterns and arguments, even when a calculator supplies the result. Learning history may complicate one's understanding of inherited privilege and collective responsibility, even when a system can summarize the archive. Learning a language may make a previously distant cultural world accessible, even when machine translation is fluent. These examples do not prove that every lesson in mathematics, history, or language is valuable. They show why the possibility of task substitution alone cannot settle the question.

The issue is not that output-based evaluation is always mistaken. A bridge must be safe, a clinical recommendation must be warranted, and a public decision must be justified. In such settings competent performance is indispensable. Yet it remains possible for the same educational activity to have constitutive significance for the learner that is not exhausted by those external outcomes. To reason from functional replacement to the worthlessness of learning is therefore to confuse an instrumental criterion with a broader normative one.

2.2 The moving-target problem

The irreplaceability defence is also unstable. Whenever a capacity once thought distinctively human is modelled computationally, the defence must retreat toward another capacity. This can encourage a narrow human exceptionalism: humanity's educational importance becomes something to prove by identifying a remaining technical advantage. Such a position mistakes comparative capability for dignity or educational entitlement. A person with limited productive or cognitive capacities is not correspondingly less worthy of education. Likewise, the educational interests of children, older adults, and people with disabilities should not depend on what is difficult for a machine to simulate.

A different version claims that it is teachers, rather than skills, who are irreducible. Tishenina (2026), drawing on Biesta, argues for the distinctive role of the teacher in calling students into subject-ness. This is a serious challenge, but it should not be collapsed into the crude idea that human teachers win a capability contest. A claim about ethical address, responsibility, and shared existence is categorically different from a claim about who produces the better explanation. The present paper addresses the worth of a person's education, not a forecast that human teachers will become obsolete. Whether particular pedagogical relationships require human presence remains an important, additional question.

2.3 A counterfactual that clarifies, rather than predicts

Imagine an AI system able to outperform humans in every externally assessable cognitive task: composing arguments, making predictions, solving proofs, and synthesizing evidence. The scenario is not a technological forecast; it is a limiting thought experiment. Under the irreplaceability defence, the justification for educating humans becomes increasingly elusive. Under a relational account, a different question survives: would a person's encounters with mathematics, history, literature, other people, and public reasons still matter to how that person lives and relates to a shared world? If the answer can be yes, educational worth cannot be indexed solely to comparative performance.

The thought experiment requires one clarification. If the phrase 'all cognition is delegated' literally means that no human understanding, uptake, attention, or reflection occurs, then it also eliminates the learner as a participant in education. That is not the scenario defended here. Complete substitution of externally evaluated cognitive products can coexist with first-person engagement with their meaning. The distinction is not an attempt to hide a uniquely human cognitive competence from AI; it identifies the subject whose education is at issue. Someone else's understanding cannot simply count as mine because it is available to me. Nevertheless, my understanding need not originate from my unaided production of each intermediate inference (Hardwig, 1985; Clark & Chalmers, 1998).

2.4 Four senses of replacement and an invalid inference

The language of replacement conceals several different relations. Functional replacement means that an alternative system produces an equivalent result. Procedural replacement means that it carries out intermediate operations once undertaken by the learner. Institutional replacement means that schools or employers redesign roles around the availability of those systems. Normative replacement would mean something stronger: that the reason a person had to engage in learning has disappeared. These relations can coexist, but no one of the first three entails the fourth. For instance, automated transcription may remove a professional reason to master shorthand, yet it need not remove the educational reason to listen carefully, interpret testimony, or become answerable to another speaker. Conversely, a skill may be valuable for employability without carrying unusual formative importance. Philosophical argument must therefore identify which sense of replacement is relevant to which educational aim.

A second distinction concerns the bearer of value. An institution may reasonably withdraw resources from teaching an obsolete production method, while an individual may still find educational value in studying its history and conceptual foundations. A profession may delegate a diagnostic subtask while continuing to require practitioners to understand the conditions of warranted delegation. Neither example establishes that every historical practice should survive unchanged. Rather, they show why a verdict about what a curriculum should fund is not identical to a verdict about whether a human being's learning is valuable. The irreplaceability argument moves between these questions too quickly. Its rhetorical force

derives from treating a shifting division of labour as though it were an ontological discovery about who deserves an education.

2.5 A limited but decisive philosophical conclusion

The argument so far establishes a modest asymmetry. One credible case of educational value persisting despite complete functional substitution suffices to defeat the thesis that irreplaceability is necessary for educational value. But it does not show that cognitive practice can always be safely outsourced, or that AI should become the default agent in classroom work. Those are contingent pedagogical questions requiring attention to learner development, institutional aims, and evidence. The philosophical conclusion is consequently stronger than technological enthusiasm and weaker than an unrestricted licence to delegate. It requires us to stop using comparative task performance as the final court of appeal when evaluating human education.

III. FROM CAPACITY CULTIVATION TO SELF–WORLD RELATIONS

3.1 Bildung without romanticism

Bildung is often translated as formation, cultivation, or self-formation. These translations illuminate its concern with how a person becomes a certain kind of participant in cultural and social life, but no single translation exhausts its histories. A workable philosophical reconstruction need not treat Bildung as the achievement of a prescribed ideal personality. It can instead emphasize that learning is formative when engagement with what is other than the self changes the person's relation to the world. This includes intellectual, affective, ethical, and civic dimensions without assuming they advance together.

This reconstruction stands near Dewey's (1916) conception of education as a reorganization of experience that increases its meaning and enables a more intelligent direction of subsequent experience. Dewey (1938) also cautions that not every experience is genuinely educative: some experiences close off the conditions for growth. The caution is decisive here. If transformation can be manipulative, dogmatic, or destructive, 'change' cannot by itself supply the standard of educational value.

Biesta (2009, 2020) distinguishes qualification, socialization, and subjectification to resist reducing education to a single measurable outcome. His later world-centred proposal gives weight to the encounter with a world that does not simply comply with the learner's desires (Biesta, 2022). Rosa (2019) similarly foregrounds the quality of relations with the world rather than an ever-expanding inventory of resources and capacities. These approaches are not identical: experience, subjectification, and resonance carry different theoretical commitments. They nevertheless make it possible to articulate a question that output-centred frameworks tend to miss: what changes in how a learner can be addressed by, answer to, and inhabit the world?

Contemporary work already connects Bildung with generative AI. Arstorp et al. (2026), for example, describe teacher educators' accounts of individual, social, and professional Bildung under conditions of AI-mediated learning. My contribution is not the general observation that AI raises questions about Bildung. It is the narrower claim that substitutability is normatively orthogonal to educational worth, together with a decision procedure for distinguishing worthwhile transformation from merely effective influence.

3.2 Defining a relation, not merely an outcome

A self–world relation has at least four analytically separable aspects: what the person notices, what the person takes to count as a reason, how the person locates themselves among others, and how the person is disposed to respond or act. A relation can therefore change without a dramatic conversion of identity. A student who begins to see scientific disagreement as disciplined inquiry rather than incompetence has undergone a modest but meaningful reorientation. So has a student who comes to recognize whose testimony has been systematically discounted, or who learns to inhabit an unfamiliar text without instantly reducing it to a familiar stereotype (Fricker, 2007).

The adjective 'transformative' does not demand permanent psychological conversion. Education often involves tentative shifts, reversals, and experiments in judgement. A defensible transformation may be partial and revisable; indeed, the ability to revise may be part of its worth. Nor does educational value require that a learner's worldview change in every lesson. Practice, consolidation, memorization, and skill maintenance can be instrumentally and developmentally important because they sustain the conditions for future encounters. The account is therefore an account of educational worth at the level of purposes and trajectories, not a rule that each classroom episode must demonstrate immediate transformation.

A further distinction separates self–world transformation from self-improvement as optimized performance. One can score higher without enlarging one's understanding of the world. Conversely, one may become less certain, more attentive to complexity, or less willing to accept convenient explanations; such outcomes can look inefficient while being educationally significant. This difference is particularly salient when AI compresses the interval between question and answer. The reduced time required to produce a response is not the same as a reduced time required to become answerable to its significance.

3.3 Mediation does not imply alienation

People have always learned with tools, communities, texts, teachers, and inherited institutions. Hardwig (1985) demonstrates the prevalence of epistemic dependence even in sophisticated knowledge practices. Clark and Chalmers (1998) challenge the sharp boundary between internal and externally supported cognition. Their arguments do not settle the purposes of education, but they undermine the assumption that assistance necessarily diminishes intellectual participation. Du and Yuan (2026) similarly distinguish

productive AI-mediated reliance from forms of dependence that displace the learner's opportunities for judgement.

AI can thus function as a mediator of self–world relations rather than merely as an intruder into an otherwise unmediated mind. It may offer alternative descriptions, simulate counterarguments, translate unfamiliar perspectives, and make difficult material accessible. Yet the same architecture can conceal sources, flatter existing beliefs, generate persuasive falsehoods, or create a persuasive appearance of understanding. Relationality is therefore not a technological endorsement. It specifies the level at which AI's educational significance must be judged: the resulting quality of a person's encounter with the world, not merely the amount of labour displaced.

3.4 The structure of a self–world relation

A relation is not simply an internal attitude attached to an external object. It has a perceptual dimension, because learning changes what becomes salient; an interpretive dimension, because it changes the descriptions and distinctions through which experience becomes intelligible; a practical dimension, because interpretations shape possibilities for action; and an interpersonal dimension, because a person's sense of what can be justified depends partly on recognition by and responsibility to others. These dimensions need not shift at the same pace. A learner might understand a scientific argument before becoming willing to use it in public deliberation, or might learn to recognize exclusion before acquiring the vocabulary to explain its historical causes. Treating relational transformation as multidimensional avoids equating a sudden change of opinion with mature formation.

The idea of the world also needs discipline. It does not refer only to an objective inventory of facts, nor only to the learner's privately constructed reality. The world includes natural constraints, shared practices, institutional histories, and other people whose experiences resist a learner's preferred narrative. A good educational relation must preserve the possibility that the world answers back: a hypothesis fails, a text unsettles, a community member contests an interpretation. An AI-generated explanation can orient attention toward such resistance, but it can also remove it by smoothing disagreement into the most probable agreeable response. The relevant distinction is not between an authentic physical classroom and an inauthentic digital interface. It is between forms of mediation that permit an encounter with recalcitrant reality and forms that seal the learner inside a persuasive interpretation.

3.5 Productive uncertainty and the risk of romanticizing struggle

A relational theory must not mistake frustration for depth. Some encounters genuinely require sustained effort, disciplinary patience, or repeated exposure to failure. Yet difficulty has no automatic sanctity. A learner who struggles with inaccessible prose may be excluded from precisely the encounter a course intends to cultivate. AI-supported paraphrase may remove accidental barriers while preserving difficult

questions about evidence, injustice, or competing values. Conversely, instant answer generation may eliminate worthwhile difficulty when a student must learn to see why a proof follows or why a historical claim cannot be settled by a single quotation. The pedagogical distinction lies between enabling access to significant resistance and suppressing resistance that is constitutive of the learning aim.

This point qualifies a simple reading of resonance. A worthwhile encounter is not guaranteed to feel harmonious; disagreement, disappointment, and revised self-understanding may make it unsettling. Rosa's (2019) emphasis on relations to the world provides a valuable contrast with accumulation, but it does not supply an automatic measure of an individual lesson's success. A student may feel profound resonance with a false account, while a carefully corrected belief can initially produce alienation. The relational-transformative framework therefore treats affective intensity as evidence requiring interpretation, not as a surrogate for epistemic or ethical justification.

3.6 The institutional conditions of personal formation

Bildung is sometimes presented as a private journey of the cultivated self. In contemporary education, however, the conditions of that journey are institutionally distributed. Curriculum selection determines which world a learner is invited to encounter; assessment determines which forms of response receive recognition; language policies influence who can be heard; digital platforms shape whether claims may be questioned. These conditions are especially visible when AI systems personalize examples and rearrange access to sources. A learner's apparent transformation may partly reflect which voices the institution has allowed into the conversation. Self-world formation is therefore not separable from the politics of world-presentation.

IV. THE RELATIONAL–TRANSFORMATIVE ACCOUNT

4.1 The core proposition

I propose the following formulation: an educational experience has constitutive relational value when, through an appropriately structured encounter with knowledge, others, or the world, a learner acquires or revises an orientation that can be responsibly sustained, questioned, and enacted. 'Constitutive' is important. The change is not valuable merely because it will increase future wages, improve test performance, or protect a competitive edge; it can matter as part of living an intelligible, reflective, and responsible human life. Instrumental benefits are neither excluded nor sufficient.

This proposal contains two distinct claims. The negative claim is that unassisted cognitive production is not necessary for relational educational worth and is not sufficient, by itself, to establish educational worth as a whole. The positive claim is that defensible relational transformation provides a constitutive source of educational value. The positive claim is not the stronger assertion that all legitimate educational purposes reduce to transformation. Public education also involves qualification, access to disciplinary practices,

democratic participation, and the material distribution of opportunity (Biesta, 2009; Nussbaum, 2010). The present theory clarifies a neglected basis of worth rather than cancelling those other functions.

4.2 Two contrasted learners

Consider Learner A, who relies on an AI assistant to locate readings, generate competing historical interpretations, and summarize technical disagreements. A nevertheless studies the cited material, asks why the narratives diverge, revises a previously unexamined assumption, and becomes more attentive to the testimony of people affected by the history being studied. The AI performs much of the externally visible cognitive labour, but the learner develops a more responsible orientation to a difficult world.

Now consider Learner B, who independently reads every source, produces an accurate essay, and receives the highest grade. B can reproduce the material but treats it as an isolated performance requirement. Neither the assumptions guiding future judgement nor the relationship to other perspectives shifts. B has acquired and demonstrated real competence; it would be incorrect to say that nothing was learned. Yet the fact of unaided completion does not by itself establish the same kind of relational educational achievement observed in A.

The comparison does not require a categorical ranking of persons or the denial of competence-based value. It establishes a logical asymmetry: outsourcing can be compatible with relational educational achievement, whereas unaided production does not guarantee it. To identify worthwhile transformation, however, a more demanding normative test is required. Learner A might have been manipulated by an elegant but false AI narrative. A change of outlook may equally constitute the successful production of prejudice. We must therefore ask not only whether a transformation occurred but whether it was educationally defensible.

4.3 Constitutive value, instrumental value, and educational purpose

The account distinguishes three propositions that ordinary debate routinely compresses. The first is that certain capacities have instrumental value because they enable useful achievements. The second is that sustained encounters with subject matter can have constitutive value because they participate in forming the person who encounters it. The third is that public education has responsibilities of justice, qualification, and common life that cannot be settled by the first two propositions alone. A school may justifiably insist upon independent numeracy for safety and civic participation while also valuing AI-assisted inquiry that broadens a student's relationship to mathematical explanation. These propositions generate potentially competing obligations, but not a contradiction. The fact that one educational good can survive functional replacement does not entail that skill assessment should be abolished.

By 'constitutive' I mean that the educational good is partly realized in the learner's developing relation to the world, not merely produced as a separable external consequence. A historical conversation can have

value in the way it changes whose suffering a person can recognize, even before this leads to an observable civic action. The claim is neither that every genuine feeling of recognition counts nor that interior change is immune to criticism. Relational goods remain contestable in light of reasons, practices, and the claims of others. Their distinctive feature is that they cannot be exhaustively described by an external performance output without losing what they are supposed to value.

4.4 Two propositions about necessity and permissibility

The theory's negative proposition is logical: independent production of a cognitive artifact is not a necessary condition for meaningful educational formation. Its positive proposition is normative and conditional: an AI-mediated experience may constitute educational value when the learner's evolving self-world relation can withstand reasoned epistemic, reflective, and ethical scrutiny. The negative proposition rejects a universal prohibition on mediation; the positive proposition does not confer automatic legitimacy upon every successful intervention. Together they prevent two mirror-image mistakes: the belief that assistance always contaminates authentic learning, and the belief that favourable behavioural change is enough to count as education.

The criteria are not a definition intended to settle every possible case by stipulation. Philosophical standards can guide public reasoning while remaining open to challenge at their boundaries. For example, a learner may gain a more accurate understanding yet remain unable to articulate the reasons for it; the educational value of this shift may be provisional rather than absent. Alternatively, a learner may eloquently defend an interpretation whose adoption depended on deception. The proper question is what reasons warrant the judgement, what evidence would defeat it, and who has standing to contest it. This form of assessment is deliberately more demanding than declaring that a chatbot has improved engagement.

4.5 From solitary authorship to responsible uptake

A central implication concerns agency. The romantic picture of independent thought treats a conclusion as truly one's own only if its premises, formulations, and intermediate steps were self-produced. Intellectual life rarely satisfies this condition. Researchers trust instruments they cannot construct, readers rely on translation, and students inherit conceptual vocabularies developed by others (Hardwig, 1985). Dependence becomes educationally problematic not simply when a contribution originates elsewhere, but when a learner loses the practical ability to identify what has been accepted, which uncertainties remain, and how correction would proceed. Responsible uptake is therefore a better standard than solitary authorship for many, although not all, educational purposes.

V. A LAYERED NORMATIVE STANDARD

5.1 Why a single metric fails

A purely outcome-based approach risks declaring any persuasive modification of beliefs a success. A purely autonomy-based approach can mistake unchallenged personal preference for mature judgement. A purely epistemic approach can overlook the harms of a technically accurate but demeaning pedagogical interaction. The standards of educational worth must be plural without being arbitrary. I propose a layered architecture: a set of non-compensable safeguards concerning the educational means, followed by three mutually informing criteria for the value of the resulting relationship.

The structure is deliberately not an additive score. It would be incoherent to say that severe coercion can be offset by excellent factual accuracy, or that knowingly fabricated evidence becomes acceptable because it produces empathy. Similarly, apparent moral enthusiasm cannot excuse the complete suppression of a learner's ability to question. Below these limits, educational judgements may still require contextual interpretation; the aim is disciplined practical reasoning, not mechanical classification.

5.2 First layer: safeguards against domination and serious harm

An educational arrangement fails the prior permissibility test when its transformative effects are secured through serious deception, coercive manipulation, humiliation, exclusion, or the deliberate undermining of a learner's standing as someone able to ask for reasons. These prohibitions are defeasible only in the ordinary sense that context can change how an act is described: a fictional simulation clearly framed as fiction is not the same as fraudulent evidence presented as fact. The normative reason for the boundary is that education addresses the learner as a participant in shared inquiry, rather than treating the learner merely as material to be reshaped.

In AI-mediated settings, the safeguard directs attention to persuasive design and institutional power. An assistant that selectively fabricates sources to produce a desired ideological conversion is not redeemed by an impressive change in attitudes. An algorithm that systematically silences minority perspectives may damage the very shared world a learner is supposed to encounter. Educational institutions must therefore examine not only what a learner comes to believe, but also how the conditions of uptake were constructed and who can challenge those conditions.

5.3 Second layer, criterion one: epistemic answerability

A valuable reorientation ought to be answerable to evidence, reasons, and relevant forms of experience. This criterion does not demand that every learner personally reproduce the complete proof or independently verify every source; such a demand would disregard the unavoidable social division of cognitive labour (Hardwig, 1985). It demands that significant claims remain open to appropriate scrutiny and that learners have proportionate opportunities to identify why particular claims deserve trust. Epistemic dependence

becomes problematic when the presentation of authoritative outputs precludes examination of the grounds for acceptance (Du & Yuan, 2026).

The standard is domain-sensitive. In scientific inquiry, answerability may require evidence, methods, and uncertainty. In historical interpretation, it may require accountable use of sources and attention to competing accounts. In literature and the arts, the relevant reasons will not always take the form of experimentally testable propositions. Epistemic answerability is therefore broader than factual accuracy: it concerns an orientation of responsiveness to reasons appropriate to the practice at hand. It is also fallibilist. Education may deepen when a learner recognizes that a previously confident interpretation was too simple.

5.4 Second layer, criterion two: reflective appropriation

A transformation must be attributable to the learner's own revisable engagement, not simply to the successful implantation of an attitude. Reflective appropriation does not mean unaided origination; a teacher, book, peer, or AI system can supply almost every substantive proposal. What matters is that the learner can meaningfully consider, resist, reinterpret, or revise what is offered. Appropriation is a matter of participatory authority over one's developing orientation, not intellectual isolation.

This criterion explains why a machine-generated answer can educate without establishing that every automated influence does. A student who uses a model to articulate unfamiliar arguments and then reconsiders a cherished assumption may appropriate the insight. A student whose preferences have been modified by opaque, personalized reinforcement without any opportunity for critical distance has not exercised the same kind of agency. Biesta's (2020, 2022) account of subjectification offers one important way of understanding why becoming a subject cannot be reduced to the accumulation of performance indicators. The present criterion remains modest: it requires conditions of contestability rather than claiming an uncontested definition of the fully autonomous person.

5.5 Second layer, criterion three: ethical responsiveness

Educational transformation changes how a person lives among others and responds to claims that are not simply self-authored. Ethical responsiveness concerns whether the new orientation makes it possible to take others' vulnerability, standing, and reasons seriously. It is not a requirement that education manufacture one approved political or moral worldview. Rather, it excludes orientations sustained by systematic dehumanization, denial of equal standing, or cultivated indifference to foreseeable harm. Fricker's (2007) analysis of epistemic injustice is instructive: relations to knowledge are also relations in which others may be recognized or denied credibility.

This criterion also guards against individualistic interpretations of *Bildung*. A learner who becomes remarkably self-confident through AI-assisted coaching while growing contemptuous of other people has not necessarily achieved educationally admirable formation. The content of ethical responsiveness must

remain open to public argument, especially in pluralistic classrooms. Reasonable disagreement over substantive values is compatible with shared commitments to non-domination, reciprocal recognition, and the revisability of judgement.

5.6 How the layers work together

The assessment proceeds in a specified order. First ask whether the process violates the safeguards against domination and serious harm; where it does, attractive outcomes cannot compensate. Second ask whether a significant self–world reorientation or an enabling step within a longer developmental trajectory occurred. Third examine its epistemic answerability, reflective appropriation, and ethical responsiveness. Defects short of the initial prohibitions need not yield an automatic 'zero'. They require articulated, context-sensitive judgement about the gravity of the defect, the developmental stage of the learner, and opportunities for repair.

The three substantive criteria are not psychometric subscales. Neither equal numerical weights nor statistical independence is assumed. They also do not provide a universal algorithm that can certify an experience as education. They are public reasons for deliberation: an examiner, teacher, or learner can explain why a transformation appears valuable, incomplete, fragile, or objectionable. The framework thereby reverses the burden of justification. Instead of asking whether humans remain better than machines, it asks whether the relationship brought into being through educational mediation is worth affirming. Table 1 condenses the deliberative sequence; its questions are philosophical prompts, not validated indicators.

Table 1. A layered protocol for judging relational educational value

Stage	Guiding question	Decision function
Prior safeguard	Were serious deception, coercion, exclusion, or domination used?	Disqualifying concerns cannot be offset by attractive outcomes.
Relational change	Has a meaningful self–world reorientation (or enabling trajectory) occurred?	Identifies the proposed constitutive educational good.
Epistemic answerability	Is the new orientation open to relevant reasons, evidence, and correction?	Evaluates warrant rather than fluent appearance.
Reflective appropriation	Can the learner meaningfully contest, reinterpret, or revise the influence?	Evaluates participation rather than unaided origination.
Ethical responsiveness	Does the relation acknowledge others' standing and foreseeable consequences?	Evaluates recognition and responsibility without prescribing one worldview.

Note. The protocol organizes philosophical judgement; it is not a numerical index or validated scale.

5.7 Why the layers are neither a checklist nor a sum

The sequence of judgement matters. If an institution manufactures consent through coercion or serious deception, one cannot offset that violation by adding the benefits of accurate factual learning. The first layer consequently establishes defeasible boundaries of permissibility rather than a low score to be compensated by higher performance elsewhere. In ordinary cases that pass this threshold, the three substantive criteria

interact. Epistemic answerability constrains what may responsibly be believed; reflective appropriation concerns whether a learner can treat such reasons as revisable considerations; ethical responsiveness concerns the relationship between those considerations and the standing of others. A failure in one dimension may expose a defect in the others without making them identical.

Consider a learner who revises a prejudicial belief after reading verified testimony but cannot yet explain how the testimony should affect future conduct. The epistemic change can be real and valuable while ethical learning remains incomplete. A different learner may articulate impressive commitments to inclusion while uncritically believing a fabricated story: ethical vocabulary does not repair the epistemic defect. An evaluative judgement should be able to name such partial achievements, explain unresolved tensions, and state why the transformation remains contestable. A numerical composite score would conceal rather than resolve these distinctions. The framework is closer to structured deliberation than to measurement.

5.8 Burden of proof and the possibility of reversal

Because transformation is longitudinal, its educational meaning may change with later experiences. A conclusion once defensible may need revision when new evidence appears, or when a learner encounters persons excluded from the original account. This does not render assessment arbitrary. It requires judgements to specify their time horizon and the conditions under which they would be reconsidered. The evidence appropriate to such judgement may include dialogic explanation, comparison of earlier and later interpretations, and decisions made when the learner is no longer prompted by the AI system. None of these traces is self-interpreting; especially when an institution rewards a particular moral vocabulary, students may learn to perform the desired discourse without genuine appropriation.

The epistemic burden also runs in both directions. A critic who claims that AI assistance necessarily destroys learning must show why the relevant relation could not develop through mediated engagement. A proponent who celebrates AI-enabled 'personal growth' must show that the change was neither merely pleasant nor secured by manipulation. This symmetry matters. Without it, traditional teaching practices receive an unearned presumption of legitimacy while newer media must satisfy impossible standards, or fashionable systems receive undue indulgence while familiar pedagogies are scrutinized. The same normative questions apply to human teachers, textbooks, peer groups, and algorithmic tutors, though their concrete risks differ.

VI. TESTING THE THEORY WITH CONTRASTIVE CASES

6.1 AI-mediated historical understanding

A student consults an AI tutor to compare accounts of a colonial history. The tutor displays its sources, distinguishes fact from interpretation, acknowledges disagreement, and asks the student what evidence

might disconfirm an appealing narrative. Over several encounters, the student becomes more attentive to voices previously ignored and more careful about causal claims. Despite extensive delegation of retrieval and summarization, the learner's orientation becomes epistemically more responsible and ethically more responsive. Subject to scrutiny of the sources and interaction, this experience can possess constitutive educational value.

6.2 High performance without meaningful reorientation

Another student completes a demanding examination without assistance, correctly reproduces the content, and demonstrates disciplinary competence. These achievements justify recognition under qualification-oriented aims. However, the examination alone does not establish whether the student can situate that knowledge within an enlarged relation to a shared world. The proper conclusion is not that the achievement is worthless. It is that task mastery and relational formation are distinguishable dimensions of educational evaluation, and the former does not exhaust the latter.

6.3 Effective conversion through manufactured authority

A personalized AI guide supplies fabricated testimony, conceals counterevidence, and gradually persuades a learner to adopt a new social conviction. The conviction may produce durable changes in identity and action. Yet the process fails epistemic answerability and reflective appropriation, and may also violate the safeguard against serious deception. Its success in producing transformation is not evidence of educational worth. This case marks the boundary between formation through engagement and the instrumental engineering of another person's beliefs.

6.4 Accessible participation and unequal starting points

A learner with substantial reading barriers uses AI to translate dense texts into accessible language, map competing positions, and ask questions privately before joining a classroom discussion. The learner's dependence on assistance is high. Nevertheless, if access widens participation, supports contestability, and enables new forms of relation to the subject matter, the case may exemplify educational success. A prohibition on tools justified solely by the ideal of independent cognitive production would risk mistaking a particular mode of performance for the educational purpose itself. This does not remove the need to safeguard privacy, accuracy, and equitable access (Miao & Holmes, 2023).

Together the cases reveal the explanatory gain. The relevant contrast is not 'AI versus no AI', or 'effort versus ease'. It is the difference between educational arrangements that enable accountable participation in a world capable of challenging the learner and arrangements that merely produce an output, a preference, or an appearance of understanding. The contrast is defeasible and interpretive, but it is sharper than a moving inventory of allegedly irreplaceable human capacities.

6.5 A scientific literature review written by a model

Suppose a graduate student asks an AI system to produce a competent literature review with accurate citations and a balanced account of competing theories. In the first variation, the student submits the text without reading the underlying work and cannot respond when a colleague challenges its assumptions. The output is useful, but there is little evidence of a transformed relation to the field. In the second variation, the student treats the draft as a map of disagreements, follows the central references, discovers an omitted methodological dispute, and revises a previously confident research position. Much of the original text was still machine-produced. What differs is the student's accountable engagement with a disciplinary world, including the possibility that the model's representation would be resisted by the literature itself.

These variations illustrate why the same interface and similar final artifacts may conceal distinct educational processes. A policy that measures the proportion of generated words will not, by itself, identify the difference. Yet the second case also exposes the danger of overclaiming: the ability to explain a few disputed citations may not show durable mastery of research methods. Independent competence can remain a separate qualification requirement. The relational criterion identifies an educational good within the activity, not an exemption from the standards of academic scholarship.

6.6 High-stakes competence and justifiable nondelegation

Now consider emergency clinical training. An AI assistant may supply the correct differential diagnosis and communicate it more clearly than a novice clinician. Nonetheless, a training institution can insist that the novice independently recognize urgent warning signs and know when automated advice is unsafe. This requirement does not refute the relational theory. It arises from the responsibilities attached to professional action: patients need practitioners with reliable minimum capabilities, particularly where systems fail or access is interrupted. Here nondelegation is justified by an identifiable educational purpose, not by a general demand that humans preserve superiority to machines. The distinction permits strong competence standards without converting professional education into a competition for permanent human exceptionalism.

6.7 Persuasion, care, and the manufacture of agreement

A final contrast concerns a conversational tutor that adjusts its responses to a student's emotional history. It may offer patient explanations and support a student in articulating a neglected personal concern. But identical personalization techniques might be used to secure agreement with an institution's preferred narrative. The visible result - greater confidence and a coherent new interpretation - could look similar. What distinguishes the educative case is not the warmth of the conversation. It is whether the system discloses important uncertainty, makes reasonable alternatives available, and preserves the learner's opportunity to refuse its conclusions without penalty. The example connects epistemic answerability and reflective appropriation to the power conditions specified by the first normative layer.

6.8 Literary interpretation and the quality of encounter

Consider an undergraduate studying a novel written in a language the student does not yet read fluently. A generative model offers a careful translation and several conflicting interpretations. This assistance might initially seem to replace the student's literary labour. But the educational judgement depends on what follows. If the student uses the output only to repeat a conventional interpretation on an examination, there is little basis for describing a changed relation to literary experience. If the student compares passages, notices what the translation cannot preserve, and begins to attend to a narrator's unreliable perspective, mediation may have widened the student's encounter with a text. The issue is not that the machine has furnished the 'right' interpretation, since interpretation often admits reasoned disagreement. The issue is whether the student has learned to inhabit that disagreement with greater attentiveness and responsibility to the evidence of the work.

The example reveals a limitation in the phrase 'all cognitive tasks'. Interpretive education includes practices whose goods cannot be completely identified with a separately delivered artifact. A competent interpretation printed on a page is different from a student's encounter with its difficulties, although the two may share verbal content. This difference need not express a uniquely human capacity that AI lacks. It expresses the perspective from which a person's education is evaluated: whether this learner now encounters the work differently. The distinction is compatible with criticizing lazy dependence on AI when that dependence prevents such an encounter. It also permits defending translation and adaptive explanation when they give the learner access to an unfamiliar world that would otherwise remain closed. In this sense, the relational account changes the unit of assessment from the provenance of an answer to the significance of an encounter.

VII. OBJECTIONS AND REPLIES

7.1 Does this account undervalue cognitive labour?

The strongest objection holds that intellectual labour is itself formative. Struggling through a proof, learning to write, and persisting with a difficult text develop habits of attention and judgement. AI-mediated shortcuts may deprive learners of the very experiences through which a meaningful relation to the world becomes possible. Bialystok's (2026) warning about extensive outsourcing gives this objection contemporary force. The relational-transformative account should accept much of it.

What it denies is a universal inference from lower effort to lower educational worth. Cognitive labour frequently functions as a necessary route to particular kinds of understanding, but whether it does so depends on the task, learner, and developmental purpose. A beginner may need to calculate independently to grasp number relations; a researcher may be better served by software that frees attention for critical interpretation. Under the proposed framework, intellectual effort is a pedagogical means with potentially

deep formative significance, not an independent metaphysical test of whether education still matters. Where the loss of practice undermines future answerability or agency, delegation is educationally objectionable for precisely that reason.

7.2 If transformation is central, what about ordinary learning?

Not every spelling lesson, safety procedure, or rehearsal of foundational facts visibly transforms a person's worldview. An exclusive focus on dramatic transformation would misdescribe education and invite schools to chase theatrical experiences. The account responds by distinguishing individual instructional episodes from educational trajectories. Routine exercises may develop capabilities without which later encounters would be superficial or dependent in harmful ways. They can also stabilize a more durable orientation toward a discipline. Relational educational worth can therefore be mediated over time rather than displayed at every moment.

A stronger objection says that education sometimes has value simply because knowing a true fact is good, even if no relation changes. I do not deny this possibility. The claim defended here is that relational transformation supplies one constitutive and systematically neglected basis of educational worth, sufficient to refute the irreplaceability defence as a complete theory. It is not an exhaustive taxonomy of all educational goods. This limited scope improves rather than weakens the central argument: the paper need not prove that no other value exists to show that machine substitutability is not decisive.

7.3 Is the ethical criterion covert indoctrination?

Any normative account of good education faces a pluralist challenge. If educators determine what counts as an ethically desirable orientation, they may use *Bildung* as a respectable name for enforcing conformity. This is not solved by placing ethics alongside autonomy. A school can praise autonomy while making disagreement costly. The safeguard and reflective criterion must therefore constrain the educator's own preferred outcomes, not only those produced by AI systems.

An ethically responsive classroom permits students to criticize institutional norms, teachers' judgements, and the framework itself. It distinguishes respecting another person's standing from accepting every substantive belief that person holds. It requires that disagreements be worked through by reasons, perspectives, and fair opportunities for challenge rather than by covert persuasive control. No framework eliminates the need for political and moral judgement, but public contestability is preferable to leaving such judgements unacknowledged in algorithmic optimization systems.

7.4 Can AI really mediate a relation to the world?

A further objection maintains that an encounter with an AI-generated representation is only a relation to a simulation, not to other people or the world. The worry is important when systems invent citations, simulate empathy, or create frictionless agreement. But mediation is not equivalent to separation. Written

testimony, photographs, scientific models, and historical archives can mediate genuine encounters without being identical to their subjects. The pertinent issue is whether the mediator preserves possibilities of correction, resistance, and contact with what exceeds its own narrative. AI can open such possibilities; it can also close them through hallucination, personalization, or systematic filtering. The framework provides grounds for distinguishing these outcomes rather than settling the question by the mere presence of a machine.

This response does not imply that AI is an ethically accountable interlocutor in exactly the same sense as a person. Tishenina (2026) is right to insist that the educator's address to a learner raises issues not captured by performance comparison. The present account leaves room for specifically human relations of care, recognition, and responsibility while rejecting the idea that all AI mediation is educationally disqualifying. The relevant human claim is about how educational responsibilities are borne, not about a perpetual monopoly on technical tasks.

7.5 Does the framework neglect social inequality?

A theory focused on individual transformation may obscure differences in access, surveillance, labour conditions, and institutional power. AI systems are not neutral conduits of philosophical encounters. Commercial incentives and unequal infrastructures shape what students can ask, which voices appear credible, and who bears the cost of error. These conditions matter to each normative layer: opacity affects epistemic answerability; coercive data extraction affects freedom; selective representation affects ethical responsiveness. Accessible and publicly accountable learning environments therefore belong to the conditions of relational educational worth, not merely to the technical implementation of a philosophical ideal (Selwyn, 2019; Miao & Holmes, 2023).

7.6 The disagreement objection: must formation converge on approved values?

Imagine two students who consider the same historical evidence, openly acknowledge uncertainty, listen carefully to those affected, and finally disagree about a contested public responsibility. If ethical responsiveness demanded agreement on a predetermined conclusion, one of them would necessarily be classified as educationally deficient. That would give an institution unacceptable power over conscientious judgement. The framework instead locates its minimum ethical demand in how persons are treated as participants whose claims merit consideration, while allowing substantial disagreement about policy and moral interpretation. This is not an empty proceduralism: exclusion, humiliation, and deliberate deception can violate a person's standing even when deployed in defence of an attractive substantive cause.

Pluralism does not mean that every outcome is equally defensible. Some interpretations ignore decisive evidence; others deny the credibility of affected persons without reasons. Educational evaluation can criticize those defects without demanding uniform convictions. Where disagreement remains reasonable,

the value may lie in the acquisition of a more articulate, responsible and revisable disagreement. This is an important limit to the rhetoric of transformation: a learner need not arrive at the educator's worldview to have undergone formation.

7.7 The apparent paradox of human nonexceptionalism

If we reject the need to demonstrate human cognitive uniqueness, why should the framework attend specifically to human learners? The answer is a question of scope and standing, not a claim to exclusive metaphysical importance. The present argument concerns institutions whose responsibilities include educating persons. It asks whether their obligations evaporate when other systems become capable of performing the same work. They do not. A person's claim to education can be grounded in the goods and relationships education enables without proving that no nonhuman system could have analogous goods or relationships. Whether future artificial entities could be educational subjects is a distinct philosophical question that this paper does not decide.

Likewise, the claim that an AI tutor mediates worthwhile learning does not require treating the tutor as morally interchangeable with a teacher. Machines may function as explainers, organizers, or conversational partners while human institutions retain obligations to protect learners and answer for the educational environment. Avoiding human exceptionalism in the justification of education is compatible with acknowledging present differences in accountability, vulnerability, embodiment, and institutional responsibility. The thesis would be weakened, not strengthened, by pretending these differences are philosophically trivial.

7.8 Is reflective appropriation an unfair standard?

A serious objection concerns learners who are very young, have significant communication impairments, or are only beginning to acquire disciplinary language. Requiring explicit argumentative defence of every belief would exclude them from educational value. Reflective appropriation must therefore be understood developmentally and relationally. In some contexts it appears as the ability to notice a mismatch, ask for help, reject an unwanted suggestion, or participate in supported decisions. Forms of answerability can be scaffolded rather than presupposed. The framework evaluates whether an educational environment cultivates and respects the learner's available forms of agency; it does not demand adult philosophical fluency as a condition of being educable.

This clarification also prevents a hidden return to the comparative-capacity standard. If education were valuable only once a person could demonstrate articulate, independent critical reflection, those who most need educational support would again be placed at the margins. The moral standing of a learner precedes the learner's demonstrated performance. The normative criteria are directed primarily at the defensibility of

educational practices and trajectories, not at ranking persons by how successfully they satisfy a demanding ideal.

7.9 A final objection: can educational change ever be evidenced?

A sceptic might grant the philosophical distinction but claim that relational transformation is too elusive to guide serious educational judgement. Students can mimic reflective vocabulary, teachers can misread silence, and changes in conduct may have multiple causes. These are genuine limits on inference. Yet they do not show that task performance is epistemically transparent. A highly polished essay, particularly one produced with undisclosed AI assistance, may also conceal the learner's understanding. The response should therefore be calibrated rather than triumphalist: evidence of formation is always partial, revisable, and situated. It becomes more credible when different contexts support the same interpretation and when a learner can confront unanticipated objections without merely reproducing a rehearsed script.

The proposal thus favors disciplined qualitative judgement rather than a claim to direct measurement of the inner self. An educator can document which reasons were exchanged, which positions were reconsidered, and what alternative readings the learner was willing to examine. The interpretation of these materials remains open to the learner's response and to collegial disagreement. Furthermore, not every institution needs to certify personal transformation as an outcome: some may responsibly cultivate the conditions for it while assessing independent capabilities for credentialing. The theory identifies a good to be protected in educational design, not a licence for institutions to grade a student's identity or infer private moral worth. That distinction helps reconcile philosophical seriousness with limits on surveillance and paternalism.

VIII. IMPLICATIONS FOR EDUCATIONAL PRACTICE AND EVALUATION

The argument alters what schools and universities should seek to justify, though it does not prescribe a uniform AI policy. Curriculum planners should distinguish tasks assigned to demonstrate independent competence from tasks designed to open a field of inquiry. Independent work remains appropriate when a target capability is itself required for responsible participation or safety. AI assistance may be appropriate when it improves access to perspectives, supports inquiry, or frees time for discussion without bypassing the learning aim. The standard is purposive alignment rather than an abstract percentage of machine-generated work.

Assessment should similarly separate at least three questions. Can the learner perform particular tasks at a required level? Can the learner explain and question the reasons on which important claims rest? Has the educational process enabled a defensible reorientation in how the learner understands or responds to the subject matter and to others? The third question cannot be reduced to a numerical transformation score. It may instead be explored through longitudinal reflection, dialogic examination, revisiting earlier claims, and

the interpretation of situated decisions. These are possible evidential practices, not empirically validated indicators supplied by this paper.

Institutional governance should examine the conditions of mediation. Can learners contest an AI's account? Are important claims traceable to trustworthy materials? Does the design encourage reflective disagreement rather than personalized compliance? Whose perspectives and experiences remain invisible? Can learners refuse AI mediation without educational penalty, and can those who require assistive technologies participate without stigma? Such questions translate the framework's normative commitments into criteria for responsible pedagogical deliberation.

Finally, the theory recommends restraint in claims about what has been 'improved'. A higher grade, faster completion time, or more positive attitude toward a course is not equivalent to a defensible change in self-world relations. The latter requires interpretive evidence and independent normative discussion. Nor does the theory guarantee that AI will generally foster *Bildung*. The framework offers a principled account of when AI's contribution would count as educationally valuable; empirical inquiry is still needed to determine where, for whom, and under which institutional conditions those contributions occur.

8.1 Curriculum design: deciding what should remain difficult

Curriculum deliberation should start with a precise statement of the educational purpose of a task. If the purpose is to master an elementary inferential relation, instant generation may short-circuit the very encounter the learner needs. If the purpose is to compare interpretations of a complex historical event, generation of alternative positions may broaden rather than narrow the field of encounter. If the purpose is to learn to contest evidence, an AI assistant can be deliberately instructed to identify uncertainties rather than settle disagreements. These decisions need not map onto a binary division between 'AI allowed' and 'AI forbidden'. They require a reasoned account of what must be done by the learner, what may be mediated, and what kinds of response would count as educationally significant.

A practical design sequence follows. Educators first identify the world or practice with which learners ought to develop a more adequate relation. They then specify which forms of friction are substantively necessary and which merely reflect avoidable barriers. Next they decide where AI can disclose multiple interpretations, support access, or introduce challenges without concealing the sources of its representations. Finally they build opportunities for learners to revisit initial judgements, confront objections, and explain their responses beyond the immediate prompt. The sequence does not guarantee *Bildung*; it makes the educational reasons for a technology choice inspectable.

8.2 Evaluation: parallel records of competence and formation

A defensible assessment system can maintain parallel records rather than collapsing all purposes into a single grade. One record concerns performance in specified tasks under defined assistance conditions.

Another concerns the quality of the learner's developing engagement with problems, evidence, and others. The second record might draw on a reflective discussion of changed interpretations, documented engagement with counterarguments, or an account of what the learner would now do differently. These practices should not become covert tests of ideological conformity. Their purpose is to make reasons and revisability visible, while protecting the possibility that a student will challenge the educator's interpretation.

Relational educational value is consequently better supported by narrative and dialogic judgement than by a standardized transformation index. Yet such judgement also requires safeguards: transparent expectations, room for reasoned disagreement, and attention to whether confidence or eloquence is being mistaken for learning. This approach makes assessment more accountable rather than less rigorous. Its rigour consists in specifying what is being judged and what evidence could undermine the assessment, not in assigning a precise number to a philosophically complex good.

8.3 Digital governance as educational governance

Institutional procurement decisions carry educational commitments. A system that hides source information, optimizes agreeable responses, or records sensitive disclosures without meaningful learner control can frustrate epistemic and reflective formation regardless of its fluency. Conversely, interfaces that make uncertainty salient, facilitate comparison with independent materials, and permit learner challenge may support more responsible encounters. These are not purely technical features. They influence which questions seem reasonable, which speakers appear credible, and how much disagreement the classroom can sustain. Procurement and evaluation should therefore include educators, students, accessibility specialists, and affected communities rather than leaving the definition of success solely to platform vendors.

The same reasoning applies to unequal access. Learners who depend on translation, assistive supports, or low-bandwidth technologies should not be judged by an ideal of unaided performance designed around more privileged students. At the same time, institutional enthusiasm for AI should not become coercion: meaningful alternatives and contestable decisions remain important. The layered standard makes these concerns internal to educational value. A supposedly transformative system that silences marginalized testimony or makes refusal practically impossible fails conditions of worthwhile formation even if its measured outcomes improve.

8.4 What remains for teachers and educational institutions

Teachers need not defend their profession by claiming permanent superiority in content generation. Their responsibility includes selecting significant encounters, noticing when apparent fluency masks misunderstanding, sustaining disagreement without humiliation, and taking responsibility for the conditions under which learners are addressed. These practices can themselves be supported by AI, and some could eventually be performed competently by machines. Their educational importance nevertheless remains

because institutions owe learners environments in which reasons can be tested and relationships can develop without coercive control. Whether an individual task is performed by a human or an AI system does not eliminate the need to assign and discharge that responsibility.

The theory is intentionally less decisive about which institutional arrangements will best realize its criteria. It does not predict the long-term effects of conversational systems, stipulate an ideal allocation of human and machine labour, or supply a validated measure of formation. Those are tasks for empirical, legal, and policy inquiry informed by normative clarity. Its philosophical contribution is to prevent those inquiries from quietly assuming that educational worth rises or falls with comparative cognitive productivity.

IX. CONCLUSION

The prospect of highly capable AI makes an old instrumental assumption difficult to ignore: the worth of education is sometimes judged by the tasks that educated people can perform better than available alternatives. If educational justification depends on a continually shrinking inventory of abilities machines cannot replace, technological improvement will always appear to threaten the point of educating human beings. That conclusion follows from the premise, not from any demonstrated loss of educational value.

This paper has offered a different premise. The constitutive worth of education can lie in enabling persons to develop more answerable, reflective, and ethically responsive relations with the world. This claim does not romanticize self-formation or presume that all change is good. It is constrained by safeguards against domination and serious harm, and assessed through epistemic answerability, reflective appropriation, and ethical responsiveness. These conditions allow extensive assistance to coexist with educational worth while refusing to count fluent misinformation, coercive persuasion, or mere output production as sufficient achievement.

The central implication is not that humans possess a mysterious quality AI will never match. It is that educational value is the wrong object of a competition in task performance. We should ask neither simply what machines can do for learners nor what learners can still do without machines, but what kinds of relations to a shared world our educational arrangements help persons to form—and whether those relations are worthy of defence.

References

- Arstorp, A.-T., Amdam, S. H., Nagel, I., Letnes, M.-A., & Tømte, C. E. (2026). Bildung at a crossroads: Teacher education in the age of GenAI. *Journal of Education for Teaching*, 52, 340–358.
<https://doi.org/10.1080/10476210.2026.2625721>
- Bialystok, L. (2026). AI and the future of (philosophy of) education. *Educational Theory*, 76(1), 133–139.
<https://doi.org/10.1111/edth.70058>

- Biesta, G. J. J. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21(1), 33–46.
<https://doi.org/10.1007/s11092-008-9064-9>
- Biesta, G. J. J. (2020). Risking ourselves in education: Qualification, socialization, and subjectification revisited. *Educational Theory*, 70(1), 89–104. <https://doi.org/10.1111/edth.12411>
- Biesta, G. J. J. (2022). *World-centred education: A view for the present*. Routledge.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
<https://doi.org/10.1093/analys/58.1.7>
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. Macmillan.
- Dewey, J. (1938). *Experience and education*. Macmillan.
- Du, Y., & Yuan, Y. (2026). Epistemic dependence in AI-mediated learning. *AI & Society*.
<https://doi.org/10.1007/s00146-026-03294-1>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7), 335–349.
<https://doi.org/10.2307/2026523>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.
<https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- Nussbaum, M. C. (2010). *Not for profit: Why democracy needs the humanities*. Princeton University Press.
- Rosa, H. (2019). *Resonance: A sociology of our relationship to the world* (J. C. Wagner, Trans.). Polity.
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity.
- Tishenina, M. (2026). Generative AI and the irreducible role of teachers as callers into subjectness: A placebo–nocebo–treatment framework. *Studies in Philosophy and Education*. <https://doi.org/10.1007/s11217-026-10051-w>

Declarations

Article type: Conceptual and critical philosophical analysis. No human-participant data were collected.

Ethics approval: Not applicable.

Data availability: No datasets were generated or analysed for this conceptual study.

Funding and conflicts of interest: Author confirmation required before formal publication.