

THE PARADOX OF HEUTAGOGY AND ARTIFICIAL INTELLIGENCE

La paradoja de la heutagogía y la inteligencia artificial

LUIS ALBERTO TRIANA LLANO*

Universidad Pontificia Bolivariana, Medellín, Colombia

<https://ror.org/02dxm8k93>

luis.triana@upb.edu.co

<https://orcid.org/0000-0003-4367-2977>

Suggested citation: Triana Llano, Luis Alberto. (2026). The Paradox of Heutagogy and Artificial Intelligence. *Sophia*, Colección de Filosofía de la Educación, (41), pp. 177-212.

Abstract

This recapitulative study examines the intersection between the use of Artificial Intelligence (AI) and heutagogy, understood as self-determined learning. This combination may give rise to the paradox of autonomy versus dependence. The analysis is justified insofar as the emergence and exponential use of AI in educational environments raise the urgent need to explore how its integration transforms learning processes and the ways in which subjects construct their cognitive subjectivity. The central problem lies in the coexistence between the emancipatory potential of AI in a heutagogical key and the risk of a new form of technological subjection that limits the student's cognitive autonomy. The thesis defended is that AI is not, in itself, either emancipatory or alienating: its educational impact depends on the pedagogical, ethical, and epistemological frameworks from which it is integrated. A qualitative interpretive methodology was used, combining a systematic review of specialised literature with case analysis, employing conceptual analysis matrices and evaluative rubrics as instruments. The conclusions indicate that the combination of heutagogy and AI (referred to here as heutagogAI) promises innovative and emancipatory educational experiences; however, it requires designing systems that promote critical and reflective learning, ensuring that students develop skills to question and analyse the information received, a fundamental condition for a truly emancipatory education.

Keywords

Artificial Intelligence, Critical Thinking, Educational Autonomy, Educational Philosophy, Heutagogy, Self-instruction.

* Ph.D. candidate in Philosophy at Universidad Pontificia Bolivariana (Medellín, Colombia); Master's degree in Philosophy for Contemporary Challenges from the Universitat Oberta de Catalunya (Spain); a degree in Philosophy from the National Open and Distance University (UNAD, Colombia), a degree in Occupational Safety and Health Engineering from the Jaime Isaza Cadavid Colombian Polytechnic (PCJIC, Colombia), and a current law student at UNAD. His research focuses on education in incarceration settings, philosophy, ethics, and risk management. Google Scholar: <https://scholar.google.es/citations?user=PezDoIUAAAAJ&hl=es> h-index: 2.

Resumen

El estudio recapitulativo acá presentado examina la intersección entre el uso de inteligencia artificial (IA) y la heutagogía, entendida como aprendizaje autodeterminado. Esta combinación puede devenir en la paradoja de la autonomía frente a la dependencia. El análisis se justifica en tanto la irrupción y el uso exponencial de la IA en entornos educativos plantean la necesidad urgente de explorar cómo su integración transforma los procesos de aprendizaje y los modos en que el sujeto construye su subjetividad cognitiva. El problema central reside en la coexistencia entre el potencial emancipador de la IA en clave heutagógica y el riesgo de una nueva forma de sujeción tecnológica que limite la autonomía cognitiva del estudiante. Se defiende que la IA no es, en sí misma, ni emancipadora ni alienante: su impacto educativo depende de los marcos pedagógicos, éticos y epistemológicos desde los cuales se integre. Se empleó una metodología cualitativa interpretativa que combina revisión sistemática de literatura especializada con análisis de casos, usando matrices de análisis conceptual y rúbricas evaluativas. Las conclusiones indican que la combinación de heutagogía e IA (denominada “heutagogIA”) promete experiencias educativas innovadoras y emancipadoras; no obstante, exige diseñar sistemas que promuevan un aprendizaje crítico y reflexivo, asegurando que los estudiantes desarrollen habilidades para cuestionar y analizar la información recibida, condición fundamental para una educación verdaderamente emancipadora.

Palabras clave

Autoaprendizaje, autonomía educativa, filosofía de la educación, heutagogía, inteligencia artificial, pensamiento crítico.

Introduction

The Sciences of the Artificial do not merely aim to understand and explain phenomena (in this case, educational ones), but rather seek to improve things by introducing new possibilities (Alonso Rodríguez, 2021, p. 132).

Contemporary education faces perhaps its greatest historical challenge: we are witnessing an unprecedented educational revolution, driven in particular by the emergence and rapid evolution of artificial intelligence (AI) —an evolution that has been accelerated by the emergence of machine-learning AI models capable of processing and generating text in a coherent, contextual manner using natural language, such as ChatGPT, Perplexity, CoPilot, Gemini, and others, which have been rapidly integrated into the teaching-learning process (Caldeiro et al., 2024; García Peñalvo, 2024; García Peñalvo et al., 2024; Peñalver Higuera et al., 2024). This integration is neither a marginal nor a future phenomenon; it is a present reality that challenges teachers, students, institutions, and philosophers of education alike.

In this context, drawing on the principles of Knowles (1970), “heutagogy” is applied as self-determined learning, a process in which the student assumes an active role in the planning, direction, execution, and evaluation of their own educational process. Heutagogy promotes the autonomy of the student as a subject, in contrast to traditional pedagogi-



cal approaches in which the teacher is the primary mediator. However, the emergence of AI in this framework introduces a tension that cannot be ignored: if the heutagogical learner increasingly relies on algorithmic systems to make decisions about what and how to learn, are they truly autonomous, or do they transfer that autonomy to an opaque computational logic that they themselves neither control nor understand?

From this tension emerges the central problem that guides this research: the coexistence of AI's emancipatory potential in a heutagogical context and the risk of a new form of technological subjugation that limits the student's cognitive autonomy. In particular, the study examines whether the growing use of AI systems in education effectively promotes self-determined learning or whether, on the contrary, it shifts decision-making away from the individual toward opaque algorithmic logics that constrain thought, creativity, and critical reflection. This question is not rhetorical; it has major pedagogical, ethical, and political implications for the design of twenty-first-century educational systems.

The research posits as its central thesis that AI is, in and of itself, neither emancipatory nor alienating; rather, its educational impact depends on the pedagogical, ethical, and epistemological frameworks within which it is integrated. In this regard, the study argues that, within a critical-reflective heutagogical approach, AI can serve as a tool to enhance self-directed learning, provided that educational experiences are designed to foster critical thinking, ethical awareness, and the students' ability to question both the content and the algorithmic systems themselves that mediate their learning. This stance avoids both uncritical enthusiasm for technology and reactive rejection, advocating instead for a reflective and politically conscious adoption of AI in education.

The importance and timeliness of this study are indisputable. With the widespread adoption of generative language models, the discussion about AI in education has shifted from being prospective to becoming urgent and immediate. International organizations such as UNESCO (2022) and the European Parliament (2024) have begun to establish regulatory and ethical frameworks to govern the use of these technologies, explicitly recognizing that their implementation without clear pedagogical criteria can deepen inequalities, perpetuate biases, and erode the individuals' intellectual autonomy. In Latin America, where digital and educational divides have historically been pronounced, this debate takes on an additional dimension of cognitive justice and social equity. Analyzing this transformation through the lens of heutagogy is particularly relevant, as it allows us to critically examine discourses on educational innovation



and place the learner's autonomy at the center in the face of risks of dependence, homogenization, and algorithmic subjugation.

The main objective of this study is to critically analyze the intersection between heutagogy and AI to determine the conditions under which the latter can genuinely promote self-determined learning rather than creating new forms of technological dependence. The discussion is organized into eight sections that progress step by step from conceptual foundations to concrete proposals.

The first section describes the interpretive qualitative methodology employed. The second establishes the theoretical framework that articulates autonomy, technology, and self-determined learning as the analytical axes of the study. Starting with the third section, the analysis unfolds in depth: it examines heutagogy and its relevance to AI, identifying the tensions that arise when self-determined learning is mediated by algorithmic systems; it then characterizes artificial intelligence as a disciplinary field and analyzes its applications in educational contexts. These two sections converge in the fifth, dedicated to "heutagogIA" as a neologism and pedagogical proposal that names this intersection.

The following three sections move toward a critical and propositional dimension: forms of resistance to algorithmic subjugation are explored under the title "Freeing Ourselves from AlgoritmoIA"; a proposal for emancipatory heutagogical education that places the subject's autonomy at the center is articulated; and a guiding decalogue for the implementation of heutagogy based on ethical, inclusive, and reflective criteria is formulated. The study concludes with a summary that outlines the objective, the research problem, the difficulties encountered, and projections for future research.

Methodology

The research adopts an interpretive qualitative approach aimed at understanding the processes, meanings, and implications associated with the use of AI in educational contexts from a heutagogical perspective. This approach is appropriate given that the object of study does not lend itself to direct quantification; rather, the goal is to explore conceptual relationships, epistemological tensions, and ethical implications that require comprehensive analysis and interpretation grounded in explicit theoretical frameworks.

The research strategies and techniques combined a systematic review of specialized literature with a particular focus on the analysis of



case studies involving the use of AI in educational settings. The literature review was conducted comparatively, allowing for the identification of trends, gaps, and relevant contributions at the intersection of AI and heutagogy, with particular attention to publications from recent years that reflect the emergence of generative language models in educational contexts.

The research procedure was structured into three sequential phases. In the first phase, an exhaustive literature review was conducted to establish the state of the art and delineate the main debates and theoretical perspectives surrounding heutagogy, AI in education, and the ethical and political implications of algorithmic systems. In the second phase, representative cases of AI applications in education that incorporate heutagogical elements were identified and analyzed, focusing on learning processes, human-machine interaction, and the possibilities for technological emancipation or subjugation. In the third phase, an integrative conceptual framework was developed based on the findings from the previous phases, culminating in the formulation of a set of ten recommendations for the design of critical, reflective, and emancipatory AI-mediated educational experiences.

The data collection and analysis tools included, first, conceptual analysis matrices designed to systematize and categorize the reviewed literature based on four dimensions:

- Conception of the learner.
- The role assigned to technology in the educational process.
- Tensions between autonomy and dependence identified in the studies.
- Proposals for—or the absence of—an emancipatory approach.

These matrices made it possible to compare perspectives from different disciplinary traditions—pedagogy, philosophy of education, computer science, and technology ethics—to identify substantive convergences and divergences.

Second, evaluation rubrics were used to analyze cases of AI implementation in education, organized around criteria such as the degree of student agency in the process, the transparency of the algorithmic system used, the presence or absence of instances of critical reflection on the tool, and alignment with principles of equity and inclusion. These rubrics made it possible to evaluate the cases not only in terms of their instrumental effectiveness but also in terms of their consistency with the values of an emancipatory education.



Autonomy, Technology, and Self-Determined Learning

This research is grounded in a theoretical framework that articulates three fundamental conceptual axes: the learner's autonomy in learning, technological mediation, and the critical dimension of education. These three axes are not independent of one another; rather, they are dialectically interrelated, generating productive tensions that constitute the philosophical core of the research problem.

The autonomy of the learner

The first conceptual axis is built upon the traditions of critical pedagogy and andragogy. Knowles (1970) identified in the adult learner a capacity for self-direction that traditional education tends to suppress rather than foster—an argument that laid the groundwork for rethinking the role of the teacher as an epistemic authority and that of the student as a passive recipient. Building on this foundation, Hase and Kenyon (2001) formulated heutagogy as a move beyond the andragogical model: whereas in andragogy the teacher facilitates adult learning within a predefined curricular framework, in heutagogy the learner fully assumes responsibility for the planning, design, development, and evaluation of their own learning process. This distinction is not merely instrumental but epistemological; it entails a conception of the learner as an agent capable of defining their own questions, selecting their own methods, and evaluating their own results, without relying on a preestablished curriculum or the teacher's mediation.

This concept resonates with Jacques Rancière's (1996) proposal in *The Ignorant Schoolmaster*, where intellectual emancipation does not stem from the teacher's instruction but from the recognition of the equality of intelligences. For Rancière, the learner does not need a teacher to impart knowledge; rather, the learner needs someone who challenges them to use their own intelligence. The "ignorant teacher" does not instruct because they do not know; they challenge because they trust in the other's capacity. This relational architecture of learning proves philosophically fruitful when AI is introduced into the heutagogical setting: an AI system does not "know" in the full sense of the term (it does not understand, it has no consciousness, and it does not experience the world); however, it can facilitate genuine learning processes in those who interact with it. In this sense, AI could function as a sort of technological "ignorant teacher"—a device that does not possess the knowledge it appears to convey but which can engage the student in ways that activate their own critical in-



telligence. The risk, however, is that this analogy obscures a fundamental difference: while Rancière's "ignorant teacher" respects and encourages the student's intellectual emancipation, a poorly designed AI system can produce exactly the opposite effect, homogenizing responses, discouraging genuine questioning, and reinforcing cognitive dependence.

Blaschke and Marín (2020) have clarified that heutagogy is not equivalent to self-regulated learning or self-directed learning, with which it is frequently confused. The fundamental difference lies in the degree of agency the learner possesses: whereas in self-regulated learning the student manages their process within a given curricular framework—selecting strategies, managing time, and monitoring their performance—in heutagogy that very framework is defined by the student. It is the learner who decides what is worth learning, through which means, and according to which assessment criteria. This distinction becomes crucial when AI is introduced, as the risk lies precisely in algorithmic systems imposing an implicit curricular framework—determined by training data and the system's objectives—which the student perceives as their own but which has in fact been designed by others. This surreptitious reimposition of a heteronomous curriculum under the guise of personalization constitutes one of the deepest tensions in heutagogy and demands constant philosophical and pedagogical vigilance.



Technological Mediation and the Production of Subjectivities

The second conceptual axis is based on the premise that all technology participates in the production of subjectivities. This assertion should not be read as a naive technological determinism—according to which technology inexorably dictates the ways in which subjects think and act—but rather as a critical acknowledgment that technological artifacts are not neutral: they embody values, interests, and power relations that condition, in more or less subtle ways, the ways in which subjects relate to knowledge and to themselves.

Foucault (2002) analyzed, in the context of disciplinary institutions, how apparatuses of knowledge-power produce docile subjects through surveillance, norms, and examination. Power is not exercised solely over the body; it is inscribed in practices, routines, and the ways of evaluating and being evaluated. The traditional school has been one of the primary spaces for this disciplinary production. Deleuze (2006), for his part, adapted this analysis to "societies of control," in which regulatory mechanisms no longer operate through confinement and direct surveillance, but rather through the continuous modulation of behaviors via networks, flows, and computerized systems. What is significant about this shift is

that subjection is no longer visible or traceable: it permeates interfaces, algorithmic recommendations, and scoring and gamification systems that silently shape the subject's decisions. This historical progression—from sovereign power to disciplinary power and from there to modulated control—takes on a new dimension with educational AI: algorithmic systems not only monitor student performance but also predict their trajectories, suggest their paths, and shape their expectations, thereby shaping what might be called algorithmic subjectivity. In this scenario, the question of heutagogical autonomy becomes philosophically pressing: Can a subject be truly autonomous when their learning decisions are guided, filtered, and fed back by a system that operates under an opaque logic?

Stiegler (2015), drawing on the philosophy of technology, has argued that any externalization of thought onto technical media carries the risk of what he calls “cognitive proletarianization”—the progressive loss of know-how and life skills when cognitive functions are delegated to devices that perform them more efficiently. This assessment is particularly relevant in the context of educational AI: if students externalize to the system the selection of what to learn, the formulation of their questions, and the evaluation of their answers, they not only lose practice in these skills but may also lose awareness that they ever possessed them. Winner (2008) and Collins and Pinch (1996), for their part, have emphasized that technological artifacts embody political and moral choices that are not apparent to the user. An AI system trained on specific datasets, designed to optimize specific metrics, and deployed in specific institutional contexts is not a transparent window into knowledge; rather, it is an artifact that reflects the interests, biases, and values of those who built it. This reality calls for critical technological literacy that goes beyond the instrumental learning of digital tools to include an understanding of the mechanisms of power at work in algorithmic systems—which is, in itself, a philosophical and pedagogical task of the highest order.

Critical Education as Emancipation

The third conceptual axis draws on the tradition of critical pedagogy and the philosophy of emancipatory education. Garcés (2020) defines emancipatory education as that which “understands education to be a path to liberation not only for a select few but for everyone” (p. 27). This definition has an inescapable political dimension: emancipation is not an individual process of self-realization, but a collective process of building the conditions for the autonomy and dignity of all individuals. To emancipate oneself does not mean merely freeing oneself from external constraints,



but developing the capacity to recognize the conditions of one's own subjugation and to act upon them with intention and discernment.

Along these lines, Freire (1970) had already warned that “banking” education—that which deposits content into the student without involving them in its construction—is a form of oppression disguised as instruction. Freire’s critique takes on new relevance in the face of AI systems that generate “personalized” content without the student having participated in defining what they need to learn or in assessing what they have learned. The risk is not merely pedagogical; it is political, insofar as it reproduces relationships of dependency under the guise of autonomy.

De Sousa Santos (2006) warns that “we have only hegemonic instruments to try to confront all of this, because today we lack the concepts to confront the new, discontinuity, rupture, and revolution” (p. 65). This warning is particularly relevant in the context of AI in education: the available tools are, for the most part, products of technology corporations operating under a market logic that does not necessarily align with the emancipatory goals of education. However, this does not mean they cannot be appropriated in a critical and transformative way. The history of education is full of examples of the subversive appropriation of hegemonic tools—such as the printing press, radio, video, and the internet; in each case, the tool was developed with specific power dynamics in mind and was later appropriated by educational and social movements that used it to challenge those very dynamics. The learner is not a passive recipient of technology; they can interrogate it, subvert it, and redefine its meaning.

This possibility of reframing refers to Deleuze and Guattari’s (2002) concept of the “line of flight”—movements that escape established territorialities and open up new possibilities for existence. Conceiving of heutagogy as a line of flight implies viewing AI not only as a learning tool but also as an object of learning: the student who learns to question the algorithmic systems that mediate their education is, in the very act of that questioning, exercising a form of technological emancipation that transcends the instrumental use of the tool.

The articulation of these three axes—the autonomy of the subject, technological mediation, and emancipatory education—constitutes the theoretical framework from which heutagogy is approached as an educational proposal. This framework does not resolve the paradox that gives this work its title, but it offers the analytical categories necessary to think through it with rigor and depth, avoiding both techno-utopian enthusiasm and Luddite pessimism, and instead advocating for a reflective, critical, and politically conscious appropriation of AI in education.



Heutagogy and Its Relevance to AI

The act of knowing is revealed to be an activity inherent to the being, in which one glimpses—or rather, acquires meaning in one’s way of being, based on gaining knowledge of that which seems unknown to them (Obando et al., 2018, p. 96).

Heutagogy did not emerge as a response to the technological revolution, but rather as a consequence of a deep dissatisfaction with teacher-centered educational models. Hase and Kenyon (2001) formulated it as an extension and improvement upon Knowles’s (1970) andragogy: whereas in andragogy the teacher facilitates a learning process whose curriculum has been institutionally defined, in heutagogy it is the student who plans, directs, designs, develops, and evaluates their own learning. This difference is not one of degree but of nature; it implies an ontological shift in the educational subject, who ceases to be the object of pedagogical action and becomes the agent of their own learning process.

It is important, however, to avoid confusion between heutagogy and related concepts that are often used as synonyms. As Narváez and Prada (2005) point out, terms such as “self-regulated learning, autonomous learning, and self-planned learning” (p. 118) are related to heutagogy but do not define or constitute it. The central distinction lies in the object of self-determination: while self-regulated or self-directed learning operates within a given curricular framework, managing resources, time, and strategies, heutagogy involves defining one’s own framework, choosing one’s own questions, and constructing one’s own assessment criteria. This distinction takes on particular relevance when AI is introduced as a mediator, since the risk lies precisely in the algorithmic system surreptitiously reimposing an implicit curriculum that the student perceives as their own without having chosen it.

Heutagogy, although it has existed as a practice long before its conceptual formalization, experienced a notable expansion following the 2020 pandemic, when educational institutions were compelled to reconfigure their teaching models for distance and virtual learning contexts. Hairi et al. (2022) document implementations in “the United Kingdom, Cyprus, New Zealand, Germany, Canada, Rwanda, and Portugal” (p. 6693), among others. In the Spanish-speaking world, the Universitat Oberta de Catalunya (UOC) in Spain and the Universidad Nacional Abierta y a Distancia (UNAD) in Colombia stand out as institutions that have incorporated heutagogical principles into their pedagogical approaches. This geographic expansion reveals that heutagogy is not a marginal academic



approach, but rather an educational trend with growing institutional and research support.

However, what gives heutagogy particular relevance in the current context is precisely the emergence of generative AI. On the one hand, AI systems seem to offer ideal conditions for self-directed learning: unlimited access to information, immediate feedback, adaptation to the user's pace and interests, and constant availability. On the other hand, these same systems can operate as mechanisms of subjugation that erode the subject's autonomy: if the heutagogical learner delegates to AI the selection of what to learn, how to learn it, and how to assess it, self-determination becomes an illusion sustained by algorithms. This tension lies at the heart of the paradox that gives this paper its title.

To paraphrase David Miller: "In order to achieve learning, it is necessary to know something and have prior knowledge, whether in physical (face-to-face) space, intangible (virtual) space, or cyberspace as a setting where epistemological and educational debate is possible" (cited in Hermann Acosta, 2011, p. 101), we encounter an argument that runs counter to the aforementioned position of Rancière (1996), who states that "one can teach what one does not know" (p. 56)—a contradiction in which heutagogy and AI might serve as mediators.

This issue takes on even greater depth when examined through Rancière's thought, which proposes that true learning arises from the recognition of the equality of intelligences and not from the transfer of knowledge from an epistemic authority to a passive recipient. The apparent convergence between Rancière's emphasis on student autonomy—as well as on the student's inherent capacity to learn on their own—and heutagogy, in which the student takes control over what and how they learn, is defended in the French philosopher's arguments when he proposes that true learning is that which arises from the equality of intelligences in a process where the teacher does not teach in the traditional sense: the "ignorant teacher" acts as a facilitator of the student's intellectual emancipation.

Both Rancière and heutagogy reject the student's dependence on the teacher as the sole source of knowledge; both promote a model of self-determined learning—in which the student decides what and how to learn—as well as critical reflection and active exploration. Both heutagogy and Rancière's thought make it clear that meaningful learning is not imposed but is constructed through the student's own will and capacity. That said, we are concerned about whether what is taught—that which is unknown—could be an AI, given that, in essence, AI does not know what it teaches; it is something like a sophisticated technological parrot that



repeats the information previously provided to it, and in that sense, it could be analogous to Rancière's ignorant teacher, who would teach what, strictly speaking, they do not know—they merely repeat it.

It is at this point that heutagogy and AI (heutagogIA) converge for the purposes of this paper. This intersection represents a transformative possibility for emancipatory education, but it also poses a fundamental paradox: the tension between heutagogical autonomy and potential technological dependence (Morduchowicz, 2023). As Mollick and Mollick (2023) warn, the algorithmic design of AI systems tends to standardize processes, which could restrict the creative exploration inherent in self-determined learning. This leads us to ask: What mechanisms would ensure that AI functions as a facilitator rather than a limiter of students' creative agency in heutagogical contexts?

As Mollick and Mollick note: "AI can be a useful tool for promoting teaching and learning practices when implemented with caution and consideration in the classroom" (2023, p. 24). This caution should not be understood as timidity toward technology, but rather as pedagogical and philosophical rigor: the very tool that can enhance autonomy can also erode it, depending on the framework of values, intentions, and practices within which it is deployed.

Artificial Intelligence

The history of AI as a multidisciplinary field drawing on mathematics, logic, computer science, and neuroscience has evolved over several decades. For the purposes of this article, it is worth highlighting three key milestones. The first dates back to 1842, when mathematician Ada Lovelace programmed what is considered the first algorithm intended to be processed by a machine (Abeliuk and Gutiérrez, 2021, p. 15). The second milestone corresponds to Alan Turing's work in 1950, as outlined in his article "Computing Machinery and Intelligence," in which he first posed the question of whether a machine can think, thereby establishing the conceptual framework that would guide decades of research. The third milestone occurred in 1956 at the Dartmouth Conference, which Abeliuk and Gutiérrez (2021) identify as the "founding moment of 'artificial intelligence,' both in terms of the term itself and the field of study" (p. 15). Since then, AI has gone through various stages: the early rule-based expert systems, the AI winter of the 1970s and 1980s, the resurgence with neural networks in the 1990s, and the deep learning revolution of the



2010s, which made applications such as facial recognition, virtual assistants, and, more recently, generative language models possible.

AI has advanced significantly thanks to the increase in available data and hardware development, which has facilitated the emergence of advanced natural language processing models. Among these, GPT (Generative Pre-trained Transformer) models stand out; they are based on deep neural networks and are capable of generating text and sustaining natural conversations. Tools such as ChatGPT, Perplexity, Copilot, and Gemini are examples of this evolution. In particular, ChatGPT has become one of the most widespread and well-known systems, using the GPT architecture to process and generate human language, thereby expanding the applications of AI in fields such as education, which is the focus of this study.

This historical trajectory is not merely a backdrop; it reveals that the development of AI has always been shaped by available computational capabilities, the training data used, and the objectives that researchers and industries have defined as priorities. Understanding these factors is essential for critically evaluating current AI systems in educational contexts.

Among the many definitions of AI, we adopt the one provided by UNESCO in Morduchowicz (2023): “AI is the ability of a system to interpret external data, use this data to achieve specific goals, and carry out actions that maximize the chances of success in a particular task” (p. 15). This deliberately functionalist definition allows us to clearly identify both the possibilities and the limits of AI in education. Thus, if the “specific goals” that an AI system optimizes are not aligned with the objectives of an emancipatory education—the development of critical thinking, intellectual autonomy, and the ability to question—the system may achieve satisfactory performance metrics while simultaneously producing pedagogically impoverished results.

In this sense, AI can provide rapid access to resources, helping students research topics of interest. This can foster the autonomy that would allow students to delve deeper into specific areas that interest them or that they need to explore to discover their interests. AI, in its current configuration, can assume multiple roles in the educational process, as summarized by Caldeiro et al. (2024, p. 8) in a classification ranging from the role of simulator and tutor to that of dynamic evaluator and co-designer (Tables 1 and 2). Each of these roles entails both pedagogical benefits—such as personalization, immediate feedback, and deliberate practice—and significant risks: “confabulation” or the generation of false but plausible information, the undue transfer of thought to the system, and incompatibility between the system’s interaction style and that of the stu-



dent. What these classifications reveal is not only a taxonomy of uses but also a map of pedagogical tensions that require reflective management on the part of teachers and students. As Mollick and Mollick (2023) suggest, “AI can be a useful tool for promoting teaching and learning practices when implemented with caution and consideration in the classroom” (p. 24). Building on this idea, Caldeiro et al. (2024) summarize some of the benefits and risks:

Table 1
Approaches Using Prompts

Use of AI	Role	Pedagogical benefit	Pedagogical risk
Simulator	Deliberate Practice	Practicing and applying knowledge helps with its transfer.	Inappropriate fidelity.
Tutor	Direct instruction	Personalized direct instruction is very effective.	The AI’s knowledge base is inconsistent. There are serious risks of confabulation.
Mentor	Provide feedback	Frequent feedback improves learning outcomes, even if not all advice is followed.	Failure to critically analyze feedback, which may contain errors.
Motivator	Encourage metacognition	Creating opportunities for introspective analysis and self-regulation, which enhances the quality of learning.	The tone or style of coaching may not match the student’s. Risk of incorrect advice.
Peer	Improving Team Performance	Providing alternative perspectives helps learning teams function more effectively.	Inaccuracies and fabrication of data. Compatibility issues when interacting with other group members.
Student	Gaining explanations	The act of explaining things to others is a highly effective learning strategy.	Fabricating information and engaging in discussions can negate the benefits of the pedagogical approach.
Tool	Carry out specific activities	It allows learners to make greater progress in the same amount of time.	Delegate thinking, rather than work.

Source: adapted from Caldeiro et al. (2024, p. 8).

Table 1 highlights some applications that an AI-supported heutagogical program could develop, ensuring that the learner is not dependent on AI, but rather can discern, critique, analyze, and even emancipate themselves from AI through AI. However, in the education sector, can any AI be used? Should AI be designed, programmed, trained, and optimized specifically for education? It seems the answer may be implicit in the Synthesis School proposal, whose slogan states, “For kids who think for themselves” (Synthesis, 2024). These children who think for themselves will be educated using the AI employed at Synthesis School to “develop students captivated by complexity and the resolution of the unknown” (Synthesis, 2024).

At first glance, this would seem to align with Rancière; however, it may not yet be possible, as suggested by statements from Chomsky, Roberts, and Watumull published in *The New York Times*:

Its most profound flaw is the absence of the most critical capacity possessed by any intelligence: to state not only what is happening, what happened, and what will happen—i.e., to describe and predict—but also what is not happening and what could and could not happen (Kuo, 2023).

It is worth noting that these are the necessary ingredients of an explanation, as “the hallmark of true intelligence” (Kuo, 2023). Therefore, the ability to explain what is absent and consider what is not obvious is crucial for informed decision-making and the resolution of complex problems. True intelligence is not limited to the superficial; it involves depth and reflection—something that, for now, cannot be achieved by AI alone—and carries the risk of “hallucinations.”

In this regard, Caldeiro et al. (2024) state that “it can produce responses that seem plausible or reasonable but are inaccurate or irrelevant because they are not based on factual data, or due to the tool’s lack of specific skills, such as solving mathematical problems” (p. 11). This phenomenon is not a minor technical glitch; it poses a direct threat to the epistemic integrity of the learning process, especially when students lack the criteria to evaluate the veracity of the information they receive. A fundamental pedagogical conclusion follows from this: the integration of AI into heutagogical contexts cannot occur without parallel training in the critical evaluation of information and in understanding the limits and mechanisms of algorithmic systems.

Hallucinations represent one of the greatest challenges and pose a risk to the AI-mediated teaching-learning process, but it is not the only one. There is also the inability to balance creativity with responsibility,



insofar as its results may yield both truths and falsehoods, as well as bi-ased arguments that could ultimately support both ethical and unethical decisions alike (Kuo, 2023).

That said, AI can assume roles such as those listed in Table 1, but these roles can be expanded even further; for example:

Table 2
Applications of ChatGPT

Role	Description	Example
Idea Generator	The AI generates al-ternatives for express-ing the same idea.	Students can use ChatGPT and the “Regen-erate Response” feature to explore different ways of phrasing ideas.
Socratic Opponent	AI challenges ideas and arguments, promoting critical thinking.	Students can ask ChatGPT questions by structuring the interaction as a debate. Teachers can ask students to use ChatGPT to prepare arguments.
Collabora-tion <i>Coach</i>	The AI assists groups in collab-orative research and problem-solving.	Working in teams, students can use Chat-GPT to search for information that helps them complete projects and assignments.
Comple-mentary Guide	AI serves as a guide for exploring both physical and concep-tual spaces.	Teachers can use ChatGPT to create class-room materials (such as discussion ques-tions) and request suggestions on how to facilitate understanding of specific topics.
Personal Tutor	AI provides individu-alized guidance and immediate feedback on student progress.	ChatGPT can offer personalized feedback to students based on information provided by them or by teachers (for example, assess-ment results).
Co-designer	AI collaborates throughout the design process.	Teachers can ask ChatGPT for ideas on designing or updating curricula (e.g., as-sessment rubrics) or on addressing specific objectives (e.g., how to improve curriculum accessibility).
Exploratory	AI provides tools for inquiry and interpretation.	Teachers can provide students with basic information, and students can then ask ChatGPT various questions to delve deeper into the topic. ChatGPT can be useful for supporting language learning.

Role	Description	Example
Study Partner	AI helps students reflect on the learning material.	Students can explain their level of understanding to ChatGPT and ask for help studying the material. ChatGPT can also help students prepare for various tasks (such as job interviews).
Motivator	AI suggests games and challenges to expand learning.	Teachers and students can ask ChatGPT for ideas on how to expand students' learning by providing a summary of their current level of knowledge (for example, quizzes or exercises).
Dynamic Assessor	The AI provides educators with a profile of each student's current knowledge.	Students can interact with ChatGPT in a tutoring-style dialogue and then ask ChatGPT to generate a summary of their current state of knowledge to share with the teacher for assessment.

Source: adapted from Caldeiro *et al.* (2024, p. 9).



The roles in Table 2 are closer to a potential combination of AI with heutagogy; of course, heutagogical effects must be considered when designing these specific roles, both in AI training and in the students' use of the system. Finally, one of the objectives is for students to be able to discern the information provided by the AI so that they can also use more precise questions to challenge the AI's model outputs. This creates a sort of spiral of refinement for the information and knowledge acquired, in pursuit of a reflective and analytical approach that encourages the exploration of deeper aspects—the latter of which is reminiscent of Gadamer's (2003) hermeneutic spiral.

Heutagogy

The intersection between AI and heutagogy—which we have termed “heutagogIA” here—is not simply the sum of two approaches, but rather an articulation that generates its own emergent properties, both in terms of pedagogical possibilities and specific risks. Thus, heutagogy promises new forms of interaction and the empowerment of self-directed learning, which allows for “optimizing education based on each student's unique needs and development, thereby enabling the creation or adaptation of personalized learning materials and experiences” (Aguilar, 2024). This personalization,

in principle, is one of the central objectives of heutagogy: that each individual can chart their own educational path. However, it is worth critically examining what kind of personalization current AI systems actually offer.

Algorithmic personalization operates on the basis of data: the system identifies patterns in user behavior, compares them with patterns from other users, and generates optimized recommendations. But this process has a fundamental epistemological limitation: it can only personalize within the space of possibilities defined by the training data. An AI system cannot recommend to a student something that has never appeared in its data, nor can it anticipate interests the student has not yet developed, nor can it evaluate creativity that challenges preexisting categories. In this sense, algorithmic personalization is, paradoxically, a form of homogenization: all students are personalized according to the same logic, optimized according to the same criteria, and steered toward territories already mapped by the data. As Morduchowicz (2023) points out, “AI is capable of storing data, organizing it, processing it, sorting it, and drafting texts, but it lacks critical thinking, curiosity, and imagination, and is not a source of free creativity” (p. 36). This limitation does not invalidate the use of AI in heutagogy, but it does require that algorithmic personalization be complemented and constantly challenged by the genuine agency of the student as a subject—and not by the student as an object.

Although AI has the potential to adapt content and methodology to the individual needs of the user, this AI-mediated self-determination may or may not enable them to make decisions about their learning and actively participate in their educational process. Furthermore, it should foster reflective capabilities, as it can promote the ability to critically question, analyze, and construct knowledge. As Lo (2023) concludes, “its use raises several concerns, such as the generation of incorrect or false information and the threat it poses to academic integrity” (p. 12).

Plagiarism and academic integrity constitute another significant source of tension. The use of AI “faces strong cultural resistance in the academic sphere, creating tensions both in teaching practice and in the future prospects of education” (Delgado et al., 2024, p. 144). This resistance is not unfounded: if a student outsources the production of texts, arguments, or solutions to AI, they not only engage in academic dishonesty but—what is pedagogically more serious—forfeit the learning process that such production was intended to generate. From a heutagogical perspective, this represents a fundamental contradiction: the learner who delegates the task of thinking to an algorithmic system is not exercising their autonomy but rather abdicating it.



However, the opposite view also exists and deserves attention. Puente et al. (2024) report that “research indicates that students who use these platforms demonstrate better knowledge retention and achieve higher grades compared to traditional methods” (p. 61). This evidence suggests that the problem is not AI itself, but rather how it is used. A student who uses AI to compare perspectives, to generate questions that they then investigate through other channels, to receive feedback on their own arguments, and to explore areas of knowledge that would otherwise be inaccessible is exercising a form of technologically mediated autonomy that is perfectly compatible with heutagogy—and may even enhance it.

The interpretive key to this distinction lies in Gadamer’s (2003) hermeneutic spiral, in that learning—as a process of progressive understanding—involves each new horizon of meaning modifying and enriching the previous understanding. Interaction with AI can serve as a catalyst for this spiral movement, provided that the learner does not stop at the system’s first response but instead uses it as a starting point for new questions, new searches, and new syntheses. From this perspective, heutagogy is not a state to be attained but a process to be constructed: a continuous practice of critical and reflective appropriation of available technological tools.

That being the case, on the one hand, there are the enthusiasts who see in heutagogy almost infinite possibilities for advancement and progress, as if it were a new Enlightenment; and on the other hand, there are the skeptics who recognize the risk of the algorithm underlying AI and the results of an AI-led educational process—which, among other things, can become a mechanism of subjugation, with the subject algorithmically subjected within heutagogy.

At this point, it should be clarified that the individual should also be capable of critiquing the educational model itself; moreover, this would enable a contextual education to the extent that learning can be related to everyday reality and the challenges in one’s environment. This combination promises a more personalized, relevant, and emancipatory learning experience. This raises the following question: How does heutagogy relate to emancipatory education? Emancipatory education is “when we understand that education is a path to liberation not only for a select few but for everyone” (Garcés, 2020, p. 27). With AI, it is possible to reach everyone—at least in theory—although for now, inequalities and difficulties in accessing ICTs, and consequently AI, persist.

This first potential of AI does not, in principle, achieve global coverage, but as time goes on, we would expect the barriers preventing uni-



versal access to these technologies—which are here to stay—to be broken down. A second potential lies in AI’s ability to act as a personalized tutor that provides tailored tools and recommendations, further empowering individuals to make informed decisions about their education; of course, there is a risk of dependency here, which directly undermines the emancipatory purpose.

However, heutagogy itself seeks emancipatory autonomy by focusing on the learner’s autonomy, which aligns with the idea that each student has different interests, paces, and learning styles. That said, this implies that heutagogy must go beyond the mere transmission of content and allow for the holistic development of individuals, fostering their autonomy, critical thinking, and the ability to understand, make connections, and act—even regarding the educational content itself—within the “glocal” context, which is both global and local. In short, we must be cautious and not simply trust AI, but rather question it, since true emancipation involves the ability to question and decide for oneself.

196



Freeing Ourselves from Algorithmic AI

The functioning of an algorithm for the development of heutagogy could be described through the following stages:

- First, the collection of dynamic data: This means that, instead of static data, AI collects real-time information about the learner’s progress, thereby adapting teaching and learning strategies based on the learner’s program, concerns, preferences, strategies, interactions, and progress.
- Second, adaptive analysis: The data is processed using deep learning algorithms that identify patterns not only in performance but also in the learner’s style and motivation. Here it is important to note the risk involved in using black-box algorithms, “whose internal workings are unknown, cannot be explained, or are not understandable. This means that the algorithm makes its decisions and processes information in a way that is hidden, even from its own developer” (Morandín Ahuerma, 2023, p. 190). These algorithms make it difficult to understand and question their decisions, raising issues of responsibility and accountability.
- Third, the generation of personalized learning paths: In this step, rather than offering a simple assessment, unique educa-

tional pathways are designed and proposed that evolve with the learner—again using a black-box algorithm. This leads to the opacity that arises from the model's decisions, which are based on millions of internal calculations and connections that cannot be easily interpreted. These pathways provide specific—and likely unique—resources and challenges that foster student autonomy and critical thinking.

- Fourth, continuous feedback: In a successive and cyclical manner, immediate and contextualized feedback is provided, optimizing the learning environment based on the student's development and achieved goals.

The steps outlined above, which are cyclical, carry the risk of algorithmic subjugation in that the student's learning and decisions may be limited by biases inherent in the algorithm's design. This is especially true with deep learning models, since the model might reinforce preexisting patterns—such as learning styles preferred by the system—thereby ignoring cognitive diversity or limiting autonomous exploration.

What this process reveals is that each of these stages can also be a stage of subjugation. The perpetual collection of data implies constant surveillance of the subject. This raises questions regarding the individual's agency in their own educational process and the danger of excessive dependence on systems that, although adaptive, are far from neutral. Therefore, it is important to consider that the codification of training biases in algorithms (García Peñalvo, 2024) carries harmful consequences, since AI has the potential to perpetuate and amplify forms of discrimination, which justifies the need to critically analyze the ethical implications of AI in education to avoid perpetuating existing inequalities.

This dynamic of subjugation does not operate visibly. It is part of what Foucault (2006) termed “governmentality” (p. 96)—understood as the set of rationalities, techniques, and institutions that guide the behavior of subjects in a subtle and continuous manner. In the digital and algorithmic version of this concept, this steering requires neither physical institutions nor human overseers: it operates through interfaces, notifications, personalized recommendations, and gamification systems that shape behaviors and incentivize certain ways of engaging with knowledge. Deleuze (2006) anticipated this dynamic when he wrote that in “societies of control, nothing is ever finished: the enterprise, education, or service are the metastable and coexisting states of a single modulation, a kind of universal shaper” (p. 280). Educational AI systems can function



precisely as that universal deformer: they never conclude their assessment of the student, never declare their learning complete, and always suggest a next step, a further optimization.

As López (n.d.), Collins and Pinch (1996), Vallès-Peris (2022), and Winner (2008) point out, technology is not neutral. It can reflect and reinforce certain values and interests—a sort of technological delegation over lives and academic decisions—which, if automated without a deep understanding of its limitations and underlying biases, has a significant impact on the life of the heutagogical learner and on society at large. For this reason, there is a call to ensure that these systems are fair, transparent, and do not perpetuate discrimination or biases inherent in the programmers' views as written in code.

Given this landscape, it is important to identify the main ethical and technical concerns that the specialized literature has highlighted regarding the use of AI in education:

- **Algorithmic bias:** This carries the potential to maintain, exacerbate, and reproduce patterns embedded during the training of algorithms, thereby increasing inequality rather than reducing it (Ramírez, 2023; García Peñalvo, 2024). If the data used to train a system reflect historical inequalities in terms of race, gender, social class, or geographic origin, the system will tend to perpetuate them and, in many cases, amplify them, as it applies them on a massive scale and with a semblance of technical objectivity that makes it difficult to challenge.
- **Lack of transparency:** This results from the use of “black box” algorithms, which makes it difficult to understand and question their decisions, and raises potential issues of responsibility and accountability (Morandín Ahuerma, 2023). When a system determines that a student has greater or lesser potential, or when it steers the student's educational path toward certain areas of knowledge, that decision is produced through millions of internal calculations that are neither explainable nor auditable. This opacity is incompatible with an education system that aims to foster critical and autonomous individuals.
- **Technological determinism:** This reflects concerns about excessive reliance on algorithms at the expense of human judgment and discretion. When teachers and institutions delegate fundamental pedagogical decisions—such as what to learn, at what pace, with what resources, and under what assessment



criteria—to AI systems, they abdicate their own responsibility and transfer to the machine a power to define human education that no technical system should possess.

- Privacy and data protection: Serious concerns arise regarding the collection, storage, use, access, or leakage of students' personal data (UNESCO, 2022; Parlamento Europeo, 2024). Educational AI systems require massive amounts of data to function, and the collection of such data on minors or on adults in vulnerable situations raises critical questions about informed consent and about who ultimately controls the information generated during learning processes.
- Inequalities in access: Limited access—due, for example, to geographic location, socioeconomic status, prior knowledge of ICT, or age—can exacerbate existing inequalities and represents a structural barrier that may worsen existing educational inequities (Morduchowicz, 2023). Heutagogy that fails to address this access issue risks being an emancipatory approach only for those who already have basic resources guaranteed.

Identifying these concerns does not lead to the conclusion that AI should be rejected as an educational tool, but rather to the conviction that its integration requires constant critical engagement that examines its conditions of possibility, its effects, and its limits. Deleuze and Guattari (2002) introduced the concept of a “line of flight” to refer to movements that escape established territorialities and open up new possibilities for existence. The question that arises here is whether it is possible to use AI itself as a line of flight: to design, train, and use it in such a way that it generates the conditions for its own questioning and transcendence.

What appears to be a growing use of —and perhaps a growing dependence on— AI not only highlights technical and ethical concerns, such as bias and a lack of transparency; it also reconfigures the dynamics of power and control. By automating processes and decisions, algorithms not only perpetuate inequalities and biases but also become devices of a new form of governmentality (Foucault, 2006): “digital governmentality.” This change goes beyond the merely technical to become part of a socio-political transformation where power relations are diffused into algorithmic networks that shape subjectivities in a more subtle and continuous manner, integrating into the logic of the societies of control described by Deleuze (2006) and Foucault (2002), and leading to a form of subjugation that is less visible but disturbingly effective.



We are entering a historical moment in which regulatory mechanisms operate fluidly and persistently through contemporary technological environments. This historical progression reveals a shift from sovereign structures toward disciplinary systems (Foucault, 2002), culminating in the current paradigm of surveillance, as Deleuze (2006) notes: “the cybernetic machines and computers of control societies” (p. 274). Almost prophetically, Deleuze (2006) anticipated and pointed out that contemporary surveillance and regulation do not require visible institutional structures but rather operate through automated systems, foreshadowing the existence of an interconnected digital ecosystem and drawing on multiple—if not thousands—of mathematical calculations and computational procedures that power devices for monitoring and influencing people. This leads to a reconfiguration of identity in which the individual today faces new mechanisms of conditioning that continuously shape their aspirations, cognitions, and behaviors.

Advanced computational environments and mathematical and algorithmic processing systems—which, in the absence of consensus, we have come to accept as “intelligent”—have evolved toward an individualized adaptation of digital interfaces and resources, characterized by inherent biases that intensify observational capabilities and undermine individual self-determination. Thus emerges a contemporary configuration featuring a new type of subject—one not identified by traditional numerical classifications, but rather as a conditioned entity.

This is an individual whose personal imprint is under constant scrutiny, adjustment, and manifestation in digital spaces, creating an algorithmically mediated identity that raises fundamental ethical questions about decision-making capacity, independence, and the possibilities of preserving a genuine sense of identity in circumstances where control manifests itself in an inherent, imperceptible, tacit, personalized, and possibly consensual manner—by accepting contractual terms upon entering digital spaces, applications, or artificial cognitive systems. One must ask: Is there a true understanding of these agreements?

With AI in education, the logic of control takes another step forward; AI has the capacity to monitor and evaluate academic performance, but it can also predict trajectories, suggest pathways, and shape expectations. In short, it creates a new type of subjectivity—algorithmic subjectivity—in which the subject’s autonomy is called into question. This AI-mediated personalized learning is a new form of the “technologies of the self” (Foucault, 1990). Furthermore, it can be interpreted as a new form of what Foucault (2002) termed “individualizing power” (p. 198), in a loop

of constant training and consultation, to which we add Deleuze's (2006) argument: "In societies of control, nothing is ever finished: the enterprise, training, or service are the metastable and coexisting states of a single modulation, a kind of universal deformer" (p. 280). Now, Deleuze and Guattari (2002) introduce the concept of the "line of flight"—what if AI itself were used as a line of flight, an emancipatory education of AI itself?

Mezzadra (2014), drawing on Marxism, has explored the ways in which individuals constitute themselves and resist through struggles and practices of emancipation—a question that must be raised in the face of an "algorithm of subjection," as algorithmic governance desubjectivizes "the subjectivity of the individual subject" (p. 94). Subjects immersed in heutagogy can seek forms of resistance and organization to challenge the subjectification imposed by AI, which fails to consider their unique and singular perspective and which, moreover, enables digital governmentality. This approach tends to ignore subjectivity and individual particularities, which can lead to a distorted representation of their capabilities and potential.



Heutagogical Emancipatory Education

Emancipatory education cannot be merely a philosophical ideal: it must translate into concrete practices, specific institutional designs, and ways of engaging with knowledge and technology that bring its principles to life. In this sense, heutagogy offers a unique opportunity: the possibility of designing systems that not only personalize learning but also equip the individual with the ability to question the conditions of that personalization.

De Sousa Santos (2006) warns of the difficulty of this undertaking: "we have only hegemonic tools at our disposal to try to address all of this" (p. 65). The AI systems currently available are, for the most part, products of technology corporations driven by market logic and optimization that do not necessarily align with the objectives of an emancipatory education. However, this observation does not preclude the possibility of a critical appropriation of these tools. The history of education is full of examples of the subversive appropriation of hegemonic instruments: writing, the printing press, radio, video, and the internet. In each case, the instrument was developed with specific power dynamics in mind and was later appropriated by educational and social movements that used it against those very same dynamics.

AI, by providing rapid data and analysis, can be a valuable tool in this process; of course, we must remember that emancipation is not just

about accessing information, but also about questioning power structures and fighting for a more just society. Heutagogy promotes the ability to learn how to learn. This “learning how to learn” can be guided by parameters that disrupt hegemonic models and tools, linking the ability to search for, process, and evaluate automated information—but in a way that is critical and emancipatory of ideologies, sensologies, biopolitics, normalizations, and other mechanisms of subjugation of the subjugated subject.

Emancipation through heutagogy requires a critical review of assessment systems. When indicators of educational quality are reduced to metrics generated by organizations such as the Organization for Economic Cooperation and Development (OECD), “with criteria based on instrumental and efficiency-oriented reasoning” (Habermas, 1999, p. 42), a fundamental contradiction arises with the objectives of emancipatory education: the system that should liberate the individual subjects them to standardized metrics that neither capture nor value the most significant aspects of their development. Heutagogy must, therefore, also generate its own modes of assessment, focused not on the efficiency of performance, but on the depth of understanding, the quality of critical thinking, and the ability to act autonomously and responsibly in the world.

With heutagogy, the autonomous individual—who is responsible for their own learning process—makes decisions about what, how, and when to study. This is where AI can be used (and properly utilized); as noted, it may be insufficient or biased, but that does not mean this is always the case. In fact, it can be properly trained—but what does it mean to train it properly, and who should do so? More unresolved questions arise, since any bias becomes entrenched in AI, according to Angwin *et al.*: “There are cases showing how intelligent systems distort data, increasing risks” (cited in Ramírez, 2023, p. 12). In any case, at this point, the risk of biased training should be clear, and for that very reason, the code can be configured to be sufficiently broad, inclusive, and bias-resistant, thereby placing the responsibility on the individual to discern and critically evaluate what is presented to them.

It is still an idea that needs refinement, but it remains interesting, challenging, and, above all, promising to consider how to combine these approaches to achieve an emancipatory, autonomous, and heutagogical education. Below are some related points.

Consider the need to use AI in an ethical and bias-free manner, similar to what is outlined in the UNESCO document (2022) and the European Parliament regulation (2024). Likewise, one can draw on manifestos and declarations from anti-establishment groups that advocate for



ICTs and an internet that are accessible to all; one could even use AI itself to detect and resolve ethical and bias-related issues, ensuring they align with ethical values and social expectations, but without succumbing to political correctness-driven censorship.

The integration of AI into education raises important ethical questions that must be considered to ensure the responsible and beneficial use of both existing and emerging technologies, focusing on how decisions affect students as well as society as a whole. In particular, issues such as data privacy, algorithmic bias, and transparency in the use of algorithms must be addressed.

AI can provide rapid access to resources and help students research what they consider relevant. Heutagogy fosters assisted autonomy, which enables the independent exploration of knowledge and can serve as a tool to transcend the limitations of hegemonic educational paradigms. Heutagogy not only facilitates access to knowledge but also fosters emancipatory and inclusive learning, empowering learners to take control of their own learning and equipping them with the ability to question and reconstruct the traditional structures of knowledge-power (Foucault, 2002) that shape a subjectivity submissive to the hegemonic order.

AI can automate administrative tasks such as grading and performance tracking. This frees up time for teachers to focus on interacting with students. However, care must be taken to ensure that assessment is not based solely on quantitative scores generated by AI, which brings us back to the need to avoid falling into the instrumental rationality (Habermas, 1999) of indicators such as those used by the OECD.

In a future educational model such as that proposed by heutagogy, it is important to know how to balance automation with depth and the development of critical thinking, fostering in students that Kantian “dare to think” —the **sapere aude**— which, according to Kant (2003), is the motto of the Enlightenment:

If I have a spiritual guide who has faith for me, if I have a doctor who judges my diet for me, and so on, then I do not need to make an effort on my own. I have no need to think, when I can simply pay (p. 1).

Kant’s warning is perfectly applicable to the context of AI in education: a system that thinks for the student, selects for them, and evaluates for them, exactly replicates the figure of the heteronomous tutor that the Enlightenment sought to overcome. Emancipatory heutagogy must, on the contrary, use AI as a condition for autonomous thinking, not as a substitute for it. Thus, we must ensure that technology does not replace



human reflection, that AI is not that “spiritual guide” or the “and so on” indicated by Kant (2003), and, on the contrary, that the student makes an effort on their own, though, of course, they may make use of technological tools to optimize the cognitive, rational, and thinking process, as has been explored in virtual philosophy education in prisons (Sarmiento *et al.*, 2024). In this sense, heutagogy can help and empower students; of course, we must remember that true emancipation involves the ability to question and decide for oneself—even regarding the AI that drives heutagogy and heutagogy itself—as Kant (2003) would say, “*sapere aude*.”

A Decalogue for Heutagogy

204



To summarize and conclude this paper, we propose some ideas for the eventual implementation of heutagogy. As demonstrated, this self-determined educational approach necessitates a focus on fostering critical autonomy, promoting inclusive education, and ensuring that technology is used as a means to empower individuals, who should in no way be subjected to AI.

1. Prioritize human thought as the starting point. While heutagogy seeks to foster self-directed education, the individual's creativity and critical reflection must be prioritized and maintained as an indispensable goal, above and beyond the codified capabilities of any algorithmic system. One possibility is to foster “rhizomatic” thinking (Deleuze and Guattari, 2002), which offers a productive model of resistance to the cognitive homogenization that poorly managed AI tends to produce, so that it does not obey hierarchical or linear patterns, and the search for and establishment of multiple and unexpected connections between ideas and concepts is enhanced. The aim is for students to develop new intellectual and creative possibilities that are not limited to reproducing preexisting patterns. Following Foucault (1990, 2002), this resistance to homogenized thought serves as a means of subverting the disciplinary systems that shape docile subjectivities.
2. Foster synergies between the individual and technology in relation to AI and its use. This must be based on the pursuit of co-creation—not as one-way instruction—valuing human singularity and thereby considering, from the perspective of Sousa Santos (2006), a way to reinvent forms of collaboration between

humans and technology, so that the use of technology is prioritized to the extent that it helps eliminate inequalities and promotes global cognitive justice. The aim is for synchronized intelligence—human and AI—to be interpreted as an assembly of the collective machine in which each part contributes what the other lacks (Deleuze and Guattari, 2002), within a plane of ethical and political consistency that prioritizes technological use that expands possibilities over that which restricts them.

3. Strengthen autonomy in accessing and evaluating information. Education in critical autonomy must begin by recognizing and denaturalizing the power structures that shape information and determine which information is visible, valid, or permitted in digital environments. This reflective access to information must be part of the emancipatory educational process (Garcés, 2020), in which the individual navigates and frees themselves not only from social and digital mechanisms of subjugation but also avoids becoming trapped in or perpetuating digital bubbles. Universal access to ICTs and AI is key at this point, accompanied by literacy that is no longer merely digital but AI-mediated; this, in itself, is an exercise in liberation rather than subordination—an exercise in emancipation.
4. Promote self-regulated and intentional learning. Heutagogy fosters individuals' ability to design, plan, develop, evaluate, and provide feedback on their own learning within a continuous "becoming" (Deleuze, 2006). This becoming is a process in which learning is neither closed nor predetermined; it is immersed in a constant dynamic of transformation in which AI is used not as a source of dependence, but, on the contrary, as an opportunity to discover new lines of flight (Deleuze, 2006) that expand one's potential.
5. Foster ethical and pluralistic argumentation. Pluralistic thinking is a prerequisite for a democratic society (Garcés, 2020). This approach resonates with Deleuze and Guattari (2002), in that it requires learning to argue ethically—which involves navigating the realm of knowledge while respecting differences and avoiding the dogmatic territorialization of ideas. In this regard, De Sousa Santos (2006) reinforces the need to construct knowledge based on dialogue among diverse epistemologies, moving beyond monocultural perspectives and promoting the world's pluralism.





6. Promote creation with transformative impact. The capacity to create content must be pursued with a critical and emancipatory approach; this involves revealing how technologies shape subjectivities (Foucault, 1990, 2002), which must be challenged, since these subjectivities—in step with ICTs—condition and perform. This creation, as a political act, allows the subject to liberate themselves to the extent that they not only reproduce information but also reinterpret it to destabilize dominant narratives and construct alternatives.
7. Expand opportunities for active participation. Participation in glocal projects—which connect the local with the global—aligns with Sousa Santos's (2006) vision of active and solidarity-based citizenship. Garcés (2020), for his part, proposes participation as a tool for emancipation that strengthens individuals' capacity to influence their realities. This process can be understood as a Deleuzian assemblage that brings together human and non-human actors in networks of transformation.
8. Educate in the principles of reflective heutagogy. Once established, heutagogy must, of course, be explained through heutagogy itself, moving beyond mere technical proficiency to focus on a deep understanding of the power structures that govern technologies. We must analyze how algorithms perpetuate inequalities and how digital platforms operate as disciplinary devices that no longer merely control the individual but produce them: "Discipline thus manufactures subjugated and trained bodies, 'docile bodies'" (Foucault, 2002, p. 142). Heutagogy must empower individuals to dismantle these structures and build inclusive alternatives, "to explore new possibilities for rethinking existing tools and using them as devices of emancipation" (Triana Llano, 2025).
9. Build inclusive and horizontal learning communities. Educational communities should be spaces for encounter that challenge traditional structures and hierarchies. These communities should promote an "ecology of knowledge" (De Sousa Santos, 2006), where marginalized forms of knowledge are recognized and valued, and capabilities and knowledge that have been deliberately hidden or disparaged are empowered. Heutagogy can facilitate these types of processes by integrating it as an ethical mediator rather than as a hegemonic tool.

10. Rethink educational objectives in light of emerging challenges. From the perspective of Deleuze and Guattari (2002), educational objectives must be open, flexible, and adaptable to the changing and complex dynamics of the contemporary world; to this end, there is a need to educate students who are prepared to face the unknown with creativity and autonomy. This includes a critical scrutiny of the ways in which AI redefines power relations in the educational sphere.

Conclusions

The main objective of this research was to critically analyze the intersection between heutagogy and AI, with the aim of determining the conditions under which AI can genuinely promote self-determined learning rather than generate new forms of technological dependence. The analysis conducted allows us to affirm that this objective has been addressed in all its complexity, although not fully resolved—a finding consistent with the paradoxical nature of the problem at hand.

Regarding the central research problem—the coexistence of AI's emancipatory potential from a heutagogical perspective and the risk of technological subjugation—the analysis shows that this coexistence is not a contradiction that must be resolved in favor of one of its terms, but rather a productive tension that must be kept active and conscious in any proposal to integrate AI into educational contexts. AI is, in and of itself, neither emancipatory nor alienating: its nature depends on the pedagogical, ethical, and epistemological frameworks within which it is integrated, on the objectives that are defined, on the individuals who use it, and on the institutional conditions in which it is deployed. This main conclusion has an immediate practical implication: it is not possible to delegate pedagogical responsibility to the technological system. The design of AI-mediated educational experiences requires intellectual, ethical, and political work that teachers, institutions, and students themselves must actively undertake.

The analysis has also identified the main challenges facing any heutagogy initiative. First, the epistemic limitations of current AI systems: their inability to engage in genuine reasoning, their tendency to generate “hallucinations,” and their reliance on training data that often reproduces historical biases. Second, the problem of algorithmic opacity, which hinders critical scrutiny of the decisions these systems make regarding



students' learning trajectories. Third, structural inequalities in access to digital technologies, which mean that heutagogy is, in current practice, a possibility reserved for those who already have resources and favorable conditions. Fourth, cultural resistance and institutional uncertainty regarding the integration of AI into educational systems, which often leads to ad hoc regulations or bans that fail to address the root of the problem.

In the face of these challenges, the proposed ten-point guide does not aim to offer definitive solutions, but rather guidelines for pedagogical work that recognizes and manages the complexity of the problem. Emancipatory heutagogy is a work in progress, not a finished model; it requires experimentation, critical evaluation, and ongoing adjustment.

Future research should move forward in at least four directions. First, the empirical evaluation of heutagogy proposals in specific educational contexts, particularly in Latin America, where conditions of access, pedagogical traditions, and power dynamics have unique characteristics that proposals from the Global North do not fully capture. Second, the development of ethical and legal frameworks that regulate the use of AI in education based on the principles of cognitive justice and equity, moving beyond the criteria of instrumental efficiency that currently predominate. Third, research on teacher training models that equip educators not only with the technical use of AI tools but also with the capacity for philosophical and ethical reflection on their pedagogical implications. Fourth, the design and testing of AI systems specifically geared toward heutagogy, which make students' critical autonomy their primary optimization criterion, rather than standardized performance metrics.

In summary, heutagogy represents a genuine opportunity to rethink education in an algorithmically mediated world. But this opportunity can only be achieved if we approach it with full awareness of the paradox that defines it: the very tool that can enhance autonomy can also erode it, and the difference between these two outcomes lies not in the technology itself, but in the quality of the critical, ethical, and pedagogical thinking with which it is approached. Education must ultimately remain a path toward freedom: and technology, including AI, must always be a means to that end, never a substitute for it.

Bibliography

ABELIUK, Ariel & GUTIÉRREZ, Carolina

2021 Historia y evolución de la inteligencia artificial. *Revista Bits de Ciencia*, (21), 14-21. <https://bit.ly/4iEDXf0>



- AGUILAR, Xavier
 2024 *Seis casos de uso de la IA en las aulas que cambiarán la educación en 2024*. Barcelona: Universitat Oberta de Catalunya. <https://bit.ly/3OVSGUZ>
- ALONSO RODRÍGUEZ, Ana
 2021 Objetividad y verdad en la ciencia de la educación como ciencia de diseño. *Sophia, Colección de Filosofía de la Educación*, (31), 113-135. <https://doi.org/10.17163/soph.n31.2021.04>
- BLASCHKE, Lisa & MARÍN, Victoria
 2020 Aplicaciones de la heutagogía en el uso educativo de e-portfolios. *Revista de Educación a Distancia*, 20(64). <https://doi.org/10.6018/red.407831>
- CALDEIRO, Gabriel, CHAMORRO, Florencia, GONZÁLEZ, Nicolás, KUITCA, Ariel & MILILLO, Cristian
 2024 *Inteligencia artificial y aprendizaje activo: investigación y diseño de estrategias de enseñanza con IA en escuelas*. Buenos Aires: Fundar; PENT-FLACSO. <https://bit.ly/4gBvXcM>
- COLLINS, Harry & PINCH, Trevor
 1996 *El Gólem: lo que todos deberíamos saber acerca de la ciencia*. Barcelona: Crítica.
- DELEUZE, Gilles
 2006 *Conversaciones*. Valencia: Pre-Textos.
- DELEUZE, Gilles & GUATTARI, Félix
 2002 *Mil mesetas: capitalismo y esquizofrenia*. Valencia: Pre-Textos.
- DE SOUSA SANTOS, Boaventura
 2006 *Renovar la teoría crítica y reinventar la emancipación social*. Buenos Aires: CLACSO.
- DELGADO, Jorge, ALMANZA, Juan & MACÍAS, José
 2024 Análisis bibliométrico de la inteligencia artificial como herramienta en la enseñanza en la educación superior. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(6), 141-153. <https://doi.org/10.56712/latam.v5i6.2997>
- FOUCAULT, Michel
 1990 *Tecnologías del yo y otros textos afines*. Barcelona: Paidós.
- FOUCAULT, Michel
 2002 *Vigilar y castigar: el nacimiento de la prisión*. México: Siglo XXI.
- FOUCAULT, Michel
 2006 *Seguridad, territorio, población*. Buenos Aires: FCE.
- FREIRE, Paulo
 1970 *Pedagogía del oprimido*. México: Siglo XXI.
- GADAMER, Hans-Georg
 2003 *Verdad y método (vol. 1)*. Salamanca: Sígueme.
- GARCÉS, Marina
 2020 *El contratiempo de la emancipación*. Recurso de aprendizaje. Barcelona: FUOC.
- GARCÍA PEÑALVO, Francisco
 2024 Inteligencia artificial generativa y educación: un análisis desde múltiples perspectivas. *Education in the Knowledge Society (EKS)*, 25, e31942. <https://doi.org/10.14201/eks.31942>



- GARCÍA PEÑALVO, Francisco, LLORENS LARGO, Faraón & VIDAL, Javier
 2024 La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *Revista Iberoamericana de Educación a Distancia*, 27(1). <https://doi.org/10.5944/ried.27.1.37716>
- HABERMAS, Jürgen
 1999 *Teoría de la acción comunicativa: racionalidad de la acción y racionalización social*. Madrid: Taurus.
- HAIIRI, Faisal, MOHAMAD, Siti, SAAD, Saiful & PINANDITA, Taufiq
 2022 A Thematic Review on the Implementation of Heutagogy in Universities. *Journal of Theoretical and Applied Information Technology*, 100(21), 6686-6701. <https://bit.ly/49EAQPZ>
- HASE, Stewart & KENYON, Chris
 2001 *From Andragogy to Heutagogy*. University of Glasgow.
- HERMANN ACOSTA, Andrés
 2011 Pedagogía del ciberespacio: hacia la construcción de un conocimiento colectivo en la sociedad red. *Sophia, Colección de Filosofía de la Educación*, (11), 83-103. <https://bit.ly/4lkQyEF>
- KANT, Immanuel
 2003 *Respuesta a la pregunta ¿qué es la Ilustración?* <https://bit.ly/3BCIo9a>
- KUO, Ryan
 2023 Noam Chomsky: The False Promise of ChatGPT. *The New York Times*. 8 de marzo. <https://bit.ly/3ZQklU>
- KNOWLES, Malcolm
 1970 *The Modern Practice of Adult Education: Andragogy versus Pedagogy*. Nueva York: Associated Press.
- LO, Chin
 2023 What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- LÓPEZ, Diego
 s. f *Filosofía de la tecnología: entre lo humano y lo técnico*. Recurso de aprendizaje textual. Barcelona: FUOC.
- MEZZADRA, Sandro
 2014 *La cocina de Marx: el sujeto y su producción*. Buenos Aires: Tinta Limón.
- MOLLIK, Ethan & MOLLIK, Lilach
 2023 *Using AI to Implement Effective Teaching Strategies in Classrooms: Five Strategies, Including Prompts*. *The Wharton School Research Paper*. 17 de marzo. <https://doi.org/10.2139/ssrn.4391243>
- MORANDÍN AHUERMA, Fabio
 2023 *Principios normativos para una ética de la inteligencia artificial*. México DF: Consejo de Ciencia y Tecnología del Estado de Puebla. <https://bit.ly/41AXs1N>
- MORDUCHOWICZ, Roxana
 2023 *La inteligencia artificial: ¿necesitamos una nueva educación?* UNESCO. <https://bit.ly/4iErNCA>
- NARVÁEZ, Margarita & PRADA, Andrés
 2005 Aprendizaje autodirigido y desempeño académico. *Tiempo de Educar*, 6(11), 115-146. <https://bit.ly/4frmOTe>

- OBANDO, Edwin, VILLAGRÁN, Norma & OBANDO, Edson
 2018 La redefinición del sujeto cognoscente: el acto de intelección en cuanto conocimiento. *Sophia, Colección de Filosofía de la Educación*, (25), 89-109. <https://bit.ly/4kSmIHJ>
- PARLAMENTO EUROPEO
 2024 *Reglamento sobre la inteligencia artificial*. <https://bit.ly/3OXoGs4>
- PEÑALVER HIGUERA, Manuel, GUERRA CASTELLANOS, Yetzy, RODRÍGUEZ, Lino & LÓPEZ, Rosario
 2024 Transformando la educación con inteligencia artificial: hacia un aprendizaje personalizado en la Era 4.0. *Revista de Ciencias Sociales*, 30(4), 416-430. <https://doi.org/10.31876/rcs.v30i4.43040>
- PUENTE, Sonia, BAJAÑA, Luis, SERRANO, Carlos & VALLEJO, Karina
 2024 La inteligencia artificial como recurso educativo en la educación superior. *Recimundo*, 8(3), 48-67. [https://doi.org/10.26820/recimundo/8.\(3\).julio.2024.48-67](https://doi.org/10.26820/recimundo/8.(3).julio.2024.48-67)
- RAMÍREZ, Rodrigo
 2023 Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial: una revisión documental. *Entretextos*, 15(39), 1-17. <https://doi.org/10.59057/iberoleon.20075316.202339664>
- RANCIÈRE, Jacques
 1996. *El maestro ignorante*. Barcelona: Laertes.
- SARMIENTO, Ingrid, ESPARZA OVIEDO, Silvia & TRIANA LLANO, Luis
 2024 Educación filosófica virtual para la reinserción social de personas privadas de la libertad. *EducaT*, 5(1), 63-67. <https://doi.org/10.22490/unad.27452115.8932>
- SYNTHESIS
 2024 *Synthesis School*. <http://bit.ly/4fiduB0>
- STIEGLER, Bernard
 2015 *La société automatique 1: L'avenir du travail*. París: Fayard.
- TRIANA LLANO, Luis
 2025 Deleuze en la era digital: poder, diferencia y subjetividad. *Revista AION*, 66. <https://lc.cx/gLzbMw>
- UNESCO
 2022 *Recomendación sobre la ética de la inteligencia artificial*. <https://bit.ly/3OSIKwK>
- VALLÈS-PERIS, Núria
 2022 *Retos tecnocientíficos*. Recurso de aprendizaje textual. Barcelona: FUOC.
- WINNER, Langdon
 2008 *La ballena y el reactor: una búsqueda de los límites en la era de la alta tecnología*. Barcelona: Gedisa.



Author Attribution Statement—CRediT Taxonomy	
Author	Contributions
Luis Alberto Triana Llano	Since this is a single-author article, the total contribution is attributed to the same author. The content presented in the article is the sole responsibility of the author.

Statement on the Use of Artificial Intelligence
The author, Luis Alberto Triana Llano, declares that the preparation of the article titled “The Paradox of Heutagogy and Artificial Intelligence” was supported by artificial intelligence (AI). The Claude Sonnet 4.6 (Anthropic) artificial intelligence system was used exclusively as an editorial assistance tool: to improve writing, organize sections, and formally develop ideas already conceived by the author. All arguments, the central thesis, the theoretical framework, the critical analysis, and the conclusions are the author’s own. Furthermore, the author assumes full responsibility for the content of the article and has verified and validated all the information presented.

Date received: December 14, 2024
Review date: February 20, 2025
Date of acceptance: May 20, 2025
Publication date: July 15, 2026