Received 16 June 2025
Revised 3 September 2025
22 October 2025
Accepted 3 November 2025

Tools or crutches? Budgeting human and machine autonomy when introducing GenAI in education

Francesco Balzan
Department of Computer Science and Engineering, University of Bologna, Bologna, Italy and
Department of Computer Science, University of Pisa, Pisa, Italy
Lorenzo Angeli
Department of Information Engineering and Computer Science, University of Trento, Trento, Italy
Ralph Meulenbroeks
Freudenthal Institute, Utrecht University, Utrecht, Netherlands, and
Federica Russo
Freudenthal Institute, Utrecht University, Utrecht, Netherlands and
Department of Science and Technology Studies, University College London, London, UK

Abstract

Purpose – Generative AI (GenAI) in education brings renewed attention to learner autonomy – that is, whether learners can think and act independently. GenAI offers the promise of learning efficiency and personalization, while raising questions about its alignment with nurturing autonomous learners. In this paper, we present a theoretical framework to investigate the relationship between GenAI and learner autonomy, to guide the design of educational environments that are safe and autonomy-supporting.

Design/methodology/approach – Our paper explores the multifaceted nature of autonomy across the cognitive, philosophical, political and computing fields, connecting theories such as self-determination theory with reflections on machine autonomy. Leveraging Latour's Actor-Network Theory, our framework aims to elucidate how autonomy is distributed between human and non-human actors in educational environments.

Findings – Our main contribution is the process of "autonomy budgeting", viewing autonomy as a resource that is allocated and traded off between an ensemble of actors. Autonomy budgeting works as a guiding conceptual tool for researchers, educators, curriculum designers and policymakers to assess and manage the autonomy trade-offs involved in integrating GenAI into educational environments.

Research limitations/implications – By re-centering the learner's agency and capacity for self-regulation, autonomy budgeting provides a way to conceptualize and operationalize autonomy within AI-mediated education, and to navigate the complex interplay between human and machine agency in education.

Originality/value – Our framework develops reflections on the socio-technical nature of educational processes, where technologies act as co-participants rather than neutral tools. Autonomy in education, becomes a multifaceted construct that spans (human) cognitive, epistemic and political domains, and must be considered vis-a-vis varying degrees of machine autonomy.

Keywords AI in education, Autonomy, Self-determination theory, Machine autonomy, Autonomy budget

Paper type Conceptual paper

Funding: This work was supported by the Italian National Recovery and Resilience Plan (NRRP) (award number: J33C22002830006).

Artificial Intelligence in Education
Vol. 2 No. 1, 2026
pp. 18-32
Emerald Publishing Limited
3049-5474
DOI 10.1108/AIIE-06-2025-0145

1. Introduction

The rise of generative AI (GenAI) technologies – such as conversational copilots, large language models and automated assistants – marks a profound shift in how technology mediates human activity. These tools, designed to assist in tasks ranging from creative writing to software development, are rapidly integrating into educational contexts, giving rise to what we term GenAI in Education (GenAIED). While GenAIED tools promise to revolutionize learning by offering personalized support, immediate feedback and access to vast knowledge repositories, their implications for the development of learner autonomy remain underexplored. This prompts us to define a first area of inquiry.

RQ1. Do GenAIED tools align with, or conflict with, the educational goal of cultivating autonomous subjects?

To address this question, we adopt a holistic approach to education, viewing it as a multi-level process: (1) biological/cognitive, as an adaptive system responding to its environment, (2) distributive/technological, shaped by interactions with other systems and tools and (3) philosophical/political, as the pursuit and negotiation of collective aims. Within this framework, self-determination theory (SDT) offers a valuable lens for analyzing the psychological conditions that support motivation and learning. According to SDT, learners thrive when three basic psychological needs–autonomy, competence and relatedness–are satisfied (Ryan and Deci, 2018, 2020; Gerlich, 2025). Designing autonomy-supportive environments is thus central to fostering meaningful and sustained engagement (Reeve and Cheon, 2021). In what follows, we explicit how each theoretical component contributes evidence or conceptual clarity for RQ1.

We argue that GenAI systems introduce novel tensions between control and autonomy–particularly by automating core cognitive processes–that invite a reinterpretation of autonomy-supportive environments in light of the partial delegation of thinking. In response, we propose to extend SDT by supplementing its psychological foundations with concepts drawn from philosophy, science technology studies and computer science. This interdisciplinary integration leads to a more distributed conceptualization of autonomy, enabling us to rethink how learner agency is shaped when cognitive tasks are shared with increasingly autonomous machines. This move grounds our second research question.

RQ2. How can we design GenAIED practices, tools and policies that preserve or enhance learner autonomy while leveraging GenAI's benefits?

This paper's key contribution lies in delineating this expanded view of education and autonomy through the concept of an "autonomy budget", which provides a framework to help educators think about how GenAIED systems allocate agency between students and algorithms in specific contexts. By describing the trade-offs in decision-making power, autonomy budgeting empowers practitioners to design AI-mediated learning environments that safeguard students' capacity for independent judgment while leveraging some of GenAIED's claimed benefits. Autonomy budgeting repositions education as a holistic endeavor, where technologies are assessed not only by short-term efficiency gains but also by their alignment with the deeper goals of human development: fostering individuals who can critically engage with tools, systems and societies. Later in the paper, we also outline how this concept can be operationalized and tested empirically (e.g. autonomy-related survey instruments, classroom experiments and mixed-methods field studies), providing a clear pathway from theory to measurement.

The paper's structure is as follows. Section 2 assembles the theoretical foundations (addressing RQ1). Section 3 extends these foundations to GenAIED and refines constructs (still addressing RQ1). Section 4 introduces the "autonomy budget" and translates it into a practical procedure with empirical touchpoints (addressing RQ2). The conclusion synthesizes implications for design and policy, delineates limitations and outlines steps for future research.

2. The goal of education: nurturing autonomous subjects

We conceptualize education as a continuum that spans three interrelated dimensions: (1) the biological and cognitive foundations of learning, rooted in individual development, (2) the distributed and technological systems that mediate educational practices, and (3) the philosophical and political values that frame the purpose of education at a societal level. Each dimension contributes uniquely to the development of autonomy, understood here as the capacity to act with critical self-determination in and on the world.

2.1 Biological and cognitive foundations

At the individual level, education is grounded in biological and cognitive mechanisms that enable learning and adaptation. Unlike most other species, humans exhibit neoteny–the retention of juvenile traits into adulthood–which fosters lifelong neuroplasticity and learning potential (Gould, 1990). This biological plasticity, coupled with evolved mechanisms for social learning and pedagogical behavior (Laland, 2004), underpins autonomy as a dynamic, self-correcting process: learners form hypotheses, assimilate feedback and iteratively refine their mental models (Laland, 2017).

However, this same plasticity renders learners vulnerable to cognitive hijacking by cultural or technological systems–such as the exploitative dynamics of attention economies in social media. This dual potential makes autonomy not merely a trait to be cultivated, but a faculty to be protected and critically guided.

2.2 Distributive/technological foundations

Human learning is fundamentally distributed: it is scaffolded by interactions with other people, tools and cultural artifacts (Tomasello, 2009; Hutchins, 2000; Clark and Chalmers, 1998). Educational practices thus extend cognition beyond the individual brain, transforming learning into a socio-technical system of delegation, coordination and feedback. This systemic perspective has profound implications: it shows that autonomy is not reducible to internal control, but must be understood in terms of strategic delegation–knowing when to rely on external systems and when to retain or reclaim cognitive agency.

This insight echoes Richard Feynman's principle of "build-to-understand": true understanding requires learners to actively deconstruct and reconstruct the tools they use. Without this engagement, the risk is not just dependency but epistemic dispossession, a concern voiced since antiquity (Cassinadri, 2024).

2.3 Philosophical/political foundations

Education is also a political act, deeply influenced by cultural values and philosophies, which determine its purpose. Throughout history, societies have used education to cultivate specific types of citizens, ensuring the survival of collective norms or the dominance of certain ideologies. As Dewey noted, education is "the fundamental method of social progress and reform" (Dewey, 2022), but it is also used to perpetuate existing power structures (Bourdieu and Passeron, 2013).

Together, these dimensions shape the autonomy of an individual, that is, their capacity to act with critical self-determination within and on the world. We claim that autonomy, as an educational goal, serves both as a lens for designing educational practices and as a measure for assessing their effectiveness. Autonomy bridges the nurturing of individual cognitive development with the transformative preservation of collective values, mediated by the critical and self-regulated utilization of external knowledge sources (as we will discuss in Sec. 3). As mentioned, autonomy is a central concept within SDT, which we now illustrate.

2.4 Self-determination theory and the central role of autonomy

SDT is a comprehensive theory of human motivation, growth and well-being that has its roots in the humanistic tradition (e.g. Rogers, 1965). Since a pioneering publication by Ryan and

Deci (2000), it has gained a central position next to the cognitivist expectancy-value theories (e.g. Eccles and Wigfield, 2020) of human behavior. The central proposition of SDT is that humans, apart from their basic physiological needs (shelter, food, sex, etc.), have three basic psychological needs (BPNs). We adopt the definitions of these three needs, as recently given by SDT scholars, as follows.

(1) Autonomy: "the experience of volition and willingness. When satisfied, one experiences a sense of integrity as when one's actions, thoughts, and feelings are self-endorsed and authentic. When frustrated, one experiences a sense of pressure and often conflict, such as feeling pushed in an unwanted direction" (Vansteenkiste *et al*., 2009, p. 3).

(2) Competence: "the experience of effectiveness and mastery. It becomes satisfied as one capably engages in activities and experiences opportunities for using and extending skills and expertise. When frustrated, one experiences a sense of ineffectiveness or even failure and helplessness" (Vansteenkiste *et al*., 2009, p. 3).

(3) Relatedness: "the experience of warmth, bonding, and care, and is satisfied by connecting to and feeling significant to others. Relatedness frustration can come with a sense of social alienation, exclusion, and loneliness." (Vansteenkiste *et al*., 2009, p. 3).

Based on the satisfaction or frustration of these three BPNs, humans experience internalization or externalization of motivation, respectively. Internalization leads to more internalized, self-determined or autonomous types of motivation, whereas externalization leads to more controlled forms of motivation, based on some form of pressure. When applied to educational settings, a teaching style that is unilateral and demanding is referred to as "controlling", whereas a teaching style that is more reciprocal and flexible is referred to as "autonomy-supportive", as it has been shown to lead to more autonomous forms of motivation (Reeve and Cheon, 2021). To put it another way: supporting all three BPNs fosters *autonomous* motivation (Meulenbroeks *et al*., 2023). Previous research has shown that more controlled forms of motivation (i.e. internal or external pressure) led to deteriorated performance across cultures and fields of work (Chen *et al*., 2015; Vansteenkiste *et al*., 2009). The reverse has also been established: more *autonomous* types of motivation, based on self-determined goals or pure fun or interest, lead to increased performance (Cerasoli *et al*., 2014).

Recent studies have applied SDT to AI education contexts. For example Xia *et al*. (2022) demonstrated that SDT-based teacher support in AI education positively influenced students' perceptions and engagement, regardless of gender or achievement level. Similarly, Xia *et al*. (2023) found that satisfying the needs for autonomy and competence mediated the relationship between students' prior knowledge and their self-regulated learning when using AI chatbots. Complementing these findings, SDT-aligned design tools and measures have emerged for GenAI contexts, including a classification tool for fostering SRL with ChatGPT (Chiu, 2024) and a validated AI Motivation Scale grounded in SDT (Li *et al*., 2025).

These findings suggest that while SDT remains a robust framework, the integration of AI technologies into education introduces new dynamics that warrant further exploration. Specifically, the automation of cognitive processes by AI systems raises questions about the perceived locus of causality (PLOC) in learning.

In the context of GenAIED, the PLOC becomes extremely relevant (Ryan and Deci, 2018), as it reflects the extent to which individuals see the origins of their actions as internally driven or externally imposed. An internal perceived locus of causality (I-PLOC) denotes that learners feel a sense of agency and ownership over their educational pursuits, engaging with tasks out of intrinsic interest or self-determined goals. In contrast, an external perceived locus of causality (E-PLOC) implies that learners experience their actions as primarily influenced by external pressures, demands, or incentives, which can undermine genuine autonomy.

In this regard, GenAIED systems pose a novel challenge: they shift the site of agency in learning processes that are not solely under the control of students. When learners offload tasks like writing, summarizing, or problem-solving to an automated agent, does this support their autonomy by scaffolding competence–or does it risk displacing agency altogether?

This question motivates the central aim of this paper: to develop a theoretical framework for studying autonomy in AI-mediated educational contexts. While we draw on SDT, we argue that GenAI technologies introduce qualitatively new dynamics–particularly the automation and/or delegation of cognitive processes–that require an expanded and distributed understanding of autonomy. In the next section, we elaborate on how such a reconceptualization can inform both empirical research and the design of autonomy-supportive learning environments.

3. Autonomy in the GenAIED era

SDT emphasizes autonomy as central to an "autonomy-supportive environment" that fosters all BPNs (Reeve and Cheon, 2021). However, considering education's broader socio-technical and political dimensions, together with and the more and more pervasive adoption of GenAIED technologies with increasing levels of autonomy calls for a significant expansion of "autonomy" beyond its original psychological scope, which we undertake in this section.

3.1 Beyond autonomy: relatedness and competence in synthetic educational settings

As already mentioned in previous sections (1) we need to re-examine autonomy in light of GenAI tools used in education and (2) this is even more needed because autonomy has been given different definitions in different fields. Since this contribution aims at laying down the theoretical framework for further empirical studies in GenAIED connected to SDT, we also need to reframe and reconceptualise "relatedness" and "competence" in the GenAIED context. This responds to recent syntheses calling for clearer constructs and measures around GenAI's effects on learning and agency (Yusuf *et al*., 2024).

To begin with, it is important to look beyond students in isolation and beyond their individual satisfaction or frustration with BPNs. Students learn in environments that are not fully social, but also *hybrid*. Classrooms have always combined social interaction with the use of tools–from the abacus to the tablet. What changes with GenAI can be understood through *Actor-Network Theory* (ANT), first developed by Latour (2007) and later elaborated by others, including Bisconti *et al*. (2024).

In short, ANT argues that society is made up of many kinds of actors: humans, natural elements, technical artefacts and institutions. What matters is not whether these actors share the same "ontological" level (human vs natural vs. machine), but how their interactions shape social relations. In this sense, society has always been *hybrid*.

Bisconti *et al*. extend this view by noting that GenAI is a special kind of actor. Unlike earlier (digital) tools, it contributes to the *creation of semantic content*. This means we interact *with* GenAI, not just through it. While LLMs remain "statistical parrots," the fact that we can converse with them and generate meaning together with them forces us to reconsider how relatedness is formed. A central question, then, is whether using GenAI for writing, editing, coding, or translating strengthens or weakens this sense of relatedness, which until recently was reserved for human-to-human interaction.

Competence is also at stake. On the one hand, students may *feel* more capable because GenAI helps them produce polished output; on the other, they may lose touch with the learning process itself, undermining their BPN of competence. Yet in some cases, careful and targeted use of GenAI could actually help students master that process.

BPNs are, in one way or another, deeply affected by the introduction of GenAI tools and while the BPNs identified since their pioneering formulation (Ryan and Deci, 2018) may remain broadly valid, the introduction of GenAIED may make necessary a recontextualization in educational settings that are, following the terminology of Bisconti *et al*., not just hybrid but entirely synthetic.

3.2 Mapping enhanced autonomy to the three dimensions of education

In Section 2, we proposed understanding human education on three levels: biological/cognitive, distributed/technological and philosophical/political. In our view, the feeling of psychological autonomy described by SDT arises from the interplay between the three levels at which education operates, and three different meanings that autonomy can have: biological/cognitive, epistemic and political. This expanded framework not only captures the subjective experience of volition and self-endorsement emphasized by SDT but also situates it within the broader socio-technical and institutional context of education described above.

3.2.1 Biological/cognitive autonomy. At the biological level, autonomy emerges through self-organization and adaptive regulation (Maturana Rumesin and Varela, 1991). Neural mechanisms–especially in the prefrontal cortex–support goal-directed behavior and error correction (Damasio, 2000), providing the basis for cognitive self-regulation. Cognitive science further defines autonomy as the capacity to control one's mental processes, allowing for deliberation, planning and critical evaluation (Deci and Ryan, 2013). Together, these mechanisms create what SDT calls psychological autonomy–a subjective sense of agency and volition. Importantly, education and learning nurture these capacities, enabling flexible, self-directed responses that go beyond genetic constraints. In sum, while biological and cognitive autonomy free individuals from genetic determinism, effectively interpreting and using environmental information requires an additional layer of autonomy: epistemic autonomy, which involves both human and material agents.

3.2.2 Epistemic autonomy. This concept aligns with Kant's call for enlightenment –*Sapere Aude*, or "Dare to know" – which advocates *intellectual autonomy*: independence from external authority (Kant, 2019). However, epistemic autonomy is not absolute: Hume emphasized the necessity of trusting expertise, while Reid highlighted the role of *intellectual solidarity*, balancing skepticism with reliance on collective knowledge (Hume, 2000; Reid, 2011). Hume and Reid's interpretations align with the distributive and technological foundations of human educational processes, where knowledge acquisition and building are conceived as relational, shaped by social and material environments (Giddens, 1992). Therefore, building upon biological and cognitive autonomy, *epistemic autonomy* refers to the ability to critically evaluate knowledge within social and technological epistemic networks (Dragos, 2021) — a process that, in turn, supports the BPN of autonomy.

3.2.3 Political autonomy. Relates to the capacity for self-governance at both individual and institutional levels (Mill, 2022). Ethically, it involves acting according to self-imposed moral principles free from coercion, while philosophically it is rooted in rational deliberation and self-rule (Kant, 2019). In education, fostering political autonomy means preparing learners to navigate and critique power structures, thereby reinforcing their ability to act as ethical and responsible agents in democratic societies. This dimension is essential for transforming individual capacities into meaningful social participation, further contributing to the emergence of psychological autonomy.

Therefore, cognitive autonomy underpins self-regulation and adaptive learning; epistemic autonomy enables critical engagement with diverse sources of knowledge; and political autonomy empowers individuals to question and transform prevailing power structures. Importantly, these levels are interdependent: without a robust cognitive foundation, epistemic development is constrained and without both, genuine political autonomy cannot emerge. This dynamic interplay gives rise to the psychological autonomy central to SDT: the internalized, subjective experience of self-determination and volition is an effect of the gradual refinement of autonomy. By reconceptualizing SDT's psychological autonomy as an emergent property [1], we can better account for the distributed and relational nature of autonomy in contemporary educational settings. This perspective challenges the traditional view of autonomy as a static, individual attribute and highlights its gradual, context-dependent formation within socio-technical systems (Bächle, 2023; Simmler and Frischknecht, 2021; Thimm *et al*., 2024).

How does this expanded understanding of education and autonomy intersect with GenAIED technologies? To address this question, we must first introduce another disciplinary perspective on autonomy: machine autonomy.

3.3 Machine autonomy

From a socio-technical perspective, understanding human-machine collaboration requires examining both the degree of autonomy present in technical systems (Beer *et al*., 2014) and the level of automation they exhibit (Kaber, 2018). These aspects are critical for assessing how responsibilities and decision-making are distributed between humans and machines (Simmler and Frischknecht, 2021). However, the concept of machine autonomy remains ambiguous, with its definition and measurement still debated. To provide clarity, we adopt the taxonomy proposed by Simmler and Frischknecht (2021), which conceptualizes machine autonomy and task automation along five levels.

At the lower level of automation – and thus of machine autonomy – systems offer decision support by suggesting options that require human selection, or execute actions only after receiving explicit human approval. As systems increase in their level of automation, they may execute actions unless vetoed by a human, or even act independently while subsequently informing the human operator. At the highest level, a system operates fully autonomously, carrying out tasks without direct human intervention.

In the discussion of machine autonomy, an important aspect is transparency – that is, the extent to which users can trace how the system transforms inputs into outputs. Generally, as a system's autonomy increases, its internal processes become less transparent (Doshi-Velez and Kim, 2017). Another dimension is determinacy: simpler systems produce consistent outputs for a given input, while more advanced systems may exhibit indeterminacy, generating variable responses in dynamic contexts. Equally important is adaptability, referring to a system's capacity to learn from experience and adjust its behavior over time – a quality that has been notably enhanced in recent LLM-based agents. These agents demonstrate open-ended behaviors by incorporating new data sources and collaborating with other systems, reflecting an elevated level of operational independence [2].

We now turn to a comparison of human and machine autonomy, drawing on the three forms of autonomy outlined above (see Table 1). In humans, the autonomy that underpins self-determination develops gradually, through a linear progression of cognitive, epistemic and political autonomy. A partial parallel can be traced in machines. Advanced GenAI systems increasingly display traits resembling human cognitive and epistemic autonomy. For example, recent Large Language Model (LLM)-based agents can move beyond rigid programming–mirroring cognitive autonomy–by using larger context windows to adapt responses to previous interactions. They also show a form of epistemic autonomy by integrating information from diverse datasets to refine their outputs.

The divergence becomes clear when we consider political autonomy. Human agents can challenge and reshape top-down norms, but machine autonomy remains limited by design and alignment constraints. Even the most advanced systems are bound by the values and restrictions set during development (Mazeika *et al*., 2025).

This limitation is especially relevant in education. GenAIED agents may promote cognitive and epistemic autonomy by offering adaptive, context-sensitive support. Yet their inability to resist or revise embedded design biases risks reinforcing subtle forms of homogenization. For instance, a GenAI coding assistant may foster cognitive autonomy by helping students debug code (cognitive layer), while at the same time constraining political autonomy by embedding corporate-driven biases into their workflows (political layer) (Balzan *et al*., 2024). As a result, students may feel more autonomous in the short term, but in the long term, they could lose the capacity for critical dissent and transformative engagement with societal norms.

Table 1. Mapping foundations of education (see Sec. 2) to autonomy types

Educational foundation	Autonomy type	Definition and key concepts
Biological/ Cognitive	Biological/Cognitive Autonomy	Ability to flexibly adapt to the external environment, detached from genetic determination
Distributed/ Technical	Epistemic Autonomy	Ability to critically evaluate knowledge within socio-technical networks
Philosophical/ Political	Political Autonomy	Capacity to act as an ethical agent within democratic systems, negotiating power structures

Note(s): Psychological autonomy, in its SDT meaning, emerges from the interplay between the three different types of autonomy

Given these complexities, it becomes essential to integrate the SDT framework with our multi-level analysis of both human and machine autonomy. To capture and balance the dynamic interactions among cognitive, epistemic, and political dimensions in GenAIED environments, we propose the concept of an "Autonomy Budget" as a pragmatic tool for this synthesis.

4. The "autonomy budget"

4.1 Budgeting autonomy: actors and procedures

By adopting an ANT view, GenAIED tools can be treated as non-human technological actors, in principle equal in their ability to mediate semantic contents as their human counterparts. From an ANT perspective, all actors exhibit some form of autonomy, including GenAIED tools. Considering autonomy, then, as *distributed* between human and non-human actors reopens the question of whether increasing the machine autonomy of GenAIED tools (see Sec. 3.3), with its benefits of cognitive offloading for human actors, *necessarily* leads to an overall increase of human autonomy. We raise this question because human autonomy, as we discussed in Sec. 3.1 and 3.2, is not only a cognitive matter, but is also epistemic and political.

We propose addressing the potential impacts of autonomy redistribution through the metaphor of an *"autonomy budget"*. By following this metaphor, we wish to show that satisfying the need for autonomy of all the human actors in the system can be structured as a process of distribution of a limited resource. This view of autonomy as something that is budgeted and distributed follows up on the reflections proposed by scholars such as Formosa (2021): while the view held by Formosa and others focuses on the potential for the creation of positive-sum games by introducing machine actors, our approach takes a step back. By defining human autonomy as cognitive, epistemic and political, we call for stronger justification of claims of autonomy gains. Viewing autonomy as something that is collectively budgeted implies that (human) gains in cognitive autonomy may be large enough to lead to an overall increase in their autonomy, but they may also be more than offset by losses in epistemic or political autonomy. This view, we posit, redefines autonomy distribution vis-a-vis the introduction of machine actors: the supposed positive-sum game of autonomy distribution needs to be broadened in scope. As a consequence of this expanded scope, then, the game may be positive-sum in optimal cases, but may otherwise be zero-sum, or even negative-sum.

What possibilities, then, does the deployment of GenAIED open for redistributing autonomy between human and non-human actors? And what possible (optimal) distributions of the "autonomy budget" may balance the efficiency gains stemming from automation with the preservation of the involved people's capacity for independent reasoning and critical engagement?

To explore these questions, we propose using the autonomy budget as a tool to inform empirical studies of perceived and enacted autonomy, which then could guide educators, communities of practices of educators, or institutions, in deciding which policies to propose for the use of GenAIED in a given setting. Operationally, this can be tested with (1) SDT-based survey measures (e.g. autonomy-support and I-PLOC/E-PLOC; see also SDT-aligned AI measures (Li *et al*., 2025)), (2) classroom or A/B experiments manipulating GenAIED affordances (retrieval vs. summarization vs. solution-generation) and comparing learning/engagement outcomes, informed by recent meta-analytic evidence on ChatGPT's mixed effects on learning and higher-order thinking (Wang and Fan, 2025), (3) learning-analytics traces (e.g. help-seeking, revision patterns) as behavioral proxies for autonomy, (4) qualitative interviews/ethnography to capture perceived agency and contextual factors and (5) qualitative methods founded in SDT such as think-aloud protocols to find out to what extent students actually experience autonomy while using GenAIED (Meulenbroeks *et al*., 2025). As each educational setting is unique, in constant flux and as mutable as the actors it gathers around itself, we argue that those interested in autonomy budgeting should see it as an *ad-hoc* exercise – to be conducted for each educational setting.

To perform autonomy budgeting in a process of educational change, we suggest loosely following the process below, starting from a mapping of the actors that participate in an educational setting and culminating with the decision of a policy on the use of GenAIED.

(1) Identify, list and map the actors that participate in the educational setting (using, for example, ANT mapping techniques).

(2) Discuss the *ex ante* allocation of autonomy. For each actor, discuss what kind of autonomy they have, are perceived as having, or should have in a given educational setting (see Sec. 3). Quantification is not necessary; rather, staying at a descriptive, qualitative level is preferable.

(3) Discuss what process of change the educational setting is going through, and the role of the GenAI actor(s) in the setting. Whether a GenAIED is introduced, being afforded extra use, or decreased, investigate how the GenAIED's autonomy may trade off with human cognitive, epistemic and political autonomy.

(4) Discuss possible *ex-post* implications of these changes in autonomy distribution within the chosen educational setting. Possible questions to discuss include:

- What gains in efficiency could be achieved by increasing the cognitive or epistemic autonomy of the GenAI?
- May any human actors lose epistemic or political autonomy? What consequences could this reduction have?
- Would all (human) actors hold a meaningful share of the autonomy budget? Is there a risk that any actor could feel deprived of their basic psychological need for autonomy?

(5) Compare the *ex ante* and *ex-post* distribution of autonomy. What actors gained, and what actors lost autonomy (of all types)? Have human actors gained or preserved their autonomy?

(6) Based on the discussion above, explore what policies for the use of GenAI the analyzed educational context should adopt.

Crucially, the view above makes it explicit that the act of distributing autonomy – "budgeting" it – implies that introducing a new actor with a non-zero amount of resources increases the overall available resources, but some actors may end at a net loss.

The assumption that introducing GenAIED increases overall autonomy mirrors reflections on human cognitive enhancement (see, e.g. Clark (2025)). An appropriate deployment of GenAIED may indeed give rise to more autonomous people, but this is not to be taken for granted. While humans may enjoy using the GenAIED's autonomy to enhance their own cognitive autonomy, the epistemic autonomy that the GenAIED absorbs from them is not necessarily returned in full. The "freeing up" of cognitive autonomy may lead humans to feel increased self-efficacy (Liang *et al*., 2023), but what remains to be investigated is whether the involved humans feel they can keep their autonomy, seen in a holistic sense, i.e. as the power to decide for themselves (Floridi and Cowls, 2021). Too strong of a shift in the autonomy budget (especially for epistemic and political autonomy) assigned to human actors may eventually lead to people feeling a shift in their Perceived Locus of Control (or PLOC, see Sec. 2.1) from internal to external, leading to a reversal of the self-efficacy increase proposed by Liang *et al*. in (Liang *et al*., 2023).

In the next section, we offer some examples to illustrate our approach.

4.2 Some examples of trade-offs in autonomy

To explore autonomy redistribution when using GenAI in educational settings, we first need to reflect on the level of machine autonomy that is granted to GenAI. In the following examples, we will explore the same GenAI technology – general-purpose LLMs – operating at increasing levels of autonomy. Going back to the "levels of automation" scale proposed by Simmler *et al*. (Simmler and Frischknecht, 2021) and discussed in Sec. 3.3, we posit that, even when LLMs are used for "decision-offering" tasks, such as proposing spelling correction or alternative ways to rewrite a sentence, the systems are operating at an "executing fully automated" level, since the LLM yields its output (of possible choices) without human intervention. This creates the need to problematise the trade-off at a different level. In this section, we propose three case studies where the GenAIED supposedly works as an enhancer of human cognitive or epistemic autonomy, leaving open the question of whether they actually achieve this goal and the potential impacts on humans' epistemic and political autonomy.

The three examples we propose here show how, when a GenAI tool is deployed to enhance the (cognitive) autonomy of a student, this is never done without the GenAI demanding a modicum of delegation in terms of epistemic autonomy. This reflection also echoes Thimm *et al*. (2024), who summarize the current debate on how the transition from automated to autonomous technologies may reduce the space for human autonomy.

When deploying the same GenAIED tool as an increasingly epistemic autonomous agent, in other words, human actors may end up getting smaller shares of the autonomy budget, leading to the need to assess people's feelings of self-determination. The following short case studies exemplify how the same tool (a general-purpose LLM) can be deployed in three different scenarios, at increasingly higher levels of human cognitive and machine epistemic autonomies – and potentially, with effects of diminishing returns.

Case 1: Information retrieval

When people use LLMs to find or retrieve information (e.g. *in lieu* of an encyclopedia, or a search engine), they are primarily using LLMs as enhancers of their cognitive autonomy, with limited epistemic delegation. Committing information to memory – or retrieving information – are often seen as menial tasks (e.g. see the discussion of 21st century skills for learning, and the shift from knowledge acquisition to knowledge evaluation). LLMs, used in this role, then become the latest tools in a long line of cognitive aids that Western society used to supplement its people's memory. Writing, libraries, the printing press and Internet search engines all follow this pattern. These aids, however, have also been discussed as having an impact on people's ability to learn as far back as the time of Plato's *Phaedrus*: if LLMs start occupying a key role in how students retrieve or store information, students may unlearn how to commit information to memory themselves. The GenAI has effectively absorbed the cognitive autonomy that was previously budgeted for the students, and the GenAI has become not a tool, but a crutch.

Case 2: Outsourcing critical thinking

The summarization of large swaths of text, or queries to LLMs such as "give me the pros and cons of *X*" are forms of outsourcing critical thinking. While a precise definition of critical thinking is beyond the scope of this article, summarization queries are a form of offloading cognitive effort (building up on the case above) where, if the output is not evaluated critically, the LLM also implicitly acquires a meaningful portion of epistemic autonomy. The LLM used in this way also operates at a higher perceived level of automation, as people receive the tool's output without knowing how it operates the selection of what information to present and what to omit. Where the summarization task normally involves two steps of knowledge evaluation – and thus of epistemic autonomy – when building the summary itself and when evaluating its goodness, delegating summarization to an LLM offloads the first part. The LLM, in this sense, absorbed a part of the epistemic autonomy normally budgeted to students, potentially contributing to diminished critical thinking abilities (Gerlich, 2025; Lee *et al*., 2025).

Case 3: Outsourcing learning

A student in a basic programming course who uses an LLM to get the solution to a programming exercise is – effectively — outsourcing their learning entirely. Building up again on the previous case, here the student has completely relinquished their epistemic autonomy. Here, even verifying the validity of the output of the LLM is trivial: executing the code immediately evaluates the validity of the output of the LLM. The question, at this point, becomes as to whether the students have kept any cognitive autonomy and, if they did, what this autonomy is used for. If the students do not have ways to reinvest this autonomy back in the learning process itself, we argue, the machine actor effectively absorbed all of the autonomy budget dedicated to the student, leaving a student who is freed of their learning task – and is thus hardly a student anymore.

The three examples above prompt us to ask several questions informed by SDT: Is the students' basic psychological need of autonomy still satisfied? And what about their sense of competence and of relatedness? Have they perceived a shift in PLOC, or do they still feel in control? And what would be needed to prompt the students to question whether they are still in control or not?

All of these questions can be the basis of empirical studies. These studies may use quantitative (e.g. conducting surveys, using learning analytics), qualitative (e.g. using interviews, educational ethnography), or mixed methods and should be aimed at investigating how autonomy has been budgeted differently before and after the introduction of a GenAIED. Specific instruments for conducting such studies may include I-PLOC/E-PLOC scales and autonomy-support measures from SDT, task-specific competence/relatedness scales, pre/post performance and transfer tests, think-aloud protocols and log-based indicators of self-regulation (e.g. prompt evolution, source checking, revision depth).

5. Conclusions and directions for future research

The interplay between distributed cognition and autonomy highlights a paradox. Autonomy in learning is not merely the ability to act independently but involves understanding one's interdependence within a broader social and cognitive framework. Autonomous learners must balance leveraging technological and collective resources with maintaining the ability to critically assess and integrate these contributions. Education's challenge lies in cultivating learners who can navigate this duality: recognizing when to trust distributed systems – their own minds' extensions (Clark and Chalmers, 1998) — and when to engage their own independent cognitive faculties. GenAIED embodies and amplifies this tension, offering opportunities for distributed intelligence but also risking the erosion of personal cognitive agency if it is over-relied upon.

In this paper, we laid down a framework to conceptualize autonomy in education in the era of GenAI. It blends important aspects of a holistic understanding of education, pointing to biological/cognitive, distributive/technological and philosophical/political foundations.

These are framed through SDT and the three Basic Psychological Needs are identified: autonomy, competence and relatedness. While these are likely to remain fundamental even in the GenAIED era, we propose re-contextualizing them as part of socio-technical systems in which GenAI tools are more than just instruments, but proper actors.

This theoretical framework opens up to empirical studies that investigate *how* autonomy changes when GenAIED becomes an educational actor. These studies will need to involve multiple disciplines (e.g. psychology, education and computing), and educational contexts (e.g. problem-solving, essay writing, literature search and coding). These studies should be used to better inform the adoption of policies for the use of GenAIED (Jin *et al*., 2025) and the renewal of well-established learning goals such as those identified in Bloom's taxonomy. We also acknowledge, in the ways of limitations, that this work is theoretical and context-sensitive; we have not yet validated the autonomy budget empirically and effects may vary across age groups, subjects and institutional constraints. Future research should test the framework across diverse settings, compare alternative GenAIED designs and refine measurement for cognitive, epistemic and political autonomy. Policy-wise, recent reviews highlight heterogeneous institutional guidance and emerging best practices (Luo (Jess), 2024).

In the current educational landscape, then, our call is for educators to approach technological integration as curators of cognitive ecosystems and as wise allocators of the autonomy resources of their educational context. Only a nuanced view, we argue, can harness the potential of GenAI while preventing the erosion of foundational learning processes.

Notes

1. Curiously enough, the layered interplay between the levels of autonomy is reminiscent of concepts in complex systems and emergence theory, where local interactions and constraints can lead to global patterns that both reflect and transcend the underlying rules (see, e.g. (Kauffman, 1993),). In both biological evolution and democratic governance, short-term sacrifices in lower-level determinism pave the way for long-term gains in freedom, adaptability and resilience.

2. However Yao *et al*. (2022), show that while LLM-based agents can exhibit open-ended behavior and task adaptability, their "learning" from interactions is mostly driven by system-level design choices rather than ongoing parameter updates within the LLM itself.

References

Bächle, T.C. (2023), "The age of machine autonomy?".

Balzan, F., Munarini, M. and Angeli, L. (2024), "Who pilots the copilots?", in *Lecture Notes in Computer Science*, Springer Nature Switzerland, Cham, pp. 448-456.

Beer, J.M., Fisk, A.D. and Rogers, W.A. (2014), "Toward a framework for levels of robot autonomy in human-robot interaction", *Journal of Human-Robot Interaction*, Vol. 3 No. 2, p. 74, doi: 10.5898/jhri.3.2.beer.

Bisconti, P., McIntyre, A. and Russo, F. (2024), "Synthetic socio-technical systems: Poiêsis as meaning making", *Philosophy and Technology*, Vol. 37 No. 3, 94, doi: 10.1007/s13347-024-00778-0.

Bourdieu, P. and Passeron, J.C. (2013), *Reproduction in Education, Society and Culture*, Published in Association with Theory, Culture and Society, 2nd ed., SAGE Publications, London, Oct 2013.

Cassinadri, G. (2024), "ChatGPT and the technology-education tension: applying contextual virtue epistemology to a cognitive artifact", *Philosophy and Technology*, Vol. 37 No. 1, 14, doi: 10.1007/s13347-024-00701-7.

Cerasoli, C.P., Nicklin, J.M. and Ford, M.T. (2014), "Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis", *Psychological Bulletin*, Vol. 140 No. 4, pp. 980-1008, doi: 10.1037/a0035661.

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E.L., Van der Kaap-Deeder, J., Duriez, B., Lens, W., Matos, L., Mouratidis, A., Ryan, R.M., Sheldon, K.M., Soenens, B., Van Petegem, S. and Verstuyf, J. (2015), "Basic psychological need satisfaction, need frustration, and need strength across four cultures", *Motivation and Emotion*, Vol. 39 No. 2, pp. 216-236, doi: 10.1007/s11031-014-9450-1.

Chiu, T.K.F. (2024), "A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: a case of ChatGPT", *Educational Technology Research and Development*, Vol. 72 No. 4, pp. 2401-2416, doi: 10.1007/s11423-024-10366-w.

Clark, A. (2025), "Extending minds with generative AI", *Nature Communications*, Vol. 16 No. 1, p. 4627, doi: 10.1038/s41467-025-59906-9.

Clark, A. and Chalmers, D. (1998), "The extended mind", *Analysis*, Vol. 58 No. 1, pp. 7-19, doi: 10.1093/analys/58.1.7.

Damasio, A. (2000), *The Feeling of what Happens*, Vintage, London.

Deci, E.L. and Ryan, R.M. (2013), "Intrinsic motivation and self-determination in human behavior", in *Perspectives in Social Psychology*, Springer, New York, NY.

Dewey, J. (2022), *Democracy and Education*, Legare Street Press, Charleston, SC.

Doshi-Velez, F. and Kim, B. (2017), "Towards a rigorous science of interpretable machine learning", arXiv [stat.ML].

Dragos, C. (2021), "Epistemic autonomy and group knowledge", *Synthese*, Vol. 198 No. 7, pp. 6259-6279, doi: 10.1007/s11229-019-02461-w.

Eccles, J.S. and Wigfield, A. (2020), "From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation", *Contemporary Educational Psychology*, Vol. 61, 101859, doi: 10.1016/j.cedpsych.2020.101859.

Floridi, L. and Cowls, J. (2021), "A unified framework of five principles for AI in society", in *Philosophical Studies Series*, Springer International Publishing, Cham, pp. 5-17.

Formosa, P. (2021), "Robot autonomy vs. human autonomy: social robots, artificial intelligence (AI), and the nature of autonomy", *Minds and Machines*, Vol. 31 No. 4, pp. 595-616, doi: 10.1007/s11023-021-09579-2.

Gerlich, M. (2025), "AI tools in society: impacts on cognitive offloading and the future of critical thinking", *Societies*, Vol. 15 No. 1, p. 6, doi: 10.3390/soc15010006.

Giddens, A. (1992), *The Giddens: Constitution Society (Cloth): Outline of the Theory of Structuration*, University of California Press, Berkeley, CA.

Gould, S.J. (1990), *Ontogeny and Phylogeny*, Belknap Press, London.

Hume, D. (2000), "A treatise of human nature", in *Oxford Philosophical Texts*, Oxford University Press, London.

Hutchins, E. (2000), "Distributed cognition", *International Encyclopedia of the Social And Behavioral Sciences*, Vol. 138 No. 1, pp. 1-10.

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

Kaber, D.B. (2018), "Issues in human–automation interaction modeling: presumptive aspects of frameworks of types and levels of automation", *Journal of Cognitive Engineering and Decision Making*, Vol. 12 No. 1, pp. 7-24, doi: 10.1177/1555343417737203.

Kant, I. (2019), "Groundwork of the metaphysics of morals", Independently Published.

Kauffman, S.A. (1993), *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York, NY.

Laland, K.N. (2004), "Extending the extended phenotype", *Biology and Philosophy*, Vol. 19 No. 3, pp. 313-325, doi: 10.1023/b:biph.0000036113.38737.d8.

Laland, K.N. (2017), *Darwin's Unfinished Symphony: How Culture Made the Human Mind*, Princeton University Press, Princeton, NJ.

Latour, B. (2007), *Reassembling the Social: An Introduction to Actor-Network-Theory*, OUP Oxford, Oxford.

Lee, H.P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. and Wilson, N. (2025), “The impact of generative ai on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers”, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1-22.

Li, J., King, R.B., Chai, C.S., Zhai, X. and Lee, V.W.Y. (2025), “The AI motivation scale (AIMS): a self-determination theory perspective”, *Journal of Research on Technology in Education*, pp. 1-22, doi: 10.1080/15391523.2025.2478424.

Liang, J., Wang, L., Luo, J., Yan, Y. and Fan, C. (2023), “The relationship between student interaction with generative artificial intelligence and learning achievement: serial mediating roles of self-efficacy and cognitive engagement”, *Frontiers in Psychology*, Vol. 14, 1285392, doi: 10.3389/fpsyg.2023.1285392.

Luo (Jess), J. (2024), “A critical review of GenAI policies in higher education assessment: a call to reconsider the ‘originality’ of students’ work”, *Assessment and Evaluation in Higher Education*, Vol. 49 No. 5, pp. 651-664, doi: 10.1080/02602938.2024.2309963.

Maturana Rumesin, H. and Varela, F.J. (1991), “Autopoiesis and cognition”, in *Boston Studies in the Philosophy and History of Science*, Kluwer Academic, Dordrecht.

Mazeika, M., Yin, X., Tamirisa, R., Lim, J., Lee, B.W., Ren, R., Phan, L., Mu, N., Khoja, A., Zhang, O. and Hendrycks, D. (2025), “Utility engineering: analyzing and controlling emergent value systems in AIs”, arXiv [cs.LG].

Meulenbroeks, R., van Rijn, R. and Reijerkerk, M. (2023), “Fostering secondary school science students’ intrinsic motivation by inquiry-based learning”, *Research in Science Education*, Vol. 54 No. 3, pp. 339-358.

Meulenbroeks, R., Del Moral-Sánchez, I., Vos, W. and Russo, F. (2025), “‘AI is something that needs to be checked’: university science students experiences with generative AI in academic writing assignments”, *AI and Education*, In press.

Mill, J.S. (2022), *On Liberty*, Salzwasser-Verlag, Bremen.

Reeve, J. and Cheon, S.H. (2021), “Autonomy-supportive teaching: its malleability, benefits, and potential to improve educational practice”, *Educational Psychology*, Vol. 56 No. 1, pp. 54-77, doi: 10.1080/00461520.2020.1862657.

Reid, T. (2011), *Cambridge Library Collection – Philosophy: An Inquiry into the Human Mind, on the Principles of Common Sense*, Cambridge University Press, Cambridge.

Rogers, C.R. (1965), “Some questions and challenges facing a humanistic psychology”, *Journal of Humanistic Psychology*, Vol. 5 No. 1, pp. 1-5, doi: 10.1177/002216786500500101.

Ryan, R.M. and Deci, E.L. (2000), “Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being”, *American Psychologist*, Vol. 55 No. 1, pp. 68-78, doi: 10.1037//0003-066x.55.1.68.

Ryan, R.M. and Deci, E.L. (2018), *Self-determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*, Vol. 756, Guilford Publications, New York, NY.

Ryan, R.M. and Deci, E.L. (2020), “Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions”, *Contemporary Educational Psychology*, Vol. 61, 101860, doi: 10.1016/j.cedpsych.2020.101860.

Simmler, M. and Frischknecht, R. (2021), “A taxonomy of human–machine collaboration: capturing automation and technical autonomy”, *AI and Society*, Vol. 36 No. 1, pp. 239-250, doi: 10.1007/s00146-020-01004-z.

Thimm, C., Gramelsberger, G., Mayer, M. and Piller, F. (2024), “From automation to autonomy: human machine relations in the age of artificial intelligence”, *Human-Machine Communication*, Vol. 9, pp. 7-24, doi: 10.30658/hmc.9.1.

Tomasello, M. (2009), *The Cultural Origins Of Human Cognition*, Harvard University Press, Cambridge, MA.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K. and Lens, W. (2009), "Motivational profiles from a self-determination perspective: the quality of motivation matters", *Journal of Educational Psychology*, Vol. 101 No. 3, pp. 671-688, doi: 10.1037/a0015083.

Wang, J. and Fan, W. (2025), "The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. Humanit", *Soc. Sci. Commun.*, Vol. 12 No. 1, pp. 1-21.

Xia, Q., Chiu, T.K.F., Lee, M., Sanusi, I.T., Dai, Y. and Chai, C.S. (2022), "A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education", *Computer Education*, Vol. 189, 104582, doi: 10.1016/j.compedu.2022.104582.

Xia, Q., Chiu, T.K.F., Chai, C.S. and Xie, K. (2023), "The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot", *British Journal of Educational Technology*, Vol. 54 No. 4, pp. 967-986, doi: 10.1111/bjet.13305.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. and Cao, Y. (2022), "ReAct: synergizing reasoning and acting in language models", arXiv [cs.CL].

Yusuf, A., Pervin, N. and Román-González, M. (2024), "Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives", *International Journal of Educational Technology in Higher Education*, Vol. 21 No. 1, 21, doi: 10.1186/s41239-024-00453-6.

Corresponding author
Lorenzo Angeli can be contacted at: lorenzo.angeli@unitn.it