frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 17 February 2026
DOI 10.3389/feduc.2026.1760626

Check for updates

OPEN ACCESS

EDITED BY
Tom Prickett,
Northumbria University, United Kingdom

REVIEWED BY
Ryan Thomas Williams,
Teesside University, United Kingdom
Marie Devlin,
Newcastle University, United Kingdom

*CORRESPONDENCE
Viviana Vinci
viviana.vinci@unifg.it

RECEIVED 18 December 2025
REVISED 21 January 2026
ACCEPTED 22 January 2026
PUBLISHED 17 February 2026

CITATION
Vinci V, Agrati LS, Berardi P and Beri A (2026) Artificial intelligence as a catalyst for transformative assessment: designing teacher literacy at the crossroads of ethics, pedagogy, and human relationships.
Front. Educ. 11:1760626.
doi: 10.3389/feduc.2026.1760626

COPYRIGHT
© 2026 Vinci, Agrati, Berardi and Beri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Artificial intelligence as a catalyst for transformative assessment: designing teacher literacy at the crossroads of ethics, pedagogy, and human relationships

Viviana Vinci[1]*, Laura Sara Agrati[2], Pierangelo Berardi[3] and Arianna Beri[4]

[1]Department of Humanities, Literature, Cultural Heritage, Education Sciences, University of Foggia, Foggia, Italy, [2]Department of Education and Sports Sciences, Telematic University Pegaso, Naples, Italy, [3]Department of Education, Psychology, Communication Sciences, University of Bari, Bari, Italy, [4]Department of Humanities and Social Sciences, University of Bergamo, Bergamo, Italy

The rapid proliferation of Generative Artificial Intelligence (GenAI) in educational settings necessitates a redefinition of assessment practices, raising crucial questions regarding the relationship between automation and pedagogical ethics. While AI offers opportunities for efficiency and personalization, there persists a risk of a technocratic drift that threatens the relational dimension of teaching. This study explores the perceptions of trainee teachers regarding the integration of AI in assessment, investigating the transformation of practices (RQ1) and the competencies required for an "AI-literate" teacher profile (RQ2). Adopting a convergent mixed-methods design, a survey was conducted on a sample of 662 teachers enrolled in qualifying courses (60 ECTS) at the University of Foggia and the University of Bergamo. Data collected via a CAWI survey were analyzed using Spearman's correlation and thematic analysis supported by text mining. The data reveal a "conscious resistance": despite possessing a good level of technical comfort, pedagogical trust in AI remains low (M = 4.83/10). Inferential analysis highlights a training paradox: trust does not correlate with the technical training received (ρ = 0.36), but rather with the perception of concrete support (ρ = 0.57). Qualitatively, concerns regarding dehumanization and a demand for a principle of subsidiarity emerge. The study suggests that technical competence is insufficient for AI adoption. A Critical AI Literacy that integrates ethics and pedagogy is required, restoring decision-making control to the teacher to ensure fair and humanizing assessment.

KEYWORDS

AI literacy, artificial intelligence in education, ethics, formative assessment, teacher agency

1 Introduction

The pervasive and rapid diffusion of Artificial Intelligence (AI), with particular reference to Generative AI (GenAI) systems, is radically redefining the foundational paradigms of education. This suggests the need for a critical revision of how learning is conceived, delivered, and, above all, assessed (OECD, 2023; UNESCO, 2023). In this continuously evolving scenario, assessment emerges as a strategic yet extremely delicate arena, raising questions that transcend the mere technical dimension to affect the epistemological and ethical spheres of the educational act (Luckin et al., 2022; Holmes and Tuomi, 2022). In this study, a distinction is drawn between AI Literacy and

traditional Digital Literacy. As observed by Ng et al. (2021), the concept of literacy is broadening to address the complexity of algorithmic environments; whereas digital literacy primarily concerns the operational proficiency required to utilize technological tools, AI Literacy constitutes a more intricate, multidimensional construct. Grounded in the frameworks proposed by Long and Magerko (2020) and expanded by Luckin et al. (2022) and Williams (2024), teacher AI Literacy is defined here as the intersection of conceptual understanding, the ability to comprehend the underlying logic of machine learning algorithms, and critical agency, which entails evaluating ethical implications such as bias and data privacy. Additionally, this competence encompasses pedagogical governance, understood as the strategic ability to discern when to delegate tasks to AI, thereby supporting the transition from passive users to critical evaluators of the technology. The introduction of algorithms capable of generating complex content should not be viewed merely as a simple functional substitution of traditional tools; rather, it may act as a catalyst prompting a rethinking of assessment practices through a more equitable, inclusive, and personalized lens, while averting the risk of a technocratic drift that would reduce formative judgment to a computational automatism.

However, the reception of these technologies within the school system is neither linear nor devoid of conflict. As highlighted by recent investigations into initial teacher training, a marked polarization is evident, oscillating between enthusiasm and resistance (Selwyn, 2016; Buchholz et al., 2020). While a fascination with innovation and the potential for personalization emerges on one side, on the other, the fear of professional "deskilling" is consolidating: the perceived risk that automation may erode fundamental pedagogical competencies, delegating the responsibility of judgment to algorithmic "black boxes" and reducing teaching to mere output management. This tension highlights how AI is perceived not merely as a tool, but as an agent that interrogates the very identity of the teacher.

Regarding opportunities, the literature highlights significant cross-cutting benefits. For teachers, GenAI represents a powerful resource for automating low-value-added and time-consuming tasks, such as the correction of structured tests or the preliminary analysis of learning data, thereby ensuring greater consistency in grading and freeing up cognitive resources to be reinvested in the educational relationship (Chan and Hu, 2023; Liu, 2024). In parallel, AI supports instructional design by offering creative insights and materials aligned with learning objectives. For students, the added value lies in access to adaptive and on-demand learning: intelligent tutoring systems can simplify complex concepts and offer immediate feedback, fostering personalized pathways that respect individual cognitive rhythms (Kasneci et al., 2023; Walter, 2024).

Nevertheless, this technological integration clashes with profound criticalities that threaten the validity of the assessment process itself. One of the most urgent concerns regards academic integrity, which is severely tested by the ease with which students can use chatbots to draft assignments, incurring practices of sophisticated plagiarism that are often difficult to detect (Gruenhagen et al., 2024; Lim et al., 2023). Even more insidious is the risk linked to the potential decline of cognitive skills: an over-reliance on GenAI for problem-solving could hinder the development of critical thinking and creativity, leading to a form of "cognitive laziness" wherein the student delegates deep processing to the machine (Çela et al., 2024; Zhai et al., 2024).

In this balance between risks and benefits, a determining factor comes into play: "familiarity" with the tool. Specific studies in the context of Teacher Education (Chai et al., 2016; Celik, 2023) suggest that the critical acceptance of AI does not depend solely on its perceived utility, but is strongly correlated with the teacher's degree of technical and pedagogical familiarity. Resistance is often not ideological but the result of scarce exposure that generates anxiety and uncertainty; conversely, greater habituation allows for the demystification of the technology and its conscious integration into teaching practices.

In light of these challenges, the need to move beyond standardized assessment methods in favor of forms of authentic assessment, based on processes that AI cannot entirely simulate, becomes increasingly relevant (Salinas-Navarro et al., 2024). This shift advocates for a deep redefinition of teachers' assessment literacy (Bearman et al., 2020). Possessing traditional assessment competencies may no longer be sufficient; it appears crucial to train an "AI-literate" teacher profile, endowed with solid critical and ethical awareness. Consequently, the integration of AI in education calls for a paradigm shift from simple automation to a Human-in-the-loop (HITL) approach (Memarian and Doleck, 2024; Mosqueira-Rey et al., 2023). However, to fully preserve teacher agency, we argue for the adoption of the more advanced AI-in-the-loop (AI^2L) perspective recently proposed by Natarajan et al. (2025). In this framework, the human expert is not merely a supervisor 'in the loop' of the algorithm, but the active owner of the educational process, while the AI is positioned 'in the loop' of the teacher as a subservient support tool. Adopting an AI^2L perspective mitigates ethical risks such as bias and opacity, ensuring that assessment remains a pedagogical act guided by human intentionality rather than an automated statistical output. Consistent with this scenario, the present research aims to explore the relationship between Artificial Intelligence (AI) and educational assessment through a critical and transformative lens, placing particular attention on the changes emerging in the professionalism of trainee teachers. Specifically, the investigation is guided by two main research questions:

- RQ1: How do trainee teachers perceive the transformation of assessment practices mediated by AI, and what tensions emerge between automation and pedagogical responsibility?
- RQ2: What specific training dimensions are central to outlining an "AI-literate" teacher profile capable of integrating AI into assessment without compromising the educational dimension?

2 Materials and methods

2.1 Research design and participants

The study employs a convergent parallel mixed-methods design, aimed at investigating the perceptions and competencies of trainee teachers regarding the integration of AI in assessment. Data collection was conducted between April and May 2025 via digital administration.

The convenience sample consists of $N = 662$ teachers enrolled in qualifying courses (60 ECTS) at the University of Foggia ($n = 523$; 79.0%) and the University of Bergamo ($n = 139$; 21.0%).

The demographic composition (Table 1) highlights a prevalence of female participants ($n = 424$; 64.0%) compared to the male component ($n = 234$; 35.3%). The distribution by age group reveals a heterogeneous population: although the modal group is represented by the 25–29 age range ($n = 176$; 26.6%), a significant presence of trainee teachers in the over-45 range ($n = 162$; 24.5%) is observed, delineating a multigenerational training context.

10.3389/feduc.2026.1760626

TABLE 1 Sociodemographic and professional characteristics of the sample (N = 662).

Characteristics	Number of participants (*n*. 662)	Percentage of participants (%)
Gender		
Male	234	35.3
Female	424	64.0
I prefer not to answer	2	0.3
Non-binary	2	0.3
Age		
Under 24 years	39	5.9
From 25 to 29 years	176	26.6
From 30 to 34 years	122	18.4
From 35 to 39 years	97	14.6
From 40 to 44 years	66	10.0
Over 45 years	162	24.5
Education		
Master's degree	464	70.0
Postgraduate degree	30	4.5
PhD	19	2.9
Other	149	22.5
Missing	189	28.6
Years of Teaching		
0–5	338	51.0
6–10	116	17.6
11–20	19	2.8

In terms of academic background, the majority of participants hold a Master's Degree (70.0%), followed by postgraduate qualifications or research doctorates (7.4% overall). Regarding professional experience, the sample is primarily divided between teachers with less than 5 years of service and those with experience ranging between 6 and 10 years (51.0%).

2.2 Data collection instrument

The data were collected between April and May 2025 using a Computer-Assisted Web Interview (CAWI) methodology. The survey was self-administered via a secure online platform (Microsoft Forms), and the link was distributed to participants through the institutional Learning Management Systems (LMS) of the involved universities. Participation was voluntary and anonymous, with an estimated completion time of approximately 20 min. The instrument consisted of 57 items organized into four logical sections, which were subsequently operationalized into five analytical dimensions (Trust, Competence, Relational Impact, Training, Perceived Utility) for the analysis:

1 Socio-Professional Profile and Motivation (Items 1–11): This section collected demographic data (gender, age), educational background, and professional status (years of teaching, school grade). It also investigated the participants' intrinsic motivation for training and their priority expectations regarding professional competence acquisition.
2 Pedagogical and Assessment Background (Items 12–25): This section explored the participants' existing "analog" practices before introducing AI. Items queried their perceived self-efficacy in instructional design, the weight attributed to assessment tools, and their preferences regarding traditional ("paper-and-pencil") versus software-mediated assessment formats.
3 AI Familiarity and Perceived Utility (Items 26–38): This section measured the "digital baseline" of the trainees. It included self-assessment scales (1–10) on comfort with digital technologies and specific AI tools (e.g., ChatGPT). It also investigated the specific perceived utility of Generative AI for tasks such as self-correction of writing skills and students' learning errors.
4 Ethical Agency and Future Scenarios (Items 39–57): The core of the survey investigated the tension between automation and human control. Through agreement scales and open-ended questions, participants were asked to define which teaching activities could be delegated to AI and which should remain exclusively human to preserve educational effectiveness. Specific items addressed the risks of dehumanization, the "trust" in AI-supported assessment, and the need for critical control over algorithmic decisions.

To ensure maximum measurement sensitivity, the response scales were adapted to the nature of the constructs. Specifically, Trust, Technical Comfort, Training Adequacy, and Relational Risks were measured using a 10-point estimation scale to capture subtle variations in sentiment and avoid the flattening effect of shorter scales. Conversely, specific constructs related to Perceived Utility (e.g., for

TABLE 2 Overview of the key variables analyzed.

Variable label (Figure 1)	Construct measured	Related item (translated)	Scale range
Trust	Trust in assessment	*What is your level of trust in the use of AI to support educational assessment?* (Item 48)	1 (very low)—10 (very high)
Support	Pedagogical support	*I believe that Artificial Intelligence can support assessment.* (Item 43)	1 (totally disagree)—10 (totally agree)
Comfort	Technical comfort	*How comfortable do you feel using Artificial Intelligence tools?* (Item 34)	1 (not at all comfortable)—10 (totally comfortable)
Training	Training adequacy	*How much do you think your current training develops skills in the ethical management of collaboration between teachers and AI?* (Item 52)	1 (not at all effective)—10 (very effective)
Usefulness	Perceived utility	*How useful do you consider GenAI for students' self-correction tasks?* (Item 26)	1 (not at all useful)—5 (very useful)
Usage	Frequency of use	*Have you ever utilized AI tools in the educational context?* (Item 32)	Binary/Freq. (yes—no)
Agency	Critical agency	*How prepared do you feel to critically decide which aspects of teaching can be entrusted to an AI?* (Item 51)	1 (not at all prepared)—10 (very prepared)

writing correction) and traditional pedagogical background utilized a standard 5-point Likert scale. Table 2 provides a detailed overview of the key variables and the measurement scales used in the present study.

2.3 Data analysis

The analysis plan was conducted entirely using the R statistical computing environment (via the RStudio interface) and was structured into distinct procedures for the two types of data. Regarding the quantitative component, descriptive statistics (frequencies, mean, and standard deviation) were initially calculated to characterize the sample profile. Subsequently, to address the research questions, Spearman's bivariate correlation analysis (ρ) was applied using the *psych* and *tidyverse* packages; this methodological choice was dictated by the ordinal nature of the variables (measured on Likert scales) and the non-normality of the data distribution, thereby ensuring greater robustness compared to parametric coefficients. On the qualitative front, open-ended responses were subjected to Inductive Thematic Analysis following the model by Braun and Clarke (2006). To ensure the trustworthiness and reduce the subjectivity of the inductive coding process, a methodological triangulation strategy was adopted. The themes emerged from the manual qualitative analysis were cross-validated against the results of the automated Text Mining analysis performed in R. Specifically, the high-frequency lexical clusters identified by the software (using tidytext and tm packages) consistently matched the semantic categories manually identified by the researchers. This convergence between human interpretation and algorithmic extraction confirms the robustness of the coding frame and mitigates the risk of researcher bias.

3 Results

3.1 Transformation of assessment and technological ambivalence

Regarding the first research question, aimed at investigating the transformation of assessment in the interaction with AI, the data outline a scenario characterized by a deep ambivalence between technical competence and pedagogical trust. Descriptive analysis highlights that, despite a moderate level of reported comfort in the general use of intelligent tools ($M = 6.04$; SD = 2.52), trust in delegating assessment support to these systems suffers a significant decline, settling on mean values below the sufficiency threshold ($M = 4.83$; SD = 2.21). Even more critical is the data regarding the perceived utility of AI for self-correction, which records a minimal mean ($M = 2.57$ on a 5-point scale), suggesting widespread skepticism toward the current didactic efficacy of such tools. Sample polarization is further confirmed by the distribution of responses on AI's capacity to support assessment ($M = 5.18$), where the modal frequency coincides with the scale's minimum value (1), indicating the presence of certain resistance. However, inferential analysis, conducted via Spearman's correlation coefficient, clarifies that trust is not independent of competence: indeed, significant positive associations emerge between the level of trust and the perception of the support role ($\rho = 0.56$), as well as between trust itself and the degree of technological comfort ($\rho = 0.54$). The qualitative analysis allows for decoding the underlying reasons for these numerical trends, revealing four semantic areas that explain the tension between acceptance and rejection:

Instrumental Efficiency (Acceptance): On the pragmatic axis, a view oriented towards optimization consolidates. Teachers recognize AI as a valuable resource for time management, the automation of repetitive correction tasks, and the enhancement of feedback speed.

Relational Depersonalization (Rejection): Conversely, on the emotional axis, a pervasive fear of "dehumanization" emerges. Participant narratives describe the risk of a drift towards an "arid," "sterile," and "distant" assessment, deprived of the empathetic component deemed essential for education.

Epistemic Opacity (Distrust): A third area concerns the lack of transparency. The recurring references to the inability to explain the algorithm's judgment ("I do not know how it decides") fuel the distrust recorded in the quantitative data, linking resistance to the "Black Box" phenomenon.

Human Irreducibility (Defense): Finally, the tension resolves in the claim of a "human reserve." Assessment is described as a hermeneutic act requiring context and sensitivity, qualities that

TABLE 3 Descriptive statistics of the main dimensions.

Analytical dimension	Reference item	Scale range	Mean (M)	Std. Dev. (SD)	Interpretation of result
Technical comfort	Item 34	1–10	6.04	2.52	Moderate/high technical familiarity
Critical agency	Item 51	1–10	5.39	2.45	Sufficient perceived capability
Training adequacy	Item 52	1–10	4.29	2.63	Insufficient preparation
Trust in AI	Item 48	1–10	4.83	2.21	Low trust (resistance area)
Perceived utility	Item 26	1–5	2.57	1.10	Low utility for self-correction
Pedagogical support	Item 43	1–10	5.18	2.80	Ambivalent/uncertain

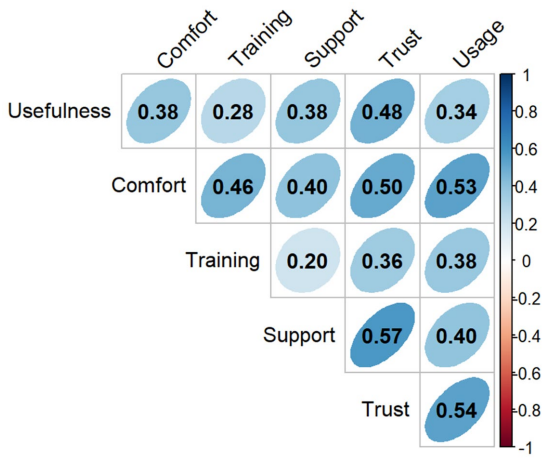

FIGURE 1
Spearman correlation matrix (N = 662) illustrating the relationships between key variables influencing AI adoption in assessment. The variables are defined as follows: Usefulness (perceived utility of AI for self-correction tasks), Comfort (technical ease of use), Training (perceived adequacy of specific training received), Support (perception of AI as a valid pedagogical support), Trust (confidence in AI-based assessment reliability), and Usage (reported frequency of AI use).

participants define as exclusively human and non-delegable to a machine.

This dialectic resolves into an explicit demand for a principle of pedagogical subsidiarity, according to which AI is legitimized exclusively as an ancillary support tool, without ever being able to replace professional judgment and the educational centrality of the teacher (see Table 3).

3.2 The "AI-literate" teacher profile: between ethics and technique

Regarding the second research question, aimed at delineating the central training dimensions for a teacher capable of integrating AI without educational compromises, the descriptive data primarily indicate a significant gap in perceived preparation. The adequacy of the training received is evaluated as insufficient, with a mean of 4.29 (SD = 2.63) and a modal frequency situated at the scale's minimum value (1), pointing to a widespread perception of structural deficiency in current training pathways. Nevertheless, participants express moderate confidence in their capacities for critical discernment (M = 5.23) and ethical management (M = 5.13), as well as in the competence to mediate the use of AI while preserving the educational relationship (M = 5.39). The inferential analysis, conducted on a sample of 662 subjects using Spearman's correlation, offers a crucial interpretive key for understanding these self-perceptions. A network of statistically significant relationships ($p < 0.001$) emerges, delineating a complex framework of technological acceptance. A finding of particular relevance concerns the role of training: while the training received correlates positively with the sense of technical comfort in using the tool (ρ = 0.46), it shows a decidedly weaker association with trust in assessment (ρ = 0.36) and a marginal association with the perception of AI as an effective didactic support (ρ = 0.20). Conversely, trust in AI as an assessment tool appears to be driven much more strongly by the perception of concrete pedagogical support (ρ = 0.57) and by the actual use of the tool (ρ = 0.54), as well as by its perceived pragmatic utility (ρ = 0.48) (see Figure 1).

These results suggest the existence of a "training paradox": current pathways appear effective in fostering technical confidence ("knowing how to use the tool"), yet prove less impactful in constructing an integrated pedagogical vision ("knowing how to assess with the tool"). Consequently, teachers' trust appears to stem not so much from theoretical instruction as from direct and pragmatic experience. This highlights the necessity to reorient training toward the development of a Critical AI Literacy that anchors technical competence in ethical and decision-making responsibility.

3.3 Qualitative analysis: the phenomenology of resistance and training needs

The qualitative inquiry, supported by lexical evidence emerging from text mining, provides a vivid snapshot not only of what teachers think but of how they position themselves professionally vis-à-vis technology. The analysis of the textual corpus allows for the construction of training needs into three interconnected dimensions: operational uncertainty, ethical urgency, and the defense of pedagogical humanism.

First and foremost, the most immediate finding is the pervasiveness of doubt. The recurrence of dubitative linguistic forms, specifically the lemma "Uncertainty," signals a state of operational paralysis (Figure 2). This term, appearing alongside the explicit demand for "Training" and technical "Knowledge," suggests that the

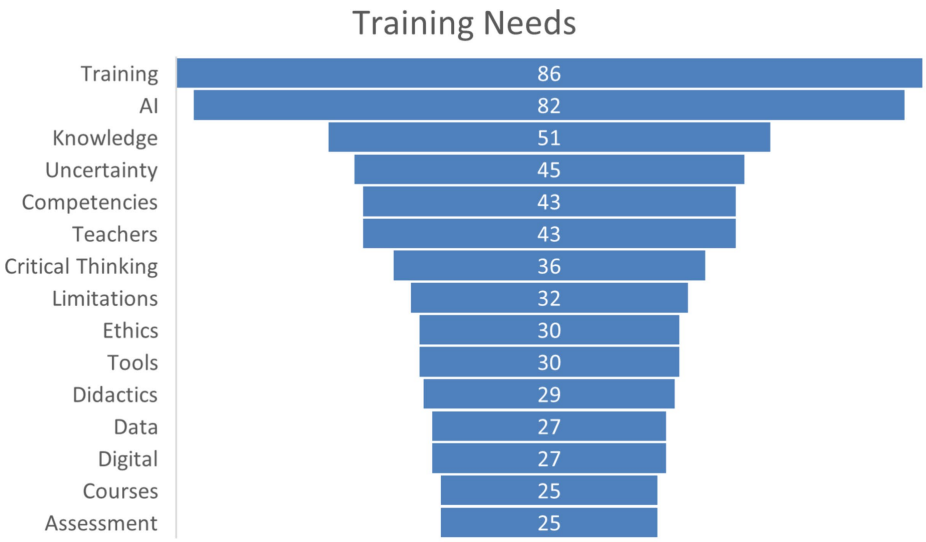

FIGURE 2
Lexical frequency of training needs. The bar chart highlights the most recurring terms regarding teachers' professional development needs, emphasizing operational uncertainty alongside the demand for training.

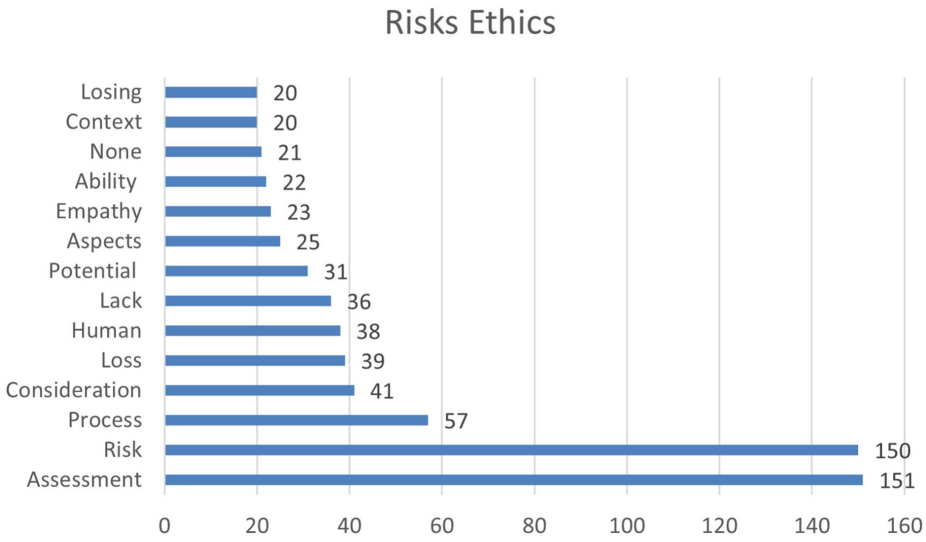

FIGURE 3
Semantic mapping of perceived risks. Frequency analysis of terms associated with the dangers of AI in assessment, highlighting ethical concerns over technical ones.

issue is not merely a lack of tools, but a difficulty of pedagogical imagination: many respondents struggle to concretely envision how AI can be integrated into their routine without disrupting it. This void of "vision" implies that current training may have been too theoretical, leaving teachers devoid of role models or practical examples of sustainable integration.

Once the initial bewilderment is overcome, the training request crystallizes in a precise direction: the demand for meaning prevails over the demand for tools. The prominence of terms such as "Critical Thinking" and "Ethics," contrasted with the high frequency of "Risk" and "Loss" (Figure 3), indicates that teachers are not asking to become computer technicians, but are claiming a deontological compass. The dominant concern does not regard the mechanics of the algorithm, but the management of its implications: data privacy, the recognition of cognitive biases, and, above all, the ability to discern when the use of technology is an added value and when it represents a loss of educational context.

This evidence supports the hypothesis that Critical AI Literacy is not an academic luxury, but a pragmatic need for those who must manage educational responsibility in the classroom.

Finally, the semantic analysis of instructional strategies reveals a clear defensive line of demarcation. The strong recurrence of the term "Relationship," flanked by "Human" and "Educational" values, unveils the teachers' adaptation strategy (Figure 4): AI is welcomed if it acts

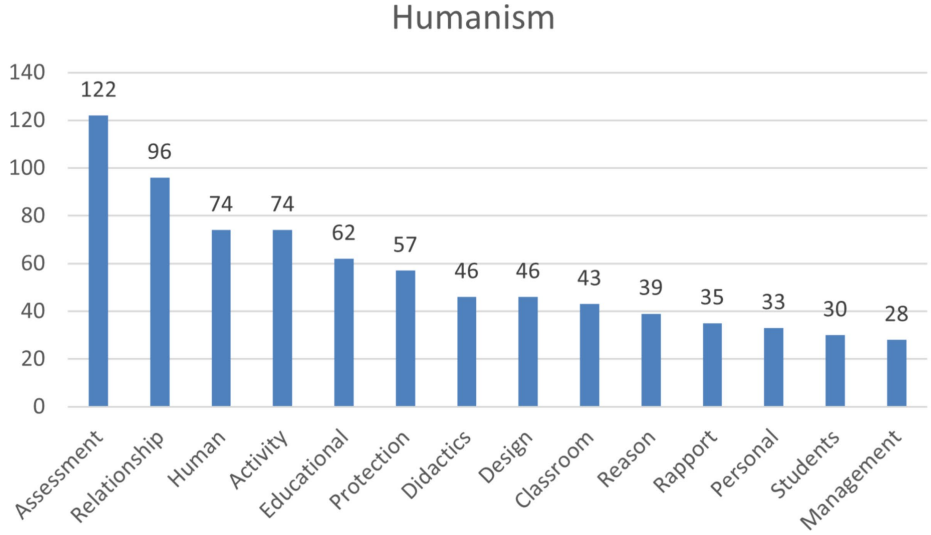

FIGURE 4
The defense of pedagogical humanism. Analysis of non-delegable tasks, showing the dominance of relational and human-centric terms.

as a "prosthesis" to enhance attention to the individual (e.g., by creating differentiated materials), but is rejected when it attempts to surrogate the relationship. In the free narratives, assessment is described as a hermeneutic act, an interpretation of student growth that requires empathy and context. Consequently, the "educational centrality" invoked by participants translates into the will to guard the emotional space of learning, delegating procedural execution to the machine while firmly retaining the monopoly on value judgment and educational care.

4 Discussion

The body of collected evidence places a crucial challenge for teacher training in the AI era at the center of the academic debate: the necessity to redefine assessment not merely as a docimological procedure, but as a reflective educational practice capable of integrating the computational power of algorithmic tools without yielding to either professional abdication or the technicization of formative processes. The research results confirm that AI, far from representing a simple instrumental opportunity, acts as a radical transformative device that deeply interrogates the very identity of the teacher and their functions of ethical and pedagogical guarantee. Primarily, the data reveal a profound tension between the perception of AI as a tool for operational efficiency and the pervasive fear of a dehumanization of the assessment gesture.

This ambivalence mirrors the dimension of "Symbolic Representation" identified in our qualitative analysis: while teachers recognize AI's potential in automating repetitive tasks, quantitative data record drastically low trust ($M = 4.83$ on a 1–10 scale) and marked negative polarization. Such a discrepancy suggests that participants are not expressing a Luddite judgment, but are actively evaluating the risks of a delegation that could compromise the relational nature of teaching. The demand for a "principle of subsidiarity," according to which AI must act exclusively as ancillary support, finds statistical validation in the positive correlation between pragmatic utility and trust ($\rho = 0.48$); this indicates that AI is accepted only if confined to the role of a "prosthesis" for efficiency. This hesitation also reflects a latent ethical concern regarding the epistemic opacity of these systems. As noted by Widder et al. (2024), current AI models often function as 'black boxes' where knowing the underlying code or weights does not guarantee interpretability of the outputs. Since these systems lack semantic transparency regarding their decision-making logic, teachers appear reluctant to validate assessment outcomes they cannot fully trace or explain to students. Responding to the need to situate these findings within the specific national context, the data allow us to reinterpret the participants' apparent "prudence" not as cultural resistance, but as a rational consequence of a systemic training gap. Contrary to the narrative of teachers fearing professional replacement, our analysis shows that only a minority ($N = 195$; 29.5%) agrees that AI might replace their skills ($M = 4.57$). The determining factor for their caution is instead found in the lack of preparation: a striking 64.2% ($N = 425$) of participants explicitly state that their training did not adequately prepare them to use AI tools ($M = 3.67$, SD = 2.55), and only 29.9% ($N = 198$) feel fully capable of critically deciding which tasks to delegate. This evidence suggests that the Italian educational context, currently in a transitional phase of digital reform, has generated a demand for innovation without providing the necessary "AI Literacy". Consequently, the low trust observed should be read as a demand for competence and ethical governance rather than a refusal of technology. Specifically, the data suggest that teachers are implicitly calling for an AI-in-the-loop (AI^2L) configuration (Natarajan et al., 2025): they do not reject the efficiency of the tool (as shown by the correlation with Utility), but they express a need for training that empowers them to maintain control over the learning loop. Their hesitation reflects a resistance to models where the human is passive, aligning instead with the Machine Teaching paradigm (Mosqueira-Rey et al., 2023), where the teacher's role is to explicitly guide and correct the AI's logic. A second element of crucial epistemological significance emerges from the analysis of the determinants of competence. The data refute the assumption that

formal training alone automatically generates technology adoption. Indeed, inferential analysis highlighted that while the training received correlates with the sense of technical comfort ($\rho = 0.46$), it shows only a moderate association with trust in the tool ($\rho = 0.36$). Conversely, the most significant statistical evidence supporting integration is the strong positive correlation ($\rho = 0.57$) found between the perception of AI as concrete pedagogical support and the level of trust. This result dismantles the technocentric narrative: professional security does not derive from abstract instruction, but from the pragmatic validation of the tool's utility in a real-world context ("Utility" vs. "Trust," $\rho = 0.48$). Thus, the profile of a teacher who becomes "AI-literate" emerges not when they learn to use the software, but when they recognize its subsidiary value for their own didactic action, exercising reflective control over assessment practices. This profile imposes a substantial revision of current initial and in-service training models. The data indicate an urgent need to move beyond technocentric instructional approaches in favor of a Ethical AI Literacy. As evidenced by the discrepancy between comfort and trust, teachers are not merely asking for technical manuals, but are claiming a deontological compass. Training programs should explicitly address the ethical dimensions of automation: data privacy, the recognition of algorithmic bias (which can reinforce social inequalities), and the moral implications of delegating evaluation. Teachers need to be trained not just to *operate* the machine, but to exercise epistemic responsibility: validating the fairness of the AI's output before it impacts the student's learning path. In this direction, the study's results not only enrich theoretical reflection but also provide operational guidelines for training policies: advanced digital competence must necessarily hybridize with ethics and pedagogy to generate a professional culture capable of governing innovation. We can affirm that assessment today constitutes the privileged locus for observing and orienting the impact of AI in educational settings. It becomes an interpretive lens and a transformative lever for exploring new balances between automation and reflexivity. However, as suggested by the discrepancy between technical comfort and pedagogical trust, this transformation cannot be sustained by mere instrumental training. It is within this liminal space that the challenge of constructing a new culture of teaching plays out: a culture that must respond to the empirical demand for 'AI-in-the-loop' governance, inhabiting the complexity of the algorithmic era not just with technical skill, but with the critical and ethical spirit claimed by the teachers themselves.

4.1 Implications for policy and practice

The evidence collected in this study suggests that the observed resistance is not a barrier to be dismantled, but a meaningful signal to be heeded, translating into specific implications for policy and practice. Consequently, we propose a strategic realignment of teacher training and school governance based on the "Training Paradox" and the demand for ethical control emerged from the data.

Regarding curriculum design, the discrepancy between high technical comfort and low trust indicates that professional development programs must urgently shift their focus from mere operational interaction ("prompt engineering") to a rigorous "Output Validation Literacy." Teachers need to be trained not as passive users but as critical evaluators capable of detecting hallucinations, logic errors, and algorithmic biases in AI-generated content. Crucially, this literacy must be grounded in a deep ethical understanding of the tool; to mitigate the distrust fueled by "epistemic opacity," training modules should demystify the "Black Box" nature of Large Language Models, clarifying their probabilistic rather than logical functioning. Only by understanding the mechanical limits of the machine can teachers exercise true epistemic responsibility without fearing professional abdication.

At the institutional level, these insights call for the standardization of "Human-in-the-loop" (AI^2L) workflows. School policies should explicitly regulate that while AI may assist in preliminary screening or feedback generation, the final assessment and its semantic explanation must remain the exclusive prerogative of the human teacher, thereby formalizing the "Principle of Subsidiarity" demanded by participants. In operational terms, we recommend adopting "Triangulated Assessment" practices, where AI is utilized not as a grader but as an adversarial critic or a "second opinion" to refine the teacher's own judgment. This approach ensures that the integration of AI enhances rather than replaces the relational and hermeneutic value of educational assessment.

4.2 Limitations of the study

Despite the significance of the findings and the sample size, the present investigation entails certain limitations that warrant a cautious interpretation of the results. First, the non-probabilistic nature of the sampling strategy and the geographical concentration of participants in two specific universities (Foggia and Bergamo) preclude the statistical generalizability of the data to the entire Italian teaching population, although they offer a representative cross-section of the current state of initial teacher training. Second, the sample composition, characterized by a prevalence of novice teachers or those in the qualification phase with less than 5 years of experience, may not fully reflect the professional stances, resistances, or consolidated pedagogical competencies of tenured teachers with greater seniority, who might approach technological innovation through different dynamics. A further limitation lies in the data collection methodology based on self-reported measures. While effective for investigating perceptions and beliefs, this approach may be subject to social desirability bias, particularly regarding sensitive topics such as ethics and digital competence. Consequently, it remains to be verified through future observational or experimental studies the extent to which declared attitudes translate into actual instructional practices in the classroom. Finally, the rapid evolution of Generative AI tools necessitates viewing these results as a snapshot situated within a specific historical-technological moment, inevitably destined to evolve as teachers' familiarity with these new devices progresses. Additionally, it must be noted that the correlational nature of the study allows us to identify significant associations between variables (e.g., Training and Trust) but does not establish direct causal links. Therefore, the interpretative models proposed (such as the Training Paradox) should be understood as probabilistic frameworks to be further verified through experimental designs.

4.3 Future research directions

In light of the emerging evidence and the outlined limitations, several promising avenues for future research arise. First, it would be beneficial to extend the investigation through longitudinal studies

monitoring the evolution of the perceptions and assessment practices of the same subjects during their professional induction, verifying whether field experience modifies trust or resistance toward AI. Second, it appears crucial to complement the survey of perceptions with direct observational or experimental analysis of classroom practices, investigating how teachers effectively integrate AI into assessment design and formative feedback. Finally, a promising line of inquiry concerns the comparison between STEM and Humanities teachers, to verify whether the epistemological nature of the subject taught influences the propensity to delegate or retain specific assessment functions. Such in-depth investigations could provide the empirical foundations for constructing an *AI Assessment Literacy* framework that is increasingly refined and grounded in educational reality.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin, in accordance with the national legislation and the institutional requirements.

Author contributions

VV: Writing – original draft, Writing – review & editing. LA: Writing – original draft, Writing – review & editing. PB: Writing – original draft, Writing – review & editing. AB: Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bearman, M., Dawson, P., Ajjawi, R., Tai, J., and Boud, D. (2020). Assessment in a digital world: challenges and recommendations. *Assess. Eval. High. Educ.* 45, 1228–1241. doi: 10.1080/02602938.2020.1772354

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Buchholz, B. A., DeHart, J., and Moorman, G. (2020). Digital citizenship during a global pandemic: moving beyond digital literacy. *J. Adolesc. Adult. Lit.* 64, 11–17. doi: 10.1002/jaal.1076

Çela, E., Fonkam, M., and Potluri, R. M. (2024). Risks of AI-assisted learning on student critical thinking. *Int. J. Risk Conting. Manag.* 12, 1–19. doi: 10.4018/IJRCM.350185

Celik, I. (2023). Towards intelligent-TPACK: an empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Comput. Hum. Behav.* 138:107468. doi: 10.1016/j.chb.2022.107468

Chai, C. S., Wong, L. H., and King, R. B. (2016). Surveying and modeling students' motivation and learning strategies for mobile-assisted seamless Chinese language learning. *Educ. Technol. Soc.* 19, 170–180.

Chan, C. K. Y., and Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int. J. Educ. Technol. High. Educ.* 20:43. doi: 10.1186/s41239-023-00411-8

Gruenhagen, J. H., Sinclair, P. M., Carroll, J.-A., Baker, P. R. A., Wilson, A., and Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: students' perceptions and use of chatbots for assistance with assessments. *Comput. Educ. Artif. Intell.* 7:100273. doi: 10.1016/j.caeai.2024.100273

Holmes, W., and Tuomi, I. (2022). State of the art and practice in AI in education. *Eur. J. Educ.* 57, 542–570. doi: 10.1111/ejed.12533

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* 103:102274. doi: 10.1016/j.lindif.2023.102274

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., and Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *Int. J. Manag. Educ.* 21:100790. doi: 10.1016/j.ijme.2023.100790

Liu, X. (2024). Navigating uncharted waters: teachers' perceptions of and reactions to AI-induced challenges to assessment. *Asia Pac. Educ. Res.* 34, 711–722. doi: 10.1007/s40299-024-00890-x

Long, D., and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–16.

Luckin, R., Cukurova, M., Kent, C., and du Boulay, B. (2022). Empowering educators to be AI-ready. *Comput. Educ. Artif. Intell.* 3:100076. doi: 10.1016/j.caeai.2022.100076

Memarian, B., and Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: a review and entity-relationship (ER) analysis. *Comput. Hum. Behav. Artif. Hum.* 2:100053. doi: 10.1016/j.chbah.2024.100053

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Pérez-Sánchez, R. C., and Moret-Bonillo, V. (2023). Human-in-the-loop machine learning: a state of the art. *Artif. Intell. Rev.* 56, 3005–3054. doi: 10.1007/s10462-022-10246-w

Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W., and Kersting, K. (2025). Human-in-the-loop or AI-in-the-loop? Automate or collaborate? *Proceedings of the AAAI conference on artificial intelligence*, 39, 28594–28600.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., and Qiao, M. S. (2021). Conceptualizing AI literacy: an exploratory review. *Comput. Educ. Artif. Intell.* 2:100041. doi: 10.1016/j.caeai.2021.100041

OECD (2023). AI and the future of skills, volume 2: methods for evaluating AI capabilities. Paris: OECD Publishing.

Salinas-Navarro, D., Vilalta-Perdomo, E., Michel-Villarreal, R., and Montesinos, L. (2024). Designing experiential learning activities with Generative Artificial Intelligence tools for authentic assessment. *Interact. Technol. Smart Educ.* 21, 708–734. doi: 10.1108/ITSE-12-2023-0236

Selwyn, N. (2016). Is technology good for education? Cambridge, UK: Polity Press.

UNESCO (2023) in Guidance for generative AI in education and research. eds. F. Miao and W. Holmes (Paris: UNESCO).

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *Int. J. Educ. Technol. High. Educ.* 21:15. doi: 10.1186/s41239-024-00448-3

Widder, D. G., Whittaker, M., and West, S. M. (2024). Why 'open' AI systems are actually closed, and why this matters. *Nature* 635, 827–833. doi: 10.1038/s41586-024-08141-1

Williams, R. (2024). Impact.AI: democratizing AI through K-12 artificial intelligence education. (doctoral dissertation). Massachusetts Institute of Technology, MIT theses. Available online at: https://dspace.mit.edu/handle/1721.1/153676 (Accessed September 23, 2025).

Zhai, C., Wibowo, S., and Li, L. D. (2024). The effects of overreliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn. Environ.* 11:28. doi: 10.1186/s40561-024-00316-7