Generative artificial intelligence and educational autonomy: historical metaphors and ethical principles for pedagogical transformation

Inteligencia artificial generativa y autonomía educativa: metáforas históricas y principios éticos para la transformación pedagógica

Marc Alier-Forment - *Universitat Politècnica de Catalunya, UPC (Spain)*
María José Casañ-Guerrero - *Universitat Politècnica de Catalunya, UPC (Spain)*
Juan Antonio Pereira-Varela - *Universidad del País Vasco, UPV/EHU (Spain)*
Francisco José García-Peñalvo - *Universidad de Salamanca, USAL (Spain)*
Faraón Llorens-Largo - *Universidad de Alicante, UA (Spain)*

ABSTRACT

This article examines the integration of generative artificial intelligence in education from a critical, historical, and ethical perspective. It highlights growing concerns about the opacity of current artificial intelligence tools, particularly in learning systems. The study adopts a metaphor-based approach to explore how technological narratives influence the adoption of educational innovations. It reviews historical metaphors used to describe educational technologies, from Multivac and Matrix to the free software Bazaar and the App Store, and proposes new conceptual frameworks that may better reflect the current context in which artificial intelligence is entering the educational sphere. Based on this metaphorical analysis, the article outlines seven fundamental ethical principles for the safe adoption of generative artificial intelligence in education, focusing on privacy, pedagogical alignment, human oversight, and technological transparency. These principles are illustrated through a practical application: the LAMB (Learning Assistant Manager and Builder) environment, an open-source software framework that enables the ethical and contextualized design of artificial intelligence-based learning assistants. The article presents real-world cases of LAMB implementation in higher education, including a controlled experience with students that demonstrates significant improvements in student autonomy and pedagogical coherence. Finally, it emphasizes how LAMB embodies the proposed ethical principles and responds to the identified critical metaphors, offering a model for technology integration centered on teacher autonomy, alignment with institutional values and practices, and meaningful student learning that prioritizes pedagogical control over technological determinism.

Keywords: educational technology; generative artificial intelligence; adaptive learning; learning assistant; ethics; instructional design.

RESUMEN

Este artículo analiza la integración de la inteligencia artificial generativa en educación desde una perspectiva crítica, histórica y ética. Se identifica una creciente preocupación por la opacidad de las herramientas de inteligencia artificial actuales, especialmente en sistemas de aprendizaje. El trabajo utiliza un enfoque basado en metáforas para entender cómo las narrativas tecnológicas influyen en la adopción de innovaciones educativas. Se revisan metáforas históricas en las tecnologías aplicadas a la educación, desde Multivac y Matrix hasta el Bazar del software libre y la App Store, y se proponen nuevas imágenes conceptuales que podrían aplicarse al contexto actual en el que la inteligencia artificial irrumpe en la educación. A partir de este análisis metafórico, se plantean siete principios éticos para una adopción segura de la inteligencia artificial generativa en educación, centrados en la privacidad, la alineación pedagógica, la supervisión humana y la transparencia tecnológica. Estos principios se ejemplifican con el entorno LAMB (Learning Assistant Manager and Builder), un marco de código abierto que permite diseñar asistentes de aprendizaje basados en inteligencia artificial de forma ética y contextualizada. Se presentan casos reales de aplicación de LAMB en educación superior, incluyendo una experiencia controlada con estudiantes que muestran mejoras significativas en autonomía y coherencia pedagógica. Finalmente, se destaca cómo LAMB encarna los principios éticos propuestos y responde a las metáforas críticas identificadas, ofreciendo un modelo de integración tecnológica centrado en la autonomía de los docentes, la alineación con los principios y prácticas de la institución educativa y el aprendizaje significativo de los estudiantes.

Palabras clave: tecnología educativa; inteligencia artificial generativa; aprendizaje adaptativo; asistente de aprendizaje; ética; diseño instruccional.

How to cite: Alier-Forment, M., Casañ-Guerrero, M. J., Pereira-Varela, J. A., García-Peñalvo, F. J., & Llorens-Largo, F. (2026). Generative artificial intelligence and educational autonomy: historical metaphors and ethical principles for pedagogical transformation [Inteligencia artificial generativa y autonomía educativa: metáforas históricas y principios éticos para la transformación pedagógica]. *RIED-Revista Iberoamericana de Educación a Distancia, 29*(1), 9-28.
https://doi.org/10.5944/ried.45536

INTRODUCTION

In September 2024, conversations were held with ten experts in educational innovation and technology regarding the "Manifesto for Safe AI in Education" (Alier, García-Peñalvo, Casañ et al., 2024). These discussions centered around a proposed evaluation framework for educational technologies based on generative artificial intelligence (GenAI). This framework resulted in a questionnaire for evaluating GenAI-based educational initiatives. The associated questionnaire focused on key aspects: data privacy, access control within institutions, and above all, the ability of tools to align with pedagogical objectives. Surprisingly, more than half of the experts considered that no currently available GenAI-based solution would meet these criteria or that it would be possible to meet them.

From these exchanges emerged a relevant observation: many technologists with extensive experience in the educational field do not view GenAI, and particularly adaptive learning systems based on AI that promise to personalize the student experience, as clear and malleable tools, but rather as powerful yet opaque artifacts, difficult to understand or control. This phenomenon led us to propose the metaphor of the *Palantír* (Tolkien, 1954-1955).

In J.R.R. Tolkien's narrative universe, the Palantír is a magical sphere that allows seeing through time and space and communicating with users of other Palantírs. However, using the Palantír also entails significant risks, as illustrated by the corruption of the wizard Saruman, which Sauron exercises through this object. This metaphor, applied to GenAI in education, captures two very present sentiments: fascination with its apparent power of personalization and prediction, and fear or skepticism about who controls it and for what purposes.

We will call this the Palantír metaphor: GenAI understood as a centralized, powerful, inevitable tool, but ultimately beyond the control of its real users (teachers, students, and institutions). This image helps explain an increasingly common attitude in the educational field: accepting the implementation of adaptive AI systems as they come, without deeply understanding how they are trained, what biases they may contain, or how decisions that affect individual learning are made.

For the purposes of this work, we use the term "GenAI-based learning assistants" to refer specifically to conversational tools that provide personalized support to students through natural language processing, distinguishing them from traditional adaptive systems that automatically adjust learning paths based on performance analysis.

The idea of technology viewed as "magic" has been widely discussed in the past. For example, Arthur C. Clarke explained in 1962 that sufficiently advanced technology could become something similar to magic (Clarke, 1962).

However, a distinction is necessary. Although the Palantír metaphor is useful for describing certain current narratives, it also has limitations. Artificial intelligence is not magic, and describing it as such may reinforce a deterministic view of technology that inhibits critical capacity and understanding of it.

Metaphors, when explicitly clarified, are valuable analytical tools. They reflect and influence how technologies are imagined, designed, implemented, and above all, how they are disseminated. As Weller and Hofstadter point out, metaphors and analogies are not just stylistic resources, but frameworks of thought that guide our decisions (Weller, 2022; Hofstadter & Sander, 2018).

For example, the metaphor of "digital natives" shaped educational policies for years, often with erroneous results. Similarly, the idea of an "Uber of education" inspired business models that ignored the complexity of learning (Adell-Segura et al., 2018).

Currently, there is growing pressure to integrate GenAI into educational systems. Faced with this situation, this work presents in section two a critical journey through the history of educational technology, with the objective of analyzing how different innovations have been introduced, what challenges they have posed, and what conceptual frameworks, often in the form of metaphors, have guided their understanding and use. Based on this historical and conceptual analysis, section three identifies fundamental ethical principles that the authors consider necessary for responsible incorporation of GenAI in educational contexts. Subsequently, section four proposes a practical application that translates these principles into concrete guidelines for the design and use of adaptive learning systems based on artificial intelligence. Finally, the article's conclusions are presented.

ANALYTICAL FRAMEWORK: METAPHORICAL EVOLUTION OF EDUCATIONAL TECHNOLOGIES

Throughout history, the introduction of new technologies in the educational field has been guided not only by technical advances, but also by narratives, metaphors, and imaginaries that have given meaning to their adoption and modeled their applications. This section explores how certain metaphors have shaped the evolution of educational technology in recent decades and what lessons can be extracted from this history to ultimately explain the current approach to GenAI (Llorens-Largo, 2019). As Weller (2022) reminds us, "Technology often comes wrapped in its own metaphors" (p. 59), and those initial metaphors can have profound consequences on how that technology develops and is applied. Along these lines, Mason (2018) empirically demonstrates how underlying metaphors structure educational technology discourse, suggesting that while it is necessary to use metaphors to understand new digital phenomena, some of the specific metaphors commonly employed may impede more reflective approaches to conceptualizing and implementing new technologies in the educational field.

This section has been divided into four key moments in the evolution of educational technologies, analyzing the dominant metaphors that accompanied each stage and their implications for learning forms, especially in relation to personalization and adaptation.

From Multivac to the Educational Mainframe: The Metaphor of Education as a Centralized System

Isaac Asimov, in his stories about Multivac, imagined a supercomputer capable of making crucial decisions for all humanity. This vision reflects a powerful idea: technology as a centralized entity that collects data, processes massive information, and directs society's actions (Asimov, 1955-1976).

This centralized and authoritarian vision of computers was also reflected in the first proposals for Computer-Assisted Education (CAE) or Computer-Assisted Instruction. B. F. Skinner, a key figure in behaviorism, proposed in his work *The Technology of Teaching* (Skinner, 1968) that computers could optimize learning through systematic application of behaviorist principles. These systems did not conceive learning as an open or exploratory process, but as a series of trajectories programmed and controlled from a single center.

Applied to GenAI-based learning systems, Multivac represents a vision where the computer system dictates the pace and content presented to the student, relegating adaptability to a rigid supervision model. The optimization of predefined results was prioritized, not genuine personalization based on interests, contexts, or individual motivations.

This metaphor remains relevant today when certain systems tend to prioritize instructional efficiency over student autonomy, reproducing a logic of centralized control rather than self-directed learning.

From Matrix to Educational Pills: The Metaphor of Learning as Instantaneous Knowledge Transfer

In the film The Matrix (Wachowski & Wachowski, 1999), characters acquire complex skills, such as piloting helicopters, through immediate download of programs into their brains. This metaphor of "instantaneous knowledge transfer" became very influential during the CD-ROM era and the first e-learning platforms during the second half of the 1990s.

In education, it translated into the development of closed and standardized learning units, the so-called "educational pills," which aimed to offer rapid, modular, and autonomous training. Under this logic, learning became "consuming" discrete information packages, not actively constructing knowledge.

The relationship with contemporary generative AI tools is clear: many AI-based approaches continue to assume that learning is a matter of administering the appropriate dose of content, adjusted to the student's speed or level, but without necessarily considering the critical, reflective, or social dimension of learning.

The Matrix metaphor feeds the promise of GenAI as a magic solution, but also hides the risks of simplifying learning to a merely receptive, passive, and decontextualized process.

From the Bazaar to the Virtual Campus: The Metaphor of Collaborative Knowledge Construction

The next revolution in educational technology came with the web, and the web cannot be understood without free software culture. This movement was born in MIT laboratories, when Richard Stallman, in response to restrictions imposed by software user licenses (EULA), which were previously shared freely among programmers, published the *GNU Manifesto* (Stallman, 1985) and founded the *Free Software Foundation (FSF)* in 1985. The FSF provided a fundamental legal framework: the GPL license, which allowed developers to share their software while guaranteeing four essential freedoms to users: to use, study, share, and improve the software. This philosophy, summarized in the slogan "free as in freedom, not as in free beer," facilitated collaboration and source code exchange on an unprecedented scale (Ceruzzi, 2003).

In contrast to previous centralized visions, the free software movement can be synthesized in the Bazaar metaphor (Raymond, 2001), which introduced a new paradigm in learning systems: open systems, where multiple actors collaborate, innovate, and adapt solutions in a decentralized and democratic manner.

In his influential essay *The Cathedral and the Bazaar* (Raymond, 2001), Eric S. Raymond observes that free software communities, where knowledge and code are shared and members gain influence based on their contributions rather than their status or position, offer a solution to the "endemic software crisis" described by Pressman in *Software Engineering: A Practitioner's Approach* (Pressman, 1991). This practical application of free software principles generates a split in the movement with the emergence of the *Open Source Initiative (OSI)*. The OSI proposed the term "open source software" as an alternative to "free software," introducing more flexible licenses (such as Apache or MIT) that were more attractive to large companies.

This philosophy inspired the creation of environments like Moodle and Sakai, learning management platforms that, being open source, allowed teachers, students, institutions, and developers to adapt tools to their specific needs. It was not just about consuming content, but about building dynamic, shared, and configurable learning environments.

In this context, the Bazaar metaphor points to the possibility of designing systems that are not black boxes, but open spaces where adaptations are understood, negotiated, and collectively reconfigured. This perspective reinforces the idea that adaptability should not only be algorithmic, but also pedagogical and social.

From the App Store to the Educational Cloud: The Metaphor of Invisible Dependency

With the emergence of smartphones and app stores (Apple's App Store and Google Play) the metaphor that "there's an app for everything" became established. The promise was to have immediate, personalized, and accessible solutions, but in exchange for almost total dependence on proprietary and opaque infrastructures, platforms, that sell security and convenience at the price of control. As Ben Thompson reminds us when explaining the true engine of these dynamics: "the reality is that platforms are not a chicken-and-egg problem: the first thing that matters is users, who attract developers despite obstacles" (Thompson, 2025). In other words, rather than each individual app, the platform itself is the center of power.

In education, there was a proliferation of educational apps whose information and management no longer occurs in systems controlled by institutions, but in "the cloud" (a new metaphor), offering customized learning experiences, but in exchange for a loss of control over data, processes, and personalization criteria. Algorithms decide what content to recommend, what routes to follow, what evaluations to apply, almost always without transparency or real participation from educators or institutions. Note how "the cloud" and the "algorithm" become subjects of analysis and criticism, when they are actually instruments operated by the companies that design and control them.

This metaphor invites us to ask: who adapts for whom? What commercial interests, what cultural or technical biases model the paths that students travel?

The platform metaphor, the App Store, warns that personalized learning can become simulated personalization, predefined by corporate interests, if mechanisms of technological and pedagogical sovereignty are not established.

Alternative Metaphors for AI-Powered Adaptive Learning

While the Palantír metaphor is useful for representing the fascination and fear surrounding the use of artificial intelligence in educational contexts, we need frameworks that allow us to think in more critical and pedagogically useful ways about integrating these technologies. For this purpose, the authors propose two alternative metaphors that can help imagine learning support systems with GenAI that are more open, controlled by educational communities, and focused on meaningful learning: the Lego building blocks set and the Prometheus dilemma.

The Lego building blocks metaphor can be used to explain that, instead of conceiving generative AI as a closed and magical artifact, the Lego metaphor poses a more accessible, modular, and comprehensible vision. Under this approach, artificial intelligence is not an "oracle" that offers automatic and unquestionable learning routes, but a set of reusable pieces that can be combined in multiple ways to design personalized, contextual learning experiences with pedagogical foundations.

Applied to GenAI-based learning tools, this implies that:

- GenAI-based tools can be configured by teachers to respond to the real needs of their students.
- The "algorithm" does not replace the educator, but offers materials with which to build alternative routes.
- Adaptations occur not only based on performance or speed, but also on interests, languages, motivations, and diverse trajectories.

This approach connects with a constructivist and participatory vision of learning, where technology does not replace pedagogical foundations, but amplifies them. In this sense, it can be linked to the *construction* metaphor, proposed by Weller (2022), which understands educational technology as a process in which users (teachers and students) actively construct their tools and learning environments, promoting autonomy and personalization. Additionally, it allows imagining futures in which systems are transparent, auditable, and created by educational communities, thus avoiding dependence on opaque platforms or automated decisions that cannot be explained.

A second relevant metaphor is the Prometheus dilemma. In Greek mythology, Prometheus steals fire from the gods and gives it to humans, an act that represents both an advance for civilization and potential punishment for humanity. "In the stalk of a reed I hid the spark, the fountain of stolen fire, which for mortals is the teacher of all arts; and for this crime I now pay the penalty, nailed with chains under the ether." (Jonas, 1984).

GenAI-powered tools can be understood as systems with the capacity to offer "personalized" recommendations, adapt content in real time, and predict educational trajectories. All of this represents a form of "fire," powerful and transformative, but also a risk if used without considering ethics and transparency.

This metaphor warns about:

- The possibility of excessively delegating educational judgment to automated algorithms, displacing the teacher's role.
- The risk of generating limited or closed trajectories for students, guided by historical data patterns, instead of opening new paths.

- The implications of reducing learning to consumption or efficiency patterns, ignoring more complex dimensions such as creativity, conflict, or exploration.

Thus, while the Lego metaphor points to the possibility of emancipatory use, the Prometheus metaphor reminds us that all powerful technology carries ethical and political tensions, especially when deployed on a large scale in educational systems.

ETHICAL PRINCIPLES FOR SAFE INTRODUCTION OF GENAI IN EDUCATION

The growing pressure to integrate GenAI into educational systems poses not only opportunities, but also ethical, legal, and pedagogical challenges of great magnitude. At the regulatory level, in the European Union, GenAI systems must guarantee safe adoption that respects the European Union's privacy regulations (European Parliament, & Council of the European Union, 2016). Additionally, its recent emergence has driven the development of specific legislation on artificial intelligence in various regions, such as the European Union's AI Act (European Parliament, 2024) or emerging regulations in China (Cyberspace Administration of China et al., 2023).

To guarantee safe adoption that respects privacy regulations and is aligned with institutional values, strategies, and practices, the authors propose in this section a series of principles to guide the evaluation and deployment of GenAI applications in education. These principles seek to ensure that GenAI-based technologies: 1) Align with institutions' educational objectives. 2) Maintain adequate levels of security, accuracy, and ethical integrity. 3) Favor quality and equitable learning, minimizing risks associated with misuse, privacy, or misinformation.

Following these principles, educational institutions will be able to take advantage of the opportunities offered by GenAI while mitigating the risks associated with its implementation in adaptive learning contexts.

Principles of Safe GenAI in Education

Beyond legal compliance, ethical integration of artificial intelligence requires practical and actionable principles that allow evaluating both the pedagogical adequacy and technological security of adopted solutions.

Based on an analysis of legal frameworks and the technical nature of GenAI, Alier, García-Peñalvo y Camba (2024) have proposed a series of principles that were subsequently expanded to seven specific guidelines for educational environments (García-Peñalvo et al., 2024). These principles allow for technical, pedagogical, and ethical evaluation of GenAI integration strategies in education. These principles are presented below.

- **(SAIE1) Confidentiality guarantee:** GenAI systems must strictly protect student data, ensuring the privacy of their identities, records, and educational interactions.
- **(SAIE2) Alignment with educational strategies:** GenAI tools must support institutional objectives and adhere to technological governance policies, facilitating learning and knowledge creation without fostering bad practices such as plagiarism or evasion of academic controls.

- **(SAIE3) Adjustment to didactic practices:** GenAI applications must operate under defined pedagogical parameters. They should not automatically solve problems or tasks, but guide the student in reasoning processes aligned with instructional design.
- **(SAIE4) Accuracy and error minimization:** AI systems must prioritize the truthfulness and relevance of their responses, especially to minimize risks of erroneous information derived from biases or model hallucinations.
- **(SAIE5) Comprehensible interface and appropriate behavior:** GenAI must present its functionalities and limitations clearly to teachers and students, avoiding false expectations or erroneous interpretations of its reliability level.
- **(SAIE6) Human supervision and responsibility:** Decisions based on GenAI must be supervised by people. Faculty must maintain control over educational processes, ensuring the possibility of appeal and review of automated decisions.
- **(SAIE7) Ethical training and transparency:** GenAI models must be trained ethically, guaranteeing transparency about their data sources and methodologies, and minimizing biases to ensure fair and explainable results.

The application of these principles has important practical consequences for educational institutions:

- **Confidentiality (SAIE1):** Institutions must guarantee control over AI platforms used, avoiding mandatory registration in external tools like ChatGPT or Gemini, which could violate student privacy.
- **Strategic alignment (SAIE2):** General-purpose tools, such as Large Language Models (LLM), present institutional integration difficulties due to their complexity of use (Willison, 2023) and risks associated with academic integrity (González-Geraldo & Ortega-López, 2024). Additionally, adding complexity to the learning process is not good pedagogical practice, as it increases students' cognitive load (Chen et al., 2023).
- **Didactic adaptation (SAIE3):** Applications must integrate coherently with educational methodologies. For example, in environments such as medical training (Hwang et al., 2024), GenAI systems must act as assistants in clinical reasoning, not as automatic substitutes for students' analytical processes.
- **Accuracy and error minimization (SAIE4):** In educational environments, it is fundamental to design adaptive AI systems capable of citing sources and providing information validation mechanisms (Tonmoy et al., 2024).
- **Interface and transparency (SAIE5):** Interface design must favor understanding of GenAI's scope and limitations, avoiding the presentation of speculative responses as absolute truths.
- **Supervision and responsibility (SAIE6-7):** Human supervision not only ensures educational quality, but protects equity in learning and guarantees that adaptive systems do not reproduce biases or systematic exclusions.

These principles have led to the development of the Safe AI Manifesto (Alier, García-Peñalvo, Casañ et al., 2024), a living document that gathers the necessary commitments for ethical and responsible integration of GenAI in education, available for consultation and subscription at https://manifesto.safeaieducation.org. Additionally, the above principles allow for technical, pedagogical, and ethical evaluation of GenAI integration strategies in education, as developed in (García-Peñalvo et al., 2024).

Relationship between Safe GenAI Principles in Education and Educational Metaphors

Each of the proposed safe GenAI principles responds, in some way, to the risks or aspirations represented by the historical metaphors analyzed in the previous section:

- **(SAIE1) Confidentiality guarantee** combats the risk of the App Store and the educational cloud: The App Store metaphor shows how dependence on opaque platforms puts privacy at risk. Confidentiality requires reversing that logic, ensuring that student data is protected and under institutional control.
- **(SAIE2) Alignment with educational strategies** responds to the danger of Matrix. Faced with the vision of learning as passive download of predefined information, aligning GenAI with educational strategies allows building systems that foster active learning processes, instead of automating content transmission.
- **(SAIE3) Adjustment to didactic practices** connects with the Bazaar metaphor. As in the free software Bazaar, the principle seeks that GenAI tools be configurable and adaptable by teachers to specific pedagogical practices, not closed products that impose standard methodologies.
- **(SAIE4) Accuracy and error minimization** combats the illusions of the Prometheus vision. The Prometheus dilemma reminds us that the "fire" of GenAI can be dangerous if not managed properly. Minimizing errors and hallucinations helps control that risk inherent to systems as powerful as they are fragile.

LAMB: A PRACTICAL PROPOSAL FOR INTEGRATING SAFE GENAI IN EDUCATION

After analyzing the evolution of metaphors that have accompanied educational technologies and defining the principles of safe GenAI (SAIE), it becomes necessary to show concrete examples that include these principles in their design and application. This section presents LAMB (Learning Assistant Manager and Builder), a software framework designed specifically to create GenAI-based learning assistants in a safe, controlled manner that is pedagogically aligned with the educational institution's principles. First, the concept of GenAI-based learning assistants is explained. Next, LAMB is defined and how this framework allows creating GenAI-based learning assistants. After that, LAMB use cases are presented as well as how this framework is aligned with the principles presented in section 3 and finally its relationship with educational metaphors.

GenAI-Based Learning Assistants

A GenAI-based learning assistant is a conversational tool that interacts with students to offer personalized educational support. Unlike a generic chatbot, a learning assistant is designed to fulfill specific functions within an educational context (Kochmar et al., 2020; Wollny et al., 2021). The emergence of tools like ChatGPT has marked a turning point in this field, transforming the possibilities for personalization and educational interaction (Bettayeb et al., 2024; Chan & Lee, 2023):

- Respond to questions about course content.
- Guide problem-solving by applying defined pedagogical methods.
- Suggest relevant support resources.
- Promote reasoning and self-explanation in students.
- Accompany formative assessment processes by providing adjusted feedback.

These assistants do not seek to replace faculty, but to amplify their capacity for accompaniment and feedback, especially in environments with large numbers of students or in hybrid and distance modalities.

LAMB

LAMB is a software environment that allows educational institutions and teachers to create their own learning assistants, adapted to each subject, course, or specific need. This framework also allows integrating these assistants simply into learning management platforms (LMS) like Moodle, through the IMS LTI (Learning Tools Interoperability) standard, so that students access learning assistants through the institution's LMS (IMS Global Learning Consortium [IMS GLC], 2014). Additionally, LAMB allows controlling and configuring the knowledge base that feeds the assistant's responses, ensuring pedagogical alignment with the subject matter. Additionally, this framework allows managing the security, privacy, and supervision of created assistants.

Alier, Pereira et al. (2024) present the technical description, as well as LAMB's detailed architecture. In summary and from a technical perspective, LAMB combines:

- Large Language Models (LLMs) that generate responses in natural language.
- Retrieval-Augmented Generation (RAG) techniques, which limit the assistant's responses to verified information from sources controlled by the institution.
- A system of interaction templates and pedagogical configurations that the teacher defines previously.

In this way, LAMB allows leveraging the power of generative AI, but under conditions of control, transparency, and human responsibility.

Examples of Assistants Created with LAMB

To illustrate LAMB's real application in educational contexts, some of the experiences carried out are presented.

First, it is important to highlight the experience developed in the Economics and Economic Environment subject, a mandatory second-year subject in the Computer Engineering Degree at the Barcelona Faculty of Informatics.

Before addressing pilot experiences with students, it is important to contextualize LAMB's development process through the assistant called "Macroeconomics Study Coach," which functioned as a prototype for the design and refinement of the LAMB framework. This assistant was not subjected to formal evaluation with students, but served as a technical and pedagogical testbed during LAMB system development iterations.

This assistant was created from 30 course lecture videos. Its main functionality consisted of answering student questions based on video transcriptions and complementary PDF documents, providing specific citations and links to videos with precise moments where the queries made were addressed.

Once framework stability was consolidated through this case, controlled experiences with students were designed and implemented, with the PESTLE assistant being the first of these empirical validations.

The PESTLE case learning assistant was designed as a resource for the Business Administration and Management Fundamentals subject, taught in the Computer Engineering degree at EPSEVG (Vilanova i la Geltrú Higher Polytechnic Engineering School). The created assistant helped students work on an evaluation case of a technology-driven business project. The assistant had been equipped with a specific knowledge base for the case. Students, organized in teams, had to analyze the case using the PESTLE methodology (Political, Economic, Social, Technological, Legal, and Environmental), which implied they had to pose questions from various dimensions before writing their final report. They posed these questions to the learning assistant created with LAMB. In previous courses, students simulated being experts with help from search engines, specific documents, and even ChatGPT in autumn 2023 (Casañ et al., 2024).

The experience with the assistant to help students in the PESTLE case was developed during the 2023-2024 academic year. The activity involved a total of 47 second-year degree students organized in groups of 6-8 members.

The experience was structured following a case study methodology focused on analyzing Tesla's Optimus humanoid robot, using the PESTLE framework (Political, Economic, Social, Technological, Legal, Environmental). The activity was developed during two 2-hour class sessions, with a one-week interval between them to allow students to continue analysis autonomously.

In the first session, the case was presented and the PESTLE methodology explained, collaborative group work to analyze the six PESTLE dimensions was conducted, as well as using the assistant to obtain expert information. During the week, students had time to continue with the analysis using the assistant from home. In the second session, SWOT analysis integrated with PESTLE was introduced. Next, students were asked to categorize the elements or aspects they had found in each PESTLE dimension as "highly relevant," "relevant," or "not very relevant." Finally, students prepared a final report with conclusions.

The data collection methodology to evaluate the assistant's effectiveness was administered through a non-mandatory 5-question questionnaire using a 5-point Likert scale (1=strongly disagree, 5=strongly agree). The questionnaire obtained 27 responses from a total of 47 students (57.4% response rate). The questionnaire questions were as follows:

1. The assistant has helped me find relevant information for the case more quickly than if I had had to do it myself using the internet
2. The additional questions suggested by the system are good prompts to search for more information about the case
3. The answers to the questions suggested by the system provide useful information

4. Being able to consult data sources has been useful
5. The assistant's interface is simple and easy to use

The results show a mostly positive reception of the assistant by students:

- **Question 1** (Efficiency in information search): 83.3% of students scored 4 or 5 (17 favorable responses out of 27 total)
- **Question 2** (Quality of suggested questions): 77.7% scored 4 or 5
- **Question 3** (Usefulness of responses): 85.2% scored 4 or 5
- **Question 4** (Access to sources): 81.5% scored 4 or 5
- **Question 5** (Interface usability): 92.6% scored 4 or 5

The results indicate that no student scored 1 or 2 on questions 1, 2, and 4, and only a small percentage did so on questions 3 and 5, evidencing a significant absence of negative evaluations.

Qualitative and quantitative results confirm several pedagogical benefits:

> In general, as positive feedback in the experience, student responses highlighted benefits such as consistency and alignment. The assistants' ability to stick to the provided knowledge base was highlighted, without generating irrelevant information or hallucinations. Regarding workload reduction, the usefulness of assistants to answer repetitive queries about course procedures was valued. In the specific case of the assistant to help students resolve an evaluation of a business project, an improvement in the quality of student reports was observed thanks to access to expert knowledge.
>
> Among the challenges and limitations, it is worth highlighting excessive dependence on these types of tools, as there was a warning about the risk of students using the assistant as a shortcut, without deepening the course materials provided by faculty. Future improvements to LAMB were also proposed for more fluid integration with the LMS, specifically to allow sending queries to the teacher when the assistant has no response. Finally, the need to support minority languages such as Catalan or Basque was pointed out.
>
> In previous courses, students simulated being experts using search engines, specific documents, and, in the autumn quarter of 2023, ChatGPT in an uncontrolled manner. The implementation of the LAMB assistant provided a more controlled and pedagogically aligned environment, with information verified and contextualized specifically for the case study.

Relationship between LAMB and Safe GenAI Principles in Education

LAMB's design directly reflects compliance with the seven safe AI principles outlined in section three:

- **SAIE1 (Confidentiality):** LAMB guarantees student data privacy by operating on institutional servers or through controlled APIs. Sensitive information is not exposed to unauthorized external platforms.
- **SAIE2 (Alignment with educational strategies):** Assistants created with LAMB only use sources approved by the institution, ensuring that content and interactions are aligned with institutional educational objectives and values.

- **SAIE3 (Adjustment to didactic practices):** LAMB allows teachers to define interaction templates and specific knowledge bases for each course or activity, integrating coherently with instructional designs, instead of imposing generic interactions.
- **SAIE4 (Accuracy and error minimization):** Through Retrieval-Augmented Generation (RAG) techniques, assistants base their responses on reliable reference documents, reducing the risk of hallucinations or errors typical of generalist LLMs.
- **SAIE5 (Comprehensible interface and appropriate behavior):** LAMB provides assistants with predictable and delimited behaviors, clarifying their function and limitations, avoiding the illusion of infallibility that often accompanies many AI systems.
- **SAIE6 (Human supervision and responsibility):** Faculty design and supervise the contents, strategies, and behaviors of the assistant, preserving their central role as pedagogical mediator. This approach is consistent with practical experiences where LLMs act as assistants that automate repetitive tasks (code review, project monitoring), but maintaining the teacher in the role of guide and supervisor of the educational process (Pereira et al., 2025).
- **SAIE7 (Ethical training and transparency):** While underlying models are from third parties, LAMB allows institutions to build their own interaction bases, which in the future can be used to adjust or train models ethically adapted to their contexts.

LAMB and Educational Metaphors

The practical implementation of LAMB and learning assistants in real contexts has revealed difficulties that significantly, validate the warnings contained in the metaphorical analysis developed in the first part of this article. This convergence between critical theory and empirical experience reinforces the importance of maintaining active vigilance over the narratives that accompany technological integration in education.

One of the most evident concerns observed in LAMB experiences is the risk of excessive dependence by students, who may use the assistant as a shortcut that prevents them from deepening the course materials provided by faculty. This practical difficulty materializes precisely the dangers anticipated in the Palantír metaphor.

This observation does not invalidate the usefulness of assistants, but reinforces the need for clear pedagogical frameworks that guide their use. As one of the consulted experts noted: "it must be clearly defined that its 'knowledge' does not go beyond the corresponding content and does not address content beyond the teacher's needs." The Palantír metaphor reminds us that the power of technology must be accompanied by structures of control and ethical reflection.

The integration difficulties with learning management systems (LMS) observed in practical experiences reflect the tension between the Bazaar and App Store metaphors analyzed earlier. Students have requested improvements for more fluid integration with the Moodle platform (the LMS used in the educational institution), especially the possibility of making queries to the teacher when the assistant has no adequate response.

This demand illustrates the inherent limitation of closed systems (App Store metaphor). When technological tools operate in isolation, without capacity for interconnection and collaboration, frustration is generated and pedagogical potential is limited. In contrast, the Bazaar metaphor suggests the need for open, modular, and collaborative systems where different tools and actors (students, teachers, systems) can interact fluidly.

LAMB, by design, attempts to respond to this tension through its modular architecture and integration with open standards like LTI. However, the practical difficulties observed underscore that technical openness is not sufficient: pedagogical openness is also required that allows connection between artificial intelligence and the human intelligence of faculty.

The identified need for support of minority languages such as Catalan and Basque constitutes a concrete manifestation of the limitations of the Matrix metaphor discussed in the conceptual framework. The Matrix vision suggests that knowledge can be "downloaded" instantaneously and universally, ignoring specific contextual, cultural, and linguistic realities.

Practical experience with LAMB demonstrates that this apparent universality is illusory. Language models, however advanced they may be, reflect the biases and limitations of their training data, which tend to favor majority languages.

From a conceptual perspective, LAMB rejects the logic of the App Store and the Palantír, where users depend on closed and opaque systems. Instead, it aligns with the Lego metaphor: assistants are modular, customizable, and transparent tools that are adaptable.

In summary, LAMB offers a realistic example of how institutions can appropriate educational GenAI as a didactic resource and not as a threat to their autonomy or pedagogical agency.

CONCLUSIONS

This article has presented a critical perspective on the incorporation of GenAI in education, combining historical, ethical, and conceptual analysis. Through the use of metaphors, it has shown how visions of educational technology have oscillated between centralized control and participatory autonomy. Metaphors such as the Palantír, Matrix, or App Store help understand current risks: opacity, technological dependence, and pedagogical misalignment. In response, alternative frameworks have been proposed, such as Lego blocks or the Prometheus dilemma, which invite critical, transparent integration controlled by the educational community.

Based on this foundation, seven principles for safe GenAI have been formulated, centered on confidentiality, institutional alignment, human supervision, and transparency. These principles are not theoretical, but have been applied in practice through LAMB, a software environment that allows institutions to design and deploy configurable and auditable learning assistants. The analyzed cases show clear benefits in pedagogical coherence, student autonomy, and reduction of repetitive tasks, but also point to challenges such as dependency or integration with educational institutions' LMS.

Collectively, this work advocates for the need to recover pedagogical control over technology, to ensure that artificial intelligence in education serves meaningful learning and not external or opaque interests.

REFERENCES

Adell-Segura, J., Castañeda, L., & Esteve-Mon, F. M. (2018). ¿Hacia la ubersidad? Conflictos y contradicciones de la universidad digital. *RIED-Revista Iberoamericana de Educación a Distancia, 21*(2), 51-68. https://doi.org/10.5944/ried.21.2.20669

Alier, M., García-Peñalvo, F. J., & Camba, J. D. (2024). Generative artificial intelligence in education: From deceptive to disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence, 8*(5), 5-14. https://doi.org/10.9781/ijimai.2024.02.011

Alier, M., García-Peñalvo, F. J., Casañ, M. J., Pereira, J. A., & Llorens-Largo, F. (2024). *Safe AI in Education Manifesto (v 0.4.0).* https://manifesto.safeaieducation.org

Alier, M., Pereira, J., García-Peñalvo, F. J., Casañ, M. J., & Cabré, J. (2024). LAMB: An open-source software framework to create artificial-intelligence assistants deployed and integrated into learning-management systems. *Computer Standards & Interfaces, 92,* 103940. https://doi.org/10.1016/j.csi.2024.103940

Asimov, I. (1955-1976). *Historias de Multivac* [Serie de relatos]. Diversas publicaciones.

Bettayeb, A. M., Abu Talib, M., Sobhe Altayasinah, A. Z., & Dakalbab, F. (2024). Exploring the impact of ChatGPT: Conversational AI in education. *Frontiers in Education, 9,* 1379796. https://doi.org/10.3389/feduc.2024.1379796

Casañ, M. J., Alier, M., Pereira, J., & García-Peñalvo, F. J. (2024). Asistentes de aprendizaje basados en inteligencia artificial: Principios de seguridad y experiencias de implementación en educación superior. In M. Navarro-Granados, J. J. Sánchez Amate, P. Berbel Oller, & C. Rodríguez Jiménez (Eds.), *Investigación y conocimientos en la educación actual* (pp. 13-35). Dykinson.

Ceruzzi, P. E. (2003). *A history of modern computing* (2nd ed.). MIT Press.

Chan, C. K. Y., & Lee, K. K. W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Chen, O., Paas, F., & Sweller, J. (2023). A cognitive-load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review, 35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

Cyberspace Administration of China, National Development and Reform Commission, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, & National Radio and Television Administration. (2023). *Interim measures for the management of generative artificial intelligence services.* Promulgadas el 13 de julio de 2023, en vigor desde el 15 de agosto de 2023. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Clarke, A. C. (1962). *Profiles of the future: An inquiry into the limits of the possible.* Harper & Row.

European Parliament. (2024, 13 de junio). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (published in OJ on 12.7.2024).* https://data.europa.eu/eli/reg/2024/1689/oj

European Parliament, & Council of the European Union. (2016, 27 de abril). *Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data* *and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016, pp. 1-88).* https://data.europa.eu/eli/reg/2016/679/oj

García-Peñalvo, F. J., Alier, M., Pereira, J., & Casañ, M. J. (2024). Safe, transparent, and ethical artificial intelligence: Keys to quality sustainable education (SDG 4). *International Journal of Educational Research and Innovation, 22,* 1-21. https://doi.org/10.46661/ijeri.11036

González-Geraldo, J. L., & Ortega-López, L. (2024). Can AI fool us? University students' ability to detect ChatGPT. *Education in the Knowledge Society, 25,* e31760. https://doi.org/10.14201/eks.31760

Hofstadter, D. R., & Sander, E. (2018). *La analogía: El motor del pensamiento* (R. Musa, Trad.). Tusquets.

Hwang, G.-J., Tang, K.-Y., & Tu, Y.-F. (2024). How artificial intelligence supports nursing education: Profiling the roles, applications and trends of AI in nursing education research (1993–2020). *Interactive Learning Environments, 32*(1), 373-392. https://doi.org/10.1080/10494820.2022.2086579

IMS Global Learning Consortium. (2014). *IMS Learning Tools Interoperability (LTI) implementation guide v2.0.*

Jonas, H. (1984). *The imperative of responsibility: In search of an ethics for the technological age.* University of Chicago Press. https://doi.org/10.7208/chicago/9780226850337.001.0001

Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., & Pineau, J. (2020). Automated personalized feedback improves learning gains in an intelligent tutoring system. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6-10, 2020, Proceedings, Part II* (Lecture Notes in Computer Science, Vol. 12164, pp. 140-146). Springer. https://doi.org/10.1007/978-3-030-52240-7_26

Llorens-Largo, F. (2019, 13 de febrero). Las tecnologías en la educación: Características deseables, efectos perversos. *Universidad, sí.* https://www.universidadsi.es/las-tecnologias-en-la-educacion-caracteristicas-deseables-efectos-perversos/

Mason, J. (2018). A critical metaphor analysis of educational technology research in the social studies. *Contemporary Issues in Technology and Teacher Education, 18*(3). https://citejournal.org/volume-18/issue-3-18/social-studies/a-critical-metaphor-analysis-of-educational-technology-research-in-the-social-studies/

Pereira, J., López-Gil, J. M., & Alier, M. (2025). The AI-powered classroom: LLMs as teacher assistants for enhanced software-engineering learning experiences. In R. Molina Carmona, C. J. Villagrá Arnedo, P. Compañ Rosique, F. García-Peñalvo, & A. García-Holgado (Eds.), *Proceedings of TEEM 2024: The Twelfth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 115-124). Springer. (Lecture Notes in Educational Technology). https://doi.org/10.1007/978-981-96-5658-5_12

Pressman, R. S. (1991). *Software engineering: A practitioner's approach* (3rd ed.). McGraw-Hill.

Raymond, E. S. (2001). *The cathedral & the bazaar.* O'Reilly.

Skinner, B. F. (1968). *The technology of teaching.* Appleton-Century-Crofts.

Stallman, R. M. (1985). *The GNU manifesto.* Free Software Foundation. https://www.gnu.org/gnu/manifesto.html

Thompson, B. (2025, 5 de mayo). *Platform power is underrated.* Stratechery. https://stratechery.com/2025/platform-power-is-underrated/

Tolkien, J. R. R. (1954-1955). *The lord of the rings* (Vols. 1-3). George Allen & Unwin.

Tonmoy, S. M. T. I., Mehedi Zaman, S. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A comprehensive survey of hallucination-mitigation techniques in large language models. *arXiv preprint arXiv:2401.01313*. https://doi.org/10.48550/arXiv.2401.01313

Wachowski, L., & Wachowski, L. (1999). *The Matrix* [Película]. Warner Bros.

Weller, M. (2022). *Metaphors of ed tech.* Athabasca University Press. https://doi.org/10.15215/aupress/9781771993500.01

Willison, S. (2023). *Prompt injection attacks: It's only going to get worse.* https://simonwillison.net/2023/Apr/25/prompt-injection

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence, 4,* 654924. https://doi.org/10.3389/frai.2021.654924

Date of reception: 1 June 2025
Date of acceptance: 27 August 2025
Date of approval for layout: 26 September 2025
Date of publication in OnlineFirst: 16 October 2025
Date of publication: 1 January 2026