Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6 n. 3 (2026)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Designing engagement in artificial intelligence education: algorithmic citizenship, agency and authentic tasks

Giovanna Amendola [1,] *, Fabio Orecchio [2] and Pio Alfredo Di Tore [1]

[1] University of Cassino and Southern Lazio, Cassino, Italy; giovanna.amendola-3@studentmail.unicas.it; pioalfredo.ditore@unicas.it
[2] Unipegaso; Fabio.orecchio@unipegaso.it
* Correspondence: giovanna.amendola-3@studentmail.unicas.it

Abstract: Artificial intelligence (AI) has entered pupils' daily lives before it has been incorporated into school curricula, yet schools tend to view it either as a risk to be contained or as a productivity tool to be mastered, neglecting pupils' ability to understand, question and transform the algorithmic systems that shape their experience. This theoretical contribution proposes a conceptual foundation for an engaging AI education, in which algorithmic citizenship is practised through authentic tasks rather than taught as content. The study conducts a critical, integrative literature review that brings together research on school engagement, self-determination theory, authentic learning, AI literacy, and critical data studies, and reinterprets the criteria for authentic learning in light of the distinctive characteristics of algorithmic objects: opacity, mutability, reflexivity, and distributed effects. The analysis identifies three shifts in the classical criteria (from relevance to the real world to relevance to the lived world; from a multiplicity of perspectives to political plurality; and from integrated assessment to intervention) and a framework of design principles structured across three levels, relating to motivational needs, scaffolding and persistence, illustrated by three task scenarios. The paper argues that strong engagement arises when students recognise that the task concerns them and that their own actions can influence it, and it outlines its limitations, ethical issues and hypotheses for empirical research.

Keywords: algorithmic citizenship; AI literacy; student engagement; agency; authentic tasks

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial intelligence (AI) became part of students' everyday lives long before schools incorporated it into curricula. Recommendation systems, conversational assistants, content filters and generative tools are increasingly shaping what young people read, listen to, buy and produce, often without them realising it. Research into everyday encounters with algorithms shows that users develop an 'algorithmic imagination', that is, a set of intuitions, feelings and strategies regarding how platforms work, which shapes their behaviour even in the absence of technical knowledge of the underlying processes (Bucher, 2017). Students, therefore, do not arrive at school as a blank slate when it comes to AI: they bring with them a history of interactions, partial theories and, often, the feeling that these systems are simply

part of the environment, rather than being artefacts designed by someone for a specific purpose.

Faced with this situation, schools tend to oscillate between two opposing approaches. In the first, AI is viewed as a risk to be contained through bans, codes of conduct and rules of use, particularly in relation to generative tools and academic integrity. In the second, AI is treated as a technical tool to be mastered for productivity, reducing education to training in prompts and applications. Although their approaches are diametrically opposed, both leave the central issue unaddressed: the student's ability to understand, question and transform the algorithmic systems that shape their experience. In the defensive stance, the learner is a subject to be protected; in the instrumental stance, a user to be trained. In neither case are they positioned as an agent in relation to the systems themselves (Selwyn, 2019).

Meanwhile, the policy framework has shifted towards recognising AI literacy as a requirement for citizenship. The European Digital Competence Framework explicitly includes interaction with AI systems amongst the knowledge, skills and attitudes required of citizens (Vuorikari et al., 2022); UNESCO (2024) has proposed a competence framework for students, centred on a human-centred perspective and the ethics of AI; the European Regulation on Artificial Intelligence introduces AI literacy as an obligation for providers and users of AI systems (Regulation (EU) 2024/1689, 2024). Research on AI literacy has mapped the competencies involved, ranging from recognising AI and understanding how it works to evaluating its outputs and ethical implications (Long & Magerko, 2020; Ng et al., 2021). These frameworks, however, say relatively little about how to design learning experiences in which such skills are practised rather than merely stated, and rarely address the motivational issue: why should students invest effort in understanding systems that seem to work perfectly well without being understood?

This paper addresses this gap. The aim is to propose a conceptual framework for engaging AI education, in which algorithmic citizenship is not taught as subject-d content, but practised through authentic tasks that require students to act on matters that directly concern them. This paper is theoretical in nature: it does not present empirical data, but proposes a framework of design principles intended to be tested in future research. In particular, it addresses three questions: How should the established criteria for authentic learning be reinterpreted when the subject of the task is an algorithmic system? What design principles can support students' engagement and agency in such tasks? What implications does this have for research, practice and educational policy? The main argument is that the strongest engagement arises when students recognise in the task an issue that concerns them and, through their own actions, the possibility of understanding and modifying it.

2. Materials and methods

This study adopts a critical review of the conceptual literature. It is neither a systematic review aimed at estimating effects, nor a scoping review aimed at mapping the volume of available studies, but rather an integrative analysis, in the sense described by Torraco (2005), in which representative literature on a topic is examined, critically assessed and synthesised to generate new frameworks and perspectives. With reference to the typology proposed by Jaakkola (2020), the study combines a theory synthesis design, which integrates concepts from different research strands into a coherent framework, with a theory adaptation design, in

which an existing theory, in this case the authentic learning model, is revisited through the lens of a different domain.

Four research traditions that are usually treated separately have been brought into dialogue: studies on school engagement; the psychology of motivation, with particular reference to self-determination theory; instructional design for authentic learning; and critical data studies, together with the literature on AI literacy. The sources were selected on the basis of three criteria: (a) their status as foundational or widely accepted references within each tradition; (b) their conceptual relevance to the relationship between students, tasks and algorithmic systems; (c) their ability to provide operationalised constructs for design or critical lenses for interpreting the specific nature of the algorithmic object. European and international policy documents were considered as context rather than as the subject of analysis.

The analytical procedure was divided into three stages. Firstly, the core constructs of each tradition were identified and their mutual compatibility was examined (Section 3.1). Secondly, the criteria for authentic learning (Herrington et al., 2010) were re-examined in the light of the distinctive properties of algorithmic systems – understood as objects of investigation – in order to identify the points at which these criteria require reinterpretation (Section 3.2). Thirdly, the findings of this re-examination were translated into design principles organised by level, the internal consistency of which was examined through the construction of illustrative task scenarios (Sections 3.3 and 3.4). The scenarios are analytical tools developed by the authors and are not accounts of classroom experiments. As the study did not involve human participants or entail data collection, approval from an ethics committee was not required.

3. Results

3.1. Four theoretical dimensions

The first axis concerns the nature of engagement. Following Fredricks et al. (2004), engagement is conceived as a multidimensional construct comprising behavioural components (participation, commitment, persistence), emotional components (interest, a sense of belonging, affective reactions to the task) and cognitive components (investment in understanding, use of deep strategies, willingness to tackle complexity). Reeve and Tseng (2011) have also proposed an agentic dimension, namely the student's intentional and proactive contribution to the learning activity. From this perspective, engagement is not a stable trait of the student, but a relationship that is constructed between the individual, the task and the context; it is therefore, at least in part, a matter of design.

The second axis provides a motivational framework. According to self-determination theory, sustained engagement depends on the fulfilment of three basic psychological needs: autonomy, competence and relatedness (Ryan & Deci, 2000). Tasks that frustrate these needs produce, at best, external compliance, whilst those that support them foster the internalisation of goals and sustained commitment. The theory is particularly relevant in this context because algorithmic systems are designed to capture attention through mechanisms that act on motivation: an education system that aims to foster agency cannot rely on the very same extrinsic levers that it seeks to make visible.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

The third pillar defines the design criteria. The authentic learning model (Herrington et al., 2010) identifies a set of conditions in which tasks resemble the complex activities of real-world practice and result in transferable learning: real-world relevance, ill-defined tasks requiring prolonged investigation, access to multiple perspectives, collaboration, reflection, articulation and integrated assessment. The model has proven robust across different contexts and technologies and provides the operational framework on which this framework is based.

The fourth axis provides both a map of the expected skills and a critical lens. The AI literacy frameworks describe what learners should know and be able to do: recognise AI, understand its main techniques and its reliance on data, evaluate its outputs, and reflect on its ethical implications (Long & Magerko, 2020; Ng et al., 2021). Critical studies on the datafication of education show that algorithmic systems are not neutral, as they incorporate choices, interests and power asymmetries, and increasingly govern educational processes themselves (Williamson, 2017). Studies on algorithmic opacity add that the inscrutability of such systems stems from various sources, including trade secrets, technical illiteracy and the disconnect between machine learning operations and forms of human reasoning (Burrell, 2016). Finally, the literature on personal data literacies emphasises that understanding data is inseparable from the ability to act upon it in one's own life (Pangrazio & Selwyn, 2019).

3.2. The algorithmic object and three shifts in the criteria for authentic learning

The central focus of this study is to re-examine the criteria for authentic learning in the light of the specific nature of the algorithmic object. An authentic task involving AI exhibits characteristics that classical models did not anticipate. The object is opaque by design, as its internal logic is not accessible to the user and is only partially accessible even to its designers (Burrell, 2016). It is changeable, because models are updated, retrained and personalised over time, so that the object observed today may behave differently tomorrow. It is reflexive because it acts upon the user whilst observing them: a recommendation system adapts its suggestions to the behaviour of the observer. Finally, its effects are distributed because they emerge from the interaction between code, data, institutions and users, none of which any single actor fully controls (Kitchin, 2017). This opacity is not an obstacle to be removed before learning can begin: it is the very ground on which learning takes place.

This reinterpretation highlights three shifts from the traditional criteria (Table 1). The criterion of relevance to the real world becomes relevance to the lived world. It is not enough for the task to refer to a plausible professional context: it must relate to systems with which students already have a history of interaction, so that the algorithmic imagination each of them has developed (Bucher, 2017) becomes the starting point for the investigation rather than an implicit residue.

The principle of multiple perspectives takes on a political dimension. An algorithmic system brings together, and often pits against one another, the interests of those who design it, those who use it, those affected by its decisions, and those who regulate it. Recognising these perspectives means acknowledging that defining what a system should optimise is a matter of value judgements and not merely one of technical efficiency (Eubanks, 2018; Mitchell et al., 2021).

Finally, the concept of integrated assessment takes on further significance. The outcome of the task is not merely an artefact to be assessed, but an intervention that can alter, albeit to a limited extent, the relationship between the student and the system under investigation, for example through a request for clarification addressed to a supplier, a public report or a proposal for rules addressed to an institution. In this sense, the task does not simulate civic action: it constitutes a small but real instance of it.

Table 1. A reinterpretation of some criteria of authentic learning in the light of the algorithmic object.

Criterion	Traditional interpretation (Herrington et al., 2010)	Reinterpretation for AI education
Relevance	Relevance to the real world: tasks reflect professional or everyday practices	Relevance to the real world: the tasks relate to systems with which pupils already interact
A variety of perspectives	Access to different viewpoints and sources	Political pluralism: designers, users, stakeholders and regulators have divergent interests
Integrated assessment	Assessment embedded within the task and its outputs	Intervention: the product influences the student–system relationship (requests, feedback, proposed rules)
Poorly defined tasks	Problems requiring prolonged investigation	Opacity as a context: the object is opaque, changeable, reflective and produces distributed effects

Note. The first three lines correspond to the displacements discussed in Section 3.2; the fourth summarises the properties of the object that give rise to them.

3.3. A three-level framework of design principles

A re-examination reveals a framework of design principles structured across three levels: motivational needs, forms of scaffolding and strategies to support persistence (Table 2).

3.3.1. Motivational needs

Autonomy is fostered when students can choose which system to investigate from amongst those they actually use, formulate their own questions and decide how to present their findings. Competence is fostered when the task is designed in such a way that the analysis yields verifiable findings, for example, identifying a pattern in the behaviour of a recommendation system, rather than merely gaining a general awareness; in terms of self-determination theory, competence requires feedback that confirms one's effectiveness (Ryan & Deci, 2000). Relationship is fostered when the work is carried out in a group and the recipients of the output are real people, whether they be classmates, families, a local authority or an online community. The three needs reinforce one another: choice without the means to

produce results leads to frustration, while results without an audience risk remaining a mere academic exercise.

3.3.2. Opacity scaffolding

As the subject matter is opaque, students need conceptual and practical tools that make visible what the system conceals. Investigative journalism and critical studies of algorithms have developed methods that can be adapted to educational contexts, such as reverse engineering based on the systematic variation of inputs and the observation of outputs (Diakopoulos, 2015; Kitchin, 2017). Within the proposed framework, these methods take four forms: observation protocols to record a system's outputs under controlled conditions; schemes for backwards reconstruction of the plausible variables that govern it; shared vocabularies for naming biases, feedback loops, objective functions and training data; and simplified models that allow the behaviour of a classifier to be simulated without requiring advanced programming skills. In line with the original notion of scaffolding (Wood et al., 1976), these supports do not simplify the subject matter, but organise access to its complexity and are gradually withdrawn as students develop investigative autonomy.

3.3.3. Support for retention

Authentic tasks involving AI are lengthy, uncertain and sometimes frustrating, because the systems resist interpretation. The literature suggests that persistence depends on the perception of progress, the opportunity to make interim work visible, and the presence of moments of recognition (Fredricks et al., 2004). The framework therefore proposes short cycles of inquiry with public feedback, research diaries documenting hypotheses and their refutations, and formative assessment that rewards the quality of questions, the transparency of the process and the soundness of conclusions (Black & Wiliam, 1998). These measures make progress tangible and reduce the risk that the task's inherent uncertainty might lead to premature disengagement, particularly amongst less confident students, for whom an inconclusive result might otherwise be experienced as a personal failure.

Table 2. Design principles for engaging AI education, by level.

Level	Design principles	Key theoretical references
Motivational needs	Choice of system and questions (autonomy); verifiable discoveries (competence); group work and real-world audience (relatedness)	Ryan & Deci (2000); Reeve & Tseng (2011)
Scaffolding	Observation protocols; retrospective reconstruction of variables; shared vocabulary; simplified models; gradual withdrawal of support	Wood et al. (1976); Diakopoulos (2015); Kitchin (2017)
Persistence	Short cycles of inquiry with public feedback; research diaries; formative assessment of questions and process	Fredricks et al. (2004); Black & Wiliam (1998)

3.4. Illustrative task scenarios

Three scenarios illustrate how to apply the principles. As noted in Section 2, these are illustrative constructs and not accounts of experiments that have been carried out.

In the first scenario, students reconstruct the recommendation systems they use on a daily basis. Working in groups, they create controlled profiles, record the content suggested over a defined period, formulate hypotheses about the variables that influence the suggestions, and compare results across groups. The outcome is a reasoned analysis of the logic behind personalisation and its effects on their own experience, which is presented to their classmates and families. This scenario fosters autonomy (choice of platform), competence (identification of patterns) and social interaction (a real audience), and makes the reflexivity of the subject explicit, as the system changes whilst it is being observed.

In the second scenario, students analyse automated decisions in public services, from the awarding of marks in ranking lists to the allocation of resources. They identify what data is used, who benefits or is adversely affected, and what forms of appeal exist, drawing on documented cases of automated decisions in welfare administration (Eubanks, 2018). The outcome is a transparency proposal addressed to a real institutional body, such as a local council department, that turns the integrated assessment criterion into an act of civic participation.

In the third scenario, students redesign fair classification criteria. Starting from a case where a classifier produces discriminatory outcomes, they discuss alternative definitions of fairness, simulate their effects on simplified data, and negotiate a criterion that makes explicit the value judgements it incorporates. The literature on algorithmic fairness shows that different definitions may be mutually incompatible, meaning that every choice involves trade-offs (Mitchell et al., 2021); for students, this realisation transforms what appears to be a technical issue into a political one.

In all three cases, the task does not ask students to learn something about AI, but to do something with AI and to challenge its opacity, in a situation with tangible consequences.

4. Discussion

The hypothesis underpinning this paper is that the strongest form of engagement arises when students recognise in a task an issue that concerns them and, through their own actions, the possibility of understanding and changing it. In this interpretation, engagement is not a motivational effect to be achieved through gamification or technological innovation, but the consequence of a relationship of responsibility between the individual and an object that has taken on a place in their life. This position is consistent with the conception of engagement as a relationship rather than a trait (Fredricks et al., 2004) and with the agentic dimension described by Reeve and Tseng (2011), but it extends both, suggesting that, when the object is algorithmic, agency must be exercised not only within the educational activity, but also beyond it, in relation to systems and institutions.

This leads to a conception of algorithmic citizenship as a practice rather than as declarative knowledge. One is an algorithmic citizen not because one knows how an algorithm works, but because one is able to question it, challenge its outcomes and participate in defining the rules that govern it. This vision complements existing

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

frameworks on AI literacy (Long & Magerko, 2020; UNESCO, 2024), which describe competencies and focus on the conditions under which these competencies are put into practice. It also responds to the critical literature on datafication (Williamson, 2017), translating its analytical insights into a pedagogical approach: the power asymmetries embedded in algorithmic systems become the subject matter of the tasks rather than a background to be recognised.

This paper has limitations that must be acknowledged. As a theoretical work, it does not provide empirical evidence regarding the effectiveness of the proposed principles, which remain hypotheses to be tested. The three scenarios are illustrative constructs and not accounts of experiments. Furthermore, the framework presupposes organisational conditions, such as ample time, teachers willing to take on the role of co-investigators, and a certain degree of institutional tolerance towards the uncertainty of results, which are not guaranteed in all school contexts. Finally, the decision to prioritise tasks that operate on real-world systems raises ethical issues concerning the protection of pupils' data, particularly when controlled profiles are created on commercial platforms, and the responsibility for actions directed at external actors. The framework highlights these issues but does not resolve them: they require specific development, including age-appropriate protocols and agreements with schools and families.

This paper has implications on three levels. For research, the framework provides operationalisable variables for future empirical studies: the dimensions of engagement can be observed in relation to specific design choices, and the distinction between the three levels makes it possible to identify which components of the task contribute most to persistence. Design-based research appears to be a suitable methodology for testing and refining these principles in real-world contexts. In terms of practice, the framework suggests that AI education does not necessarily require new subjects, but rather a different organisation of activities, which may span civic education, mathematics, science and the humanities. Finally, in terms of policy, the study calls for engagement to be viewed not as an indicator of satisfaction, but as a prerequisite for the effective exercise of digital rights, with implications for the design of guidelines and curricula, including the implementation of the provisions on AI literacy introduced at European level.

5. Conclusions

This paper proposes a conceptual framework for engaging AI education, integrating research on engagement, self-determination theory, authentic learning, and critical data studies. By re-examining the criteria of authentic learning in the light of the opacity, mutability, reflexivity and distributed effects of algorithmic systems, it has identified three shifts: towards relevance to the lived world, towards the political plurality of perspectives, and towards assessment as intervention; and has translated these into a framework of design principles across three levels. Restoring to young people the agency that digital everyday life tends to take away is, from this perspective, both the educational objective and the quality criterion for design.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Herrington, J., Reeves, T. C., & Oliver, R. (2010). *A guide to authentic e-learning*. Routledge.

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, *8*, 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualising AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, *21*(2), 419–437. https://doi.org/10.1177/1461444818799523

Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, *36*(4), 257–267. https://doi.org/10.1016/j.cedpsych.2011.05.002

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (: Artificial Intelligence Act). (2024). *Official Journal of the European Union*, L 2024/1689. http://data.europa.eu/eli/reg/2024/1689/oj

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, *4*(3), 356–367. https://doi.org/10.1177/1534484305278283

UNESCO. (2024). *AI competency framework for students*. UNESCO.

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens. With new examples of knowledge, skills and attitudes*. Publications Office of the European Union. https://doi.org/10.2760/115376

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. SAGE.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.14