Beyond Tools: Evaluating the impact of AI and foundational AI literacy

Marina Bertolini*

Riassunto

L'intelligenza artificiale sta trasformando rapidamente l'istruzione superiore, influenzando le pratiche didattiche, i processi di apprendimento e le competenze richieste ai laureati. Il contributo esamina le sfide metodologiche e operative legate alla valutazione dell'impatto educativo dell'IA, evidenziando i limiti degli approcci valutativi tradizionali in un contesto tecnologico in rapida evoluzione. Si sostiene che i cambiamenti negli esiti di apprendimento, la crescente diffusione dell'IA e l'assenza di quadri di valutazione stabili rendano difficili confronti significativi sull'efficacia educativa. In risposta, il lavoro propone lo sviluppo sistematico di competenze di base sull'IA quale obiettivo istituzionale condiviso, per promuovere un utilizzo critico, etico e consapevole dell'IA e favorire pratiche valutative più coerenti tra le discipline.
Parole chiave: Intelligenza artificiale; AI literacy; Istruzione superiore; Esiti di apprendimento; Valutazione educativa

Abstract

Artificial intelligence is rapidly transforming higher education, affecting teaching practices, learning processes, and the competencies expected of graduates. This paper examines the methodological and practical challenges involved in evaluating AI's educational impact, highlighting the limitations of traditional assessment approaches in a rapidly evolving technological context. It argues that changes in learning outcomes, widespread AI adoption, and the

* Department of Statistical Sciences, University of Padova.

Excellence and Innovation in Learning and Teaching (ISSNe 2499-507X), 2026, 1 Special Issue

Doi: 10.3280/exioa1-2026oa23978

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial – No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

absence of stable evaluation frameworks hinder meaningful comparisons of educational effectiveness. In response, the paper proposes the systematic development of foundational AI literacy as a shared institutional objective, enabling universities to foster critical, ethical, and informed AI use while supporting more consistent assessment across disciplines.
Keywords: Artificial intelligence; AI literacy; Higher education; Learning outcomes; Educational assessment

Article submitted: 03/07/2026; accepted: 03/09/2026

AI in the learning environment

The rapid diffusion of Artificial Intelligence (AI) tools in higher education is reshaping both teaching and learning practices. While professors increasingly employ AI to support content creation, lesson planning, feedback generation, and assessment design, students use the same technologies to enhance studying, complete assignments, access individualized support, and receive immediate explanations or feedback. This convergence transforms the educational process into an ecosystem in which both instructors and learners interact with AI, often simultaneously and for complementary purposes (Ogunleye et al., 2024).

As AI becomes embedded in everyday academic activities, it challenges traditional assumptions about knowledge acquisition and the role of educators; these developments also raise important questions regarding the evolving nature of the educational relationship (Xu & Ouyang, 2022), the effectiveness of AI-enabled learning environments (Sailer et al., 2024), and the responsibilities of higher education institutions in guiding the adoption of these technologies (UNESCO, 2023; OECD, 2026).

Beyond their direct influence on teaching and learning, AI systems are also changing expectations about what graduates should know and be able to do: in this sense, it is necessary to work on the revision of learning outcomes and educational objectives. In particular, universities shall consider new skills required in the changing labour market, such as critical thinking, problem formulation, source evaluation, ethical reasoning, and effective human-AI collaboration.

Challenges in evaluating the impact of AI

Understanding the impact of AI on learning outcomes is particularly challenging. Rigorous evaluation typically requires experimental or quasi-

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

experimental research designs, adequate sample sizes, reliable measures, and the ability to establish causal relationships between interventions and outcomes (Shadish et al., 2002).

These requirements are difficult to satisfy in contemporary educational settings, particularly while AI technologies are evolving rapidly and are widely accessible both inside and outside the classroom. Students may use different AI tools independently, making it difficult to control exposure to the intervention or distinguish institutional initiatives from spontaneous adoption.

Effective evaluation also requires baseline data and sufficiently long observation periods. However, these are difficult to establish in the current AI landscape, where tools evolve rapidly and differ significantly from one academic year to the next, making longitudinal comparisons particularly challenging: AI platforms are continuously updated, meaning that the intervention itself may change during the course of a study, further reducing the stability and reproducibility of empirical findings (Bertolini, 2026).

Practical and ethical constraints also affect a proper experimental design. Educational institutions cannot easily prevent some students from accessing AI simply to create a control group, nor can they fully monitor the intensity and quality of AI use outside formal learning environments.

A further complication concerns the definition of educational outcomes themselves. As AI increasingly supports tasks within the learning process, institutions must reconsider which competencies remain essential and how they should be assessed. Skills traditionally associated with academic performance, such as information retrieval, drafting, or routine problem solving, may become less central, while abilities related to critical judgment, verification of AI-generated content, creativity, and decision-making gain greater importance. This evolution makes comparisons of student performance before and after the introduction of AI less meaningful, since the learning outcomes themselves may have changed and are therefore no longer directly comparable. Simply put, we do not have truly comparable *ex ante* data, and the educational context in which previous evidence was generated has substantially changed.

An additional challenge lies in disentangling improvements in productivity from improvements in learning. Students may complete assignments more efficiently with AI assistance, but greater efficiency does not necessarily imply deeper conceptual understanding and long-term retention.

Limitations of current approaches

Revising and defining learning outcomes is a complex process that requires substantial effort from the academic community, institutional coordination, and

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

considerable time: this work cannot be implemented as fast as the technological innovation. As a result, many institutions are experimenting with AI while continuing to evaluate students according to frameworks that were designed before the widespread availability of generative AI.

The existing literature reflects this transitional phase. It contains numerous AI-related initiatives and pilot projects (Bond et al., 2024), but many lack robust evaluation frameworks, longitudinal designs, or appropriate comparison groups. Although experimentation is both necessary and valuable during periods of technological change, it inevitably limits the production of reliable and generalizable evidence. Results obtained in one discipline, institution, or educational context are often difficult to transfer to others because pedagogical objectives, teaching methods, and patterns of AI adoption differ substantially.

Moreover, the most important educational consequences of AI may only become apparent over time, as sustained interaction with AI has the potential to influence learners' cognitive engagement, self-regulation, and patterns of cognitive offloading, effects that require longitudinal investigation (Skulmowski, 2023; Bond et al., 2024) rather than performance in individual courses. Some consequences may be positive, such as improved self-regulated learning or greater access to personalized support, while others may involve overreliance on AI, reduced independent problem-solving, or changes in students' motivation. These long-term effects require longitudinal observation that extends well beyond a single academic semester.

Assessment is further complicated by disciplinary diversity. AI can support programming, writing, data analysis, design, or language learning in fundamentally different ways, making it unlikely that a single evaluation framework will adequately capture its educational impact across all fields. Consequently, institutions require evaluation strategies that combine general indicators with discipline-specific measures while remaining sufficiently flexible to accommodate continuous technological change.

A coordinated approach: a proposal

Given these challenges, a promising institutional objective is the systematic development of foundational AI literacy among all students, regardless of their discipline. AI literacy extends well beyond the effective use of AI tools, as it includes the understanding the basic principles underlying AI systems, recognizing their biases and limitations. This helps in critically evaluating AI-generated outputs, which is one of the relevant skills students should be trained on. Also, protecting privacy and intellectual integrity shall become a priority, making people to understand the value of data they are providing (data value is

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

a concept that has wider impact, see Bertolini, 2024). Being able to make informed decisions about when and how AI should be incorporated into academic and professional tasks is relevant to undertake efficient and cost-effective decisions.

The ability to make informed decisions is a key factor in preparing students to engage with AI responsibly. Rather than encouraging either unconditional adoption or complete avoidance of AI, higher education should enable students to understand the trade-offs associated with different uses of these technologies and to exercise informed professional judgment. Such competencies are likely to remain valuable even as specific AI tools evolve or become obsolete.

Several leading universities, including the University of Oxford, MIT, Stanford University, Harvard University, ETH Zurich, Imperial College London, and the National University of Singapore, have adopted institution-wide strategies that combine governance, faculty development, AI literacy, ethical guidance, and the dissemination of best practices. These initiatives reflect a shift from regulating AI use toward fostering an institutional culture of responsible AI adoption.

This perspective also offers practical advantages for educational evaluation. General training on AI foundations can be designed around relatively stable learning objectives, allowing institutions to develop common assessment instruments and compare outcomes across disciplines and over time. Measuring improvements in AI literacy, critical evaluation skills, and responsible use may therefore prove more feasible than attempting to isolate the impact of individual AI applications on course-specific performance. Jin et al. (2025) found proactive approaches by universities towards generative AI integration, emphasizing academic integrity, teaching and learning enhancement, and equity: it seems still missing a systematic program that includes testing of the results of these programs.

The central question is not simply whether AI improves learning outcomes in the short term, but how it reshapes the competencies, responsibilities, and purposes of education in an AI-enabled society. Universities must therefore move beyond evaluating AI as a technological innovation and instead consider its implications for curriculum design, assessment practices, institutional governance, and graduate preparedness.

Developing shared frameworks for AI literacy and its assessment may represent an important first step toward producing cumulative evidence while ensuring that higher education continues to fulfil its educational mission in an increasingly AI-mediated world.

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

References

Bertolini M. (2026). AI literacy and new learning strategies. In: W. Leal Filho, T. Wall, H. Müller, P. Yang & A. Klimczuk (Eds.). *Artificial Intelligence in Sustainability Teaching and Research.* World Sustainability Series. Cham: Springer.

Bertolini M., Castellini M., Moretto M., & Vergalli S. (2024). Smart grids in the renewable energy community era: New markets and the role of information. In: *Sustainable Development of Renewable Energy: Latest Advances in Production, Storage, and Integration* (pp. 401-412). Amsterdam: Elsevier. Doi: 10.1016/B978-0-443-13613-9.00006-4.

Bond M., Khosravi H., De Laat M., Bergdahl N., Negrea V., Oxley E., Pham P., Chong S. W., & Siemens G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education,* 21, 4. Doi: 10.1186/s41239-023-00436-z.

Jin Y., Yan L., Echeverria V., Gašević D., & Martinez-Maldonado R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8, 100348. Doi: 10.1016/j.caeai.2024.100348.

OECD (2026). Policies supporting responsible and systematic GenAI adoption in higher education. *OECD Education Spotlights,* No. 23. Doi: 10.1787/c4e5621f-en.

Ogunleye B., Zakariyyah K. I., Ajao O., Olayinka O., & Sharma H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences,* 14(6), 636. Doi: 10.3390/educsci14060636.

Sailer M., Maier R., Berger S., Kastorff T., & Stegmann K. (2024). Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education. *Learning and Individual Differences,* 112, 102446. Doi: 10.1016/j.lindif.2024.102446.

Shadish W. R., Cook T. D., & Campbell D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Skulmowski A. (2023). The cognitive architecture of digital externalization. *Educational Psychology Review,* 35, 101. Doi: 10.1007/s10648-023-09818-1.

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

Xu W., & Ouyang F. (2022). A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies,* 27(3): 4195-4223. Doi: 10.1007/s10639-021-10774-y.

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org