

Ethical considerations of integrating new educational technologies into the classroom – Technological ethics construction based on the protection of "thinking sovereignty"

Chenyu Lv

Sichuan Normal University, Chengdu, China

1448301863@qq.com

Abstract. With the large-scale integration of new educational technologies such as generative artificial intelligence, adaptive learning systems, and affective computing into classrooms, the corresponding ethical discussions have primarily focused on privacy, fairness, and interpersonal relationships. This paper raises a deeper ethical issue that has not yet received sufficient attention – the implicit relinquishment of students' "thinking sovereignty". Drawing on the theory of technological intentionality, the article reveals how educational technology, under the legitimate goals of "precision", "efficiency", and "adaptability", has shifted from being an auxiliary tool to a cognitive substitute, thereby eroding students' metacognitive abilities, creative will, and independent thinking rights. To address this crisis, the article constructs three ethical principles – "cognitive friction retention", "sovereignty transparency", and "developmental autonomy" – in hopes of preserving an irrevocable thinking territory for education in the technological era.

Keywords: Educational Technology, Classroom Ethics, Reflective Sovereignty, Cognitive Substitution, Technological Intentionality.

1. Introduction

New educational technologies are reshaping classrooms at an unprecedented pace. From intelligent grading and adaptive question generation, to generative AI-assisted writing, VR immersion, and affective computing, "precision," "personalization," and "efficiency" have become the promises technology bestows upon education. Schools eagerly introduce smart classrooms, and policies promote digital transformation. Beneath this technological fervour, however, ethical concerns persist. Current discussions mostly focus on privacy leakage, algorithmic bias, digital divides, or the thinning of emotional teacher-student communication. These are real dilemmas, but they primarily view students as objects to be protected – from data abuse, discrimination, or isolation. This article advances the discussion by raising a more subtle and fundamental ethical question: when educational technology transforms from an assistant into a substitute for thinking, students unknowingly relinquish a more basic right – the right to think autonomously. This is no longer about "whether technology is safe" or "whether algorithms are fair," but an ethical crisis of "whether humans are still thinking for themselves." By shifting the ethical lens from external protection to internal cognitive autonomy, this paper draws an insurmountable red line for integrating new technologies into classrooms.

2. Traditional ethical challenges posed by new educational technologies

Data privacy and algorithmic discrimination, digital divides and equity, and teacher-student alienation constitute the main ethical issues in current educational technology discourse. However, these examinations focus mostly on external infringements and have not touched upon a more subtle crisis – technology replacing students' own thinking processes.

2.1. Data privacy and algorithmic discrimination

Smart classrooms rely on real-time collection of behaviour, performance, and physiological data. Face recognition, eye tracking, and clickstream data are all transformed into assets. Privacy

infringement has become routine, and algorithms trained on historical data may amplify gender, racial, or urban-rural stereotypes, forming systematic algorithmic discrimination. Technology that should eliminate inequality may instead solidify class differences.

2.2. Digital divide and educational equity

High-end adaptive systems, VR equipment, and high-speed networks are not affordable for all schools and families. Technological integration may inadvertently widen the “technological gap” between schools, regions, and families, making equity reappear in a digital guise. Classroom technological advantages easily translate into academic advancement advantages.

2.3. Alienation and emotional estrangement in teacher-student relationships

When intelligent systems provide instant feedback, diagnosis, and even emotional encouragement, teachers’ humanistic care appears replaceable. The intersubjectivity between teachers and students faces challenges; empathy, role modelling, and moral influence gradually give way to screen-based human-computer interaction, making emotional education harder to implement.

These dilemmas have triggered widespread alarm, but they mostly remain within the framework of “technology as an external force that infringes upon existing educational ethical orders.” The hidden aspect is that technology has begun to shape students’ internal cognitive processes, touching education’s core ethical bottom line – human independent thinking.

3. From the perspective of innovation ethics: The crisis of “cognitive sovereignty” – cognitive substitution and technological intentionality

The impairment of sovereignty is not abstract theoretical deduction; it has observable cognitive manifestations. Drawing on cognitive and educational psychology, this article classifies it into three dimensions: excessive cognitive disengagement, metacognitive decline, and suppressed creativity.

3.1. Intentionality of educational technology: sliding from “assistance” to “replacement”

“Cognitive offloading” refers to shifting cognitive tasks to external tools to reduce internal load – using notes for memory or calculators for arithmetic. Moderate offloading is normal and efficient, releasing resources for higher-level thinking. Risko and Gilbert (2016) [7] systematically examined its types and mechanisms, noting that humans readily rely on external tools to save internal resources.

However, new educational technologies have qualitatively changed the degree of offloading. Traditional tools offloaded lower-level processes – memory and calculation; new technologies offload higher-level processes – reasoning, analysis, synthesis, and creative connections. When students use generative AI to complete essays, the offloaded processes are not spelling and grammar but idea condensation, argument organisation, and evidence development. When they rely on adaptive systems for step-by-step maths hints, the offloaded processes are not calculation itself but the metacognitive decision of “where to start with an unfamiliar problem.” When offloading extends to higher-level cognition, the brain becomes not a training ground for thinking but an input device for instructions – this is where thinking sovereignty is fundamentally undermined. Furthermore, this shift disrupts the formation of durable cognitive schemas. According to cognitive load theory, deep learning requires active manipulation of information within working memory to construct long-term mental models. When AI performs the organising and synthesising work, the learner’s working memory is never sufficiently challenged to build these robust schemas; consequently, when the AI is removed, the learner lacks not only the immediate answer but also the internalised mental structure needed to generate one. In essence, the technology does not just save effort – it prevents the very cognitive labour that builds competence.

Gerlich (2025) [2] found that students who frequently use AI tools scored significantly lower in critical thinking tests than infrequent users, with cognitive offloading mediating this negative

correlation – the more one offloads thinking to AI, the weaker the critical thinking ability. This provides preliminary empirical evidence for damaged thinking sovereignty.

3.2. Reflecting on the ethical connotation of sovereignty

Metacognition – “thinking about thinking” – is the ability to monitor and regulate one’s own cognitive processes. It develops through continuous reflection on cognitive experiences: making mistakes, identifying them, analysing causes, and adjusting strategies. This cycle drives metacognitive growth.

Yet a core design philosophy of intelligent educational technology is precisely “reducing errors and smoothing the path”. Adaptive systems predict error nodes and intervene in advance; intelligent tutors provide prompts the moment students get stuck, avoiding prolonged “ineffective struggle”. These designs improve efficiency and protect confidence, but their potential cost is profound: students encounter fewer cognitive conflicts and fewer “metacognitive moments” of realising “I was wrong”. Over time, they lose sensitivity to and control over their own cognition. Once the system is removed – as in unaided exams – they lack metacognitive ability: they do not know where to start, how to check, or where they are confused.

Park (2023) [5] sharply observed: “When artificial intelligence does our thinking for us, we stop thinking for ourselves – and we will gradually forget how to think for ourselves.” Wang et al. (2023) [9] report similarly warns that intelligent tools may “suppress students’ ability to reflect on their own learning process,” weakening metacognitive development.

Some researchers argue that well-designed AI scaffolding can enhance metacognition by prompting reflection on reasoning steps, acting as a “cognitive mirror” rather than a substitute (Molenaar, 2022) [4]. This is a valid concern. However, current AI systems rarely incorporate gradual withdrawal; instead, they provide continuous, invisible assistance without requiring active self-monitoring. The absence of a clear “exit mechanism” means students seldom experience unassisted problem-solving, thus losing opportunities to develop internal self-regulation. Sovereignty remains compromised unless scaffolding explicitly fades and students are made aware of what is being taken over – which most systems fail to do.

3.3. Consider the triple manifestations of sovereignty impairment

Creativity – generating novel and valuable ideas – requires deviating from conventional paths, attempting unusual combinations, and tolerating ambiguity. Yet mainstream generative AI models operate on “high probability” – predicting the most likely next word, the most conventional expression, and the safest output. This naturally inclines AI to provide average, smooth, risk-averse content. When students habitually accept AI-generated content, their thinking is drawn onto the smooth track of “thinking like most people”.

The concealment of creativity goes beyond AI replacing students’ creation; it affects their aesthetic standards of “what constitutes a good thinking result.” Repeated exposure to AI-generated “standard answers” and “model essays” may cause students to internalise those styles as the “correct form”. Consequently, unconventional, adventurous, and rough-but-original ideas are self-censored because they fail the “smooth AI-style standard”. This suppression is particularly insidious because creativity thrives on what psychologists call “divergent thinking” – the willingness to generate multiple unusual solutions. AI’s probabilistic nature inherently penalises divergence; it smooths away outliers and reinforces the statistical mean. When students internalise this smoothed standard, they begin to self-censor their own divergent impulses before they even reach conscious articulation, thereby extinguishing the very spark of originality. Creativity is not externally suppressed but silently obscured by an internalised conformity to machine-generated norms.

Excessive cognitive deactivation deprives thinking of “practice rights”; metacognitive shrinkage deprives it of “reflection rights”; and creativity concealment deprives it of “adventure rights”. The damage to thinking sovereignty is thus three-dimensional and systematic – a product of technological logic colluding with educational utilitarianism, not an incidental side-effect.

4. Typical case: Erosion of thinking sovereignty in generative AI and adaptive learning systems

The theoretical framework needs testing in specific contexts. This section selects two typical applications – generative AI-assisted writing and adaptive learning systems – to illustrate sovereignty erosion mechanisms.

4.1. Case One: Generative AI in the composition class

A typical scenario: a teacher assigns an argumentative essay; students use AI, enter a prompt, and receive a well-structured article within seconds. They adjust some wording and submit it, receiving a good grade. From a sovereignty perspective, the problem exceeds academic dishonesty. Throughout the process, students' thinking participation is minimised: viewpoint condensation, logical structure, and evidence organisation are all AI-replaced. Students merely select from AI output and perform surface polishing. The autonomy to initiate thinking is eliminated by AI's immediate response; the choice of thinking path is replaced by AI's default; the tolerance of cognitive setbacks is bypassed; and ownership of the result becomes ambiguous – to what extent is this “the student's own”? Crucially, the AI's immediate response eliminates the necessary “incubation period” – the time during which ideas gestate, connect unconsciously, and mature. Cognitive psychology has long established that creative problem-solving benefits from periods of unconscious associative processing; by providing a polished product within seconds, the AI short-circuits this essential cognitive phase. What students lose is not merely the exercise of writing, but the deeply human process of wrestling with half-formed thoughts, abandoning weak arguments, and discovering unexpected connections through sustained mental effort. If repeated daily, students may gradually “forget” how to independently construct an argument from scratch, creating a structural deficit in cognitive ability.

4.2. Case Two: Adaptive learning system in mathematics class

A typical scenario: a mathematics class uses an adaptive system. It pushes a moderately difficult question based on prior performance. The student hesitates; the system detects a timeout and pops up a hint. The student follows the hint, solves it, and the system marks “mastered” and proceeds. Accuracy stays high, and the system reports “good learning state”. Yet from a sovereignty view, the student never truly struggles. That question could have been solved with more time and trial-and-error, but the system intervenes at the moment of near-confusion – a “cognitive greenhouse”: constant temperature, no adversity; seedlings appear vigorous but never develop resilience. Educational psychology has long recognised that errors are not failures but essential data points for learning. The adaptive system's pre-emptive hints prevent the student from generating and testing their own hypotheses. When a student tries a wrong path and corrects themselves, they strengthen their understanding of why the correct path works; this self-correction process deepens mental representations far more than passive reception of a hint. By eliminating this trial-and-error cycle, the system produces high accuracy but fragile knowledge – knowledge that fails when context changes or hints are withdrawn. The core of mathematical thinking lies in facing unfamiliar, unguided problems and exploring solution paths independently – not in correctly solving after hints. The adaptive system optimises efficiency indicators (accuracy, speed, coverage) at the expense of depth indicators (courage to face the unknown, ability to find structure, resilience). Cognitive friction is minimised, yet it is precisely this friction that is the necessary load for thinking to grow.

Across both cases, the underlying logic is consistent: efficiency prevails over development; “assistance” masks “substitution”; short-term gains obscure long-term costs. The current evaluation system, with its emphasis on standardised test scores, rewards any technology that boosts performance while its ethical costs are easily overlooked. Protecting sovereignty is thus not merely a technical ethics issue but also a question of educational evaluation philosophy.

5. Construction of ethical principles for educational technology based on “thinking sovereignty”

5.1. Principle of retaining cognitive friction

Educational technology design should not merely pursue smoothness, zero frustration, and immediate satisfaction. Instead, it should deliberately retain moderate “cognitive friction” – delaying answers, hiding some paths, providing incomplete information, and encouraging trial-and-error before feedback – to ensure substantial independent thinking during technology use.

From Piaget’s (1970) [6] “cognitive conflict” theory to contemporary “productive failure” research (Kapur, 2008) [3], educational theory consistently affirms that true learning often occurs when cognition encounters resistance and must reorganise itself. If technology makes learning too smooth, it undermines the very conditions for learning.

At the design level, this principle translates into specific strategies: photo-search should not directly present the complete answer but first push hints, relevant knowledge points, and similar examples, with a mandatory “thinking time”; adaptive systems can include open-ended challenging questions simulating real-world uncertainty; and systems can include error-analysis functions guiding students to review their mistake processes. At the classroom level, teachers should establish “moderation” – for instance, requiring that first drafts and first problem-solving attempts be done independently, positioning technology as an auxiliary improvement tool after independent effort.

5.2. Principle of sovereign transparency

Educational technology must clearly and understandably disclose to students: which cognitive processes the system is taking over, the logic behind that takeover, and that students have the option to turn it off and return to independent thinking. Cognitive takeover should not be silent or default; it must be transparent and require active confirmation.

This principle draws on the “informed consent” principle in medical ethics (Beauchamp & Childress, 2019) [1]. If transferring thinking sovereignty is a “cognitive major intervention”, students have the right to know what they are giving up, to whom, and with what consequences. The common problem is that students have no idea why an algorithm recommends this question over that one, why AI generates this text rather than another, or what long-term impact these interventions have.

At the design level, adaptive systems can attach brief explanations: “The system recommends this proof question because your error rate in geometry proofs has been high.” Generative AI tools should clearly label AI-generated content and provide a “turn off AI assistance” option. A “thinking sovereignty dashboard” could visually show the proportion of cognitive processes completed independently versus assisted. At the classroom level, teachers should cultivate critical awareness by guiding discussions like “When you use AI, what is your brain doing? What aspects is AI replacing?” – helping students move from unconscious dependence to conscious coexistence.

5.3. Principle of developmental autonomy

Technological intervention in cognitive sovereignty must follow the developmental logic of “gradual withdrawal,” with the ultimate goal not permanent replacement but enhanced capacity for independent thought. The system must have a clear “exit mechanism” – gradually removing scaffolding as students’ abilities grow, until sovereignty is fully restored.

This principle stems directly from Vygotsky’s (1978) [8] “scaffolding” theory, whose core characteristics are “temporary” and “gradual withdrawal” – provided when needed and removed as ability improves. If scaffolding becomes permanent, it ceases to be scaffolding and becomes a constraint. Applied to educational technology: if a system creates permanent cognitive dependence, it is ethically flawed because it violates education’s fundamental purpose – fostering independence.

At the design level, adaptive systems should adjust prompts not solely based on “current accuracy” but also on “capability growth” – as students practise and error rates decline, the system should reduce prompts and increase open-ended questions. Optimisation objectives should include “independence”

metrics – tasks completed without prompts, first-attempt accuracy, and self-correction rates – alongside learning speed and accuracy. The system should also schedule regular “unassisted assessments”. At the classroom level, teachers should apply a gradual withdrawal strategy – more assistance early in the semester, progressively defining “technology-free zones” and “prohibited periods”. At the evaluation level, critical assessments can include a dedicated “unassisted evaluation” component examining students’ unaided thinking quality, giving it appropriate weight.

6. Conclusion and Discussion

The integration of new educational technologies is irreversible. This article does not aim to halt this trend but to install an ethical brake on its momentum – when we rejoice in enhanced efficiency and personalised experiences, we must ask: what rights are students trading away for these conveniences?

The answer is: students are trading their “thinking sovereignty” – autonomous control over their cognitive processes – for technological efficiency. This exchange is silent, gradual, and cloaked as “assistance,” making it far more covert than privacy violations or algorithmic discrimination. Once sovereignty is systematically ceded, education’s foundation is undermined: what is cultivated is no longer independent thinkers but instruction-following agents reliant on technological systems.

To address this crisis, this paper introduces the analytical concept of “thinking sovereignty” and constructs an ethical framework from three dimensions: cognitive friction retention, sovereignty transparency, and developmental autonomy. The core spirit can be distilled into a single question: Is every technological assistance merely replacing student thinking, or is it genuinely helping students think more independently? Only by adhering to the latter can educational technology remain on the ethical track.

Certainly, this discussion has limitations. Sovereignty protection must be examined within broader political economy and critical pedagogy perspectives, analysing how power dynamics among technology, capital, and education shape the ecology of “thought substitution.” Moreover, the three principles remain at the conceptual stage, requiring further operationalisation through design-based research and collaboration with developers to test their impact on independent thinking in real classrooms. Future research must move beyond conceptual critique to empirical validation, developing measurable indicators of cognitive friction and sovereignty retention, and testing these principles through longitudinal classroom interventions that track both short-term performance and long-term cognitive development. Such empirical work is essential to translate ethical principles into actionable design standards.

Finally, safeguarding the right to think is the shared duty of educators, technologists, policymakers, parents, and students themselves. Education’s ultimate goal has never been to make people more efficient at completing tasks, but to cultivate individuals capable of independent thinking, autonomous judgment, and free creation. In an era of pervasive technology, this ancient ideal is more crucial than ever to be remembered and protected.

References

- [1] Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.
- [2] Gerlich, M. (2025). AI tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- [3] Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.
- [4] Molenaar, I. (2022). The concept of scaffolding in the age of artificial intelligence. *Educational Psychology Review*, 34(4), 2153–2172.
- [5] Park, S. (2023). When AI thinks for us: the erosion of human cognition. *Skeptic*, 28(2), 34–41.
- [6] Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.

- [7] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688.
- [8] Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Harvard University Press.
- [9] Wang, T., Zheng, J., Tan, C., & Lajoie, S. P. (2023). Computer-based scaffoldings influence students' metacognitive monitoring and problem-solving efficiency in an intelligent tutoring system. *Journal of Computer Assisted Learning*, 39(5), 1652–1665. <https://doi.org/10.1111/jcal.12824>