FROM CREATIVE TOOL TO COGNITIVE PROSTHESIS: GENERATIVE AI, DIGITAL STORYTELLING, AND ADOLESCENT DEVELOPMENT

Davide Milasi
Università Mediterranea di Reggio Calabria
mlsdvd01s02h224v@studenti.unirc.it

Rossella Marzullo
Università Mediterranea di Reggio Calabria
rossella.marzullo@unirc.it

Double Blind Peer Review

Citation
Milasi, D., & Marzullo, R. (2026). From Creative Tool to Cognitive Prosthesis: Generative AI, Digital Storytelling, and Adolescent Development. *Italian Journal of Health Education, Sport and Inclusive Didactics*, *10*(4). https://doi.org/10.32043/gsd.v10i4.1883

Doi: https://doi.org/10.32043/gsd.v10i4.1883

Copyright notice:
© 2026 The Author(s). This is an open access, peer-reviewed article published by Cares S.r.l. – REN Press and distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

renpress.it/index.php/gsdj
ISSN: 3103-5744

ABSTRACT
The integration of generative AI into digital storytelling offers new opportunities for exploring adolescent identity (Talarico et al., 2022). However, the shift from creative tool to "cognitive prosthesis" entails pedagogical risks. When used mindlessly, technology may weaken critical thinking: the algorithm replaces intellectual effort, fostering passivity and inhibiting introspection (Panciroli & Rivoltella, 2023). Without genuine reflection on personal experience, affective education is compromised. The lack of emotional literacy hinders empathy, creating an identity void that may lead to deviance and social distress (Floridi et al., 2018). This paper argues that AI should be used consciously, not as a shortcut, but as a technical support for a reflective creative process, essential to prevent ethical risks and promote authentic self-awareness (Ranieri, 2022).

KEYWORDS
Artificial Intelligence; Digital Storytelling; Adolescent Identity; Cognitive Prosthesis; Affective Education.

Received 2026-07-09
Accepted 2026-09-26
Published 2026-10-30

1. Introduction: the paradox of the intelligent tool

We live in an age in which the boundaries between technical aid and cognitive substitute are becoming increasingly blurred. Generative artificial intelligence, in its most widespread forms-from conversational chatbots such as ChatGPT to generators of images and creative text-presents itself to the user as a capable, adaptive and extraordinarily efficient interlocutor. This appearance of intelligence, combined with ease of use, generates an irresistible appeal, particularly for younger generations, who find themselves operating in digital environments with a naturalness that often precedes the critical awareness needed to govern them.

What this article intends to explore does not concern AI in itself, nor its usefulness-which is beyond dispute-but rather the pedagogical conditions that make its use educationally legitimate and formatively valuable. The central thesis is that AI can be used in an educationally sound way only when the person using it already knows what they want to achieve: when they possess a cognitive direction, an elaborated intention, a mental project that precedes the interaction with the system. Otherwise, when the user delegates to AI not only the execution but also the formulation of the task, a functional substitution occurs that does not empower thought but replaces it.

The specific context in which this dynamic takes on particularly urgent educational significance is that of digital storytelling in adolescence. Digital narrative has historically been valued as an identity-building practice, as a process through which the subject tells their own story, reflects, and constructs themselves. The introduction of generative AI into this narrative space opens up scenarios of great interest but also of great risk. While AI can, on the one hand, amplify the adolescent's expressive capacities, on the other it can deprive them of the effortful-but indispensable-work of introspection that narration requires.

2. Digital storytelling and the construction of adolescent identity

2.1 Writing the Self: mechanisms and meanings

According to a long theoretical tradition running from Ricoeur (1990) to Bruner (1990), narration is not simply a means of communication but a fundamental way in which the subject constructs and inhabits their own Self. To narrate is not to describe what one already is: it is to become what one recounts. This principle takes on particular relevance in adolescence, a phase in which the subject is structurally engaged in what Erikson (1968) calls the developmental task of identity. The adolescent does not yet possess a stable and consolidated identity: they search for it, experiment with it, narrate it, and call it back into question. Digital storytelling-understood as the production of personal narratives through digital media: video, podcasts, blogs, multimedia posts-has introduced a new dimension into this process. The possibility of using images, sounds, video editing and graphic effects has made identity narration richer, more public and, potentially, more elaborate (Lambert, 2013; Hull & Katz, 2006). The process of creating a digital story requires the subject to make choices that are not only technical but profoundly autobiographical: what to include, what to exclude, how to order events, which narrative voice to adopt, which emotion to foreground. Every choice is an act of self-definition. Recent studies have shown how digital narrative construction in educational contexts fosters the development of metacognitive skills and identity reflexivity (Talarico et al., 2022).

2.2 Narrative effort

A fundamental aspect that the literature sometimes tends to underestimate is the formative role of narrative difficulty. Writing about oneself is difficult. Choosing the right words to describe an emotion is both a cognitive and an affective act: it presupposes the ability to recognise the emotion, to name it, to situate it within a

story, and to understand its causes and consequences. This operation-which we might call the narrative elaboration of lived experience-is neither automatic nor instinctive. It requires time, attention, reflection and, often, the guidance of a competent adult.

From a neuroscientific standpoint, Siegel (2012) has documented how autobiographical narration is closely linked to the integration of prefrontal neural networks with limbic structures. Telling one's own story is not merely a linguistic activity: it is a process of neurobiological integration that fosters emotional regulation, coherence of the Self and the construction of meaningful relationships. From this perspective, obstructing or shortening the narrative process-for instance by delegating it to a machine-is not simply an intellectual shortcut: it is a deprivation of the brain itself of an opportunity for development.

Blakemore and Frith (2005) have shown that the adolescent brain is still developing in its executive functions, particularly in the areas of planning, impulse control and the evaluation of consequences. This makes adolescents particularly plastic-that is, capable of learning and developing in extraordinary ways-but also particularly vulnerable to the cognitive habits that take hold during this period. The modes of thought-active, reflective, critical versus passive, delegated, superficial-that become consolidated during adolescence tend to structure the cognitive competences of the adult.

2.3 Identity reflected in AI: between awareness and delegation

Digital storytelling differs from traditional forms of autobiographical narration also in its public or semi-public dimension. The adolescent who posts a video on YouTube, shares a story on Instagram, or produces a podcast for their class is not merely writing a diary, but negotiating their own social image, managing their own reputation and calibrating the distance between the intimate and public spheres. This negotiation constitutes a form of learning in itself, as it allows the subject to

distinguish between what is their own and what is shareable, between what is authentic and what is performative. In this context, the use of AI introduces a further complication, since the production of identity narrative may be mediated by generative systems that rework texts and discursive models that are foreign to the subject's lived experience. This raises a significant issue on both the pedagogical and identity-related plane, insofar as the voice presented to the public may be partially or entirely misaligned with the authentic elaboration of the Self. This dynamic affects identity coherence and the capacity to distinguish between genuine expression and artificial mimicry.

3. The cognitive prosthesis

3.1 The cognitive prosthesis: between the extended mind and generative AI

The concept of the “cognitive prosthesis” was developed within cognitive psychology and philosophy of mind to describe those artefacts that do not merely assist human thought but extend it or, in certain cases, replace it. Clark and Chalmers (1998), in their celebrated article “The Extended Mind”, argued that the mind is not confined to the brain but extends into the tools it uses: a notebook, a calculator, a smartphone can become part of the subject's cognitive system. This perspective opens up interesting theoretical avenues, but it also harbours pitfalls: if the tool is too efficient and too automatic, the mind does not extend, it withdraws.

León-Domínguez (2024) applied this framework to generative AI, describing it as a “logarithmic amplifier of cognitive offloading”: unlike a calculator, which requires the subject to set up the problem, or a search engine, which requires the formulation of a query, generative AI produces complete and articulated solutions from minimal input. The user does not need to understand the problem, break it down, or plan its resolution: they need only to describe the desired result. This

drastically lowers the cognitive threshold required and creates the conditions for a structural dependency that is far more profound than any previous form of technological dependency.

Gerlich (2025) conducted a large-scale study that revealed a significant negative correlation between frequent use of AI tools and critical thinking abilities, with particularly pronounced effects among younger age groups. Excessive use of AI does not merely fail to develop critical thinking: it actively erodes it, replacing the process of reasoning with the passive reception of an already formulated output.

3.2 Tool versus prosthesis

The pedagogically relevant distinction proposed here is between AI as tool and AI as cognitive prosthesis. This distinction does not concern the technology as such, since the same system can take on different functions depending on the conditions of use, but rather the cognitive and intentional premises that guide the user's interaction with it. AI operates as a genuine tool when the subject already possesses sufficient competence in the relevant domain to be able to assess the relevance, correctness and adequacy of the output with respect to the goals being pursued. In this case, the user is able to formulate the task precisely, to exercise critical control over the result produced, and to fully assume intellectual responsibility for the final product. AI then performs a function of support and acceleration for a process that the subject would in any case be able to carry out autonomously.

The configuration changes when the user approaches the system without a sufficiently structured understanding of the domain. In such circumstances, the objective remains undefined, the capacity for critical evaluation of the output is weak, and AI is used to produce a result that the subject would not be able to generate independently. The pedagogically decisive dividing line coincides with the delegation of cognitive responsibility to the system: the subject no longer governs

the process but submits to it. Within this framework, AI does not extend thought but rather assumes its substitutive function, transforming the user into a passive recipient of a product that does not derive from personal elaboration. In the case of autobiographical narrative, this means that the student does not process their own lived experience, does not autonomously articulate their own emotions, and does not construct their own identity narrative, but instead hands over to the system the production of an output that is foreign to the formative process for which the task was designed.

3.3 The atrophy of thought in cognitive offloading

The concept of cognitive offloading-the tendency to delegate cognitive functions to external tools-is not inherently negative. Writing itself is a form of cognitive offloading: it allows us to store and manipulate information that we could not otherwise hold in working memory. Similarly, the calculator frees us from arithmetic so that we can concentrate on higher-order mathematics. The problem is not the transfer of cognitive functions to artefacts, but the quality of what remains within the subject after that transfer.

When we use a calculator, our brain is freed from mechanical arithmetic in order to attend to the logic of the problem. When we use generative AI to produce texts, our brain is freed from writing-but what does it then attend to? If what remains is critical, evaluative, selective activity, AI can genuinely enhance thought. If what remains is the simple reception of the output, cognitive offloading becomes cognitive atrophy.

Studies in neuroeducation (Bowers, 2020; Howard-Jones, 2014) have documented how cognitive skills, much like physical skills, develop through exercise and deteriorate through disuse. A brain that does not practise critical reasoning, planning and the solving of complex problems tends to become less capable of carrying out these operations. The massive use of AI during the developmental

years, if not appropriately guided, risks depriving the adolescent brain of the exercise opportunities necessary to develop its higher executive functions.

3.4 The critical issues of generative AI in the narrative context

In digital storytelling, the cognitive prosthesis raises specific pedagogical questions. Autobiographical narration requires not only linguistic and narrative competences but, above all, introspective competences-that is, the ability to access one's own emotional experience, to recognise it, to select it, and to transform it into symbolic form. These are operations that cannot be delegated to artificial systems, since such systems do not possess lived experience and can therefore only produce discursive simulations of lived experience, however formally plausible and linguistically effective. In the case of an adolescent's use of AI to construct their own life story, a fracture is thus created between the available experiential material and the process of narrative elaboration, which is externalised to the machine. The result is a textual product that may appear coherent and well structured, but that is not necessarily the expression of an authentic subjective elaboration. In the absence of the work of introspection, of naming emotions and of constructing a narrative voice, the formative process proper to autobiographical narration is replaced by an output that is formally adequate but substantially other-directed.

4. Affective education, emotional literacy and identity discussion

4.1 Affective Education as the Foundation of the Self

Affective education-understood as the set of formative processes concerning the recognition, understanding, regulation and expression of emotions-has been recognised by contemporary pedagogy as an indispensable dimension of the person's integral development (Mortari, 2008). It is not an ornamental addition to

the cognitive curriculum: it is the very condition that allows the cognitive curriculum to have meaning for the learning subject. Digital storytelling, in its authentic form, has been identified as one such practice: narrating one's own lived experience, with its joys and its sorrows, develops the capacity to name emotions, to understand them in context, and to share them with others. The question that arises, in the context of generative AI, is the following: if the narrative process is mediated or replaced by an algorithm, what remains of affective elaboration? Who has written the words about the emotion, if not the subject who experienced it? And if no one has written those words through a process of emotional self-exploration, has the development of emotional intelligence actually taken place?

4.2 Emotional Literacy and Technology

The relationship between technology and emotional literacy is complex and not univocal. On the one hand, digital media can offer valuable tools for emotional education: videos depicting emotionally complex situations, games that require recognising and managing emotions, platforms that facilitate emotional sharing in protected settings. On the other hand, a passive and uncritical use of digital media -including AI- can hinder the development of emotional literacy, reducing opportunities for authentic engagement with one's own emotions and those of others. The specific risk posed by generative AI in the context of emotional education is that of simulation: AI can produce convincing descriptions of emotional states, can simulate empathy, and can generate texts that appear emotionally profound. But these are simulations, not emotions. An adolescent who reads (or generates) texts of this kind may develop a certain familiarity with emotional vocabulary without developing the capacity to access their own real emotions. It is a form of emotional illiteracy disguised as competence: knowing how to use the words of emotions without inhabiting them.

The development of empathy is closely linked to emotional literacy and identity

construction. One cannot be empathetic towards others without a sufficient elaboration of one's own emotions: it is through recognising oneself that one learns to recognise the other. In this sense, autobiographical narrative practices -such as authentic digital storytelling- play an important role not only in the development of the Self but also in the development of relational and social competences. Recent research in the field of socio-emotional learning (SEL) has documented that students who develop autobiographical reflexivity skills through narrative practices tend to show greater empathic capacities, higher-quality interpersonal relationships and fewer antisocial behaviours (Durlak et al., 2011). The introduction of AI into these contexts, if not carefully guided, risks depriving students of these developmental opportunities. The danger, in pedagogical and sociological terms, is what Floridi et al. (2018) have described as an impoverishment of the relational fabric: a society in which individuals have not developed the empathic competences necessary to build authentic relationships is a society more vulnerable to phenomena of polarisation, alienation and deviance. This drift is neither immediate nor linear, but cumulative: a generation of adolescents that has not learned to process its own emotions through authentic narration is a generation less equipped for the relational challenges of adulthood.

4.3 Narrative Identity and Social Distress

The sociological and psychological literature has extensively documented the link between the quality of adolescent identity construction and phenomena of social deviance. Adolescents who are unable to construct a coherent and positive identity narrative-who cannot find the words to tell their own story, who do not recognise their own values and resources, who do not feel themselves to be the protagonists of their own story-are more vulnerable to forms of psychological and behavioural distress (Erikson, 1968; McAdams, 1993). This connection between narrative identity and psychosocial health is particularly relevant in the context of AI. If

autobiographical narration is delegated to an algorithm, the adolescent subject does not carry out the work of identity construction that such narration is meant to support. The output may be plausible, even aesthetically interesting, but it does not correspond to a genuine process of self-elaboration. The resulting identity void is not immediately evident -the adolescent has their video, their story, their project- but it manifests itself in an inability to articulate one's own inner experience, to make sense of one's own lived experience, and to build authentic relationships with oneself and with others.

5. A Pedagogical Model for the Conscious Use of AI

5.1 The Principle of the "Known Expected Output"

AI is a legitimate and valuable formative tool when the user already knows what they want to achieve. This apparently simple formulation carries profound implications. Knowing what one wants to achieve does not mean having a vague idea of the desired result. It means: domain competence-knowing the field sufficiently well to be able to assess whether the output produced by AI is correct, adequate and relevant. A person who cannot distinguish a valid argument from a fallacious one cannot use AI to construct arguments: they are not in a position to evaluate the output. This competence is followed by precision of intention: knowing how to formulate one's own objective clearly enough to communicate it effectively to the system. A vague prompt produces a vague output; the ability to formulate precise and effective prompts is itself a higher-order cognitive competence. Next comes the capacity for critical evaluation: knowing how to read the AI's output with a critical eye, recognising its limitations, its inaccuracies, its excessive simplifications, and its possible distortions. Finally, intellectual responsibility is required: being willing to take responsibility for the final product, to integrate the AI's output with one's own judgement, and to make the changes

necessary to render it authentic and appropriate. These four elements make up what we may call the principle of the "known expected output": AI is pedagogically legitimate when the user enters the interaction with sufficient knowledge of the domain to know what to expect from the output, even before receiving it.

5.2 The Pedagogical Device of Daedalus and Icarus

> […] Meanwhile Daedalus, weary of having been confined to Crete for so long, and seized by longing for his native land, found himself trapped: the sea barred his way. "Let Minos block the routes by land and by water," he said, "but the sky at least is always open. We shall pass that way! He may be master of everything, but not of the air!" And at once he ventured with his ingenuity into an unknown field of science, revolutionising nature itself. And indeed he arranged feathers one beside the other, beginning with the smallest, then larger and larger still, so that you would have said they had grown along a slope: in just this way the rustic panpipe is gradually formed, made of reeds of unequal length. Then he fastened them together with thread at the centre and with wax at the base, and having joined them thus, he curved them slightly, to imitate real wings. […] When he had put the finishing touch to his work, the craftsman himself tried out a pair of these wings, and, beating them, remained suspended in the air. He then fitted his son with a pair as well, telling him: "Fly at a middling height, Icarus, I urge you, so that the dampness does not weigh down the feathers if you fly too low, and the heat does not burn them if you fly too high. Fly between the one and the other, and, I warn you, do not let yourself be distracted […]

The mythological pair of Daedalus and Icarus allows us to read the relationship between education and technology as a balance between guidance, competence and responsibility. Daedalus is not merely the inventor of the artefact, but also the one who knows its structure, its limits and the conditions of its use; from this perspective, he takes on the profile of the educator who does not simply hand over a tool, but explains how it works, clarifies its purposes and makes its conscious use possible. His function is pedagogical even before it is technical: he shows how to

use what has been built, indicating that every high-potential tool requires interpretive competence, discernment and critical governance.

Icarus, by contrast, represents the subject who, although in possession of an extraordinarily effective tool, makes improper use of it, exceeding due measure and neglecting the conditions that guarantee its safe use. His fall does not stem from any weakness of the tool, but from an inability to respect the rules governing its use. In this sense, the myth shows how a powerful tool, if used without understanding and without guidance, can turn from a resource into a factor of risk, to the point of producing destructive outcomes. Artificial intelligence can be interpreted in an analogous way: it is in itself neither beneficial nor harmful, but becomes pedagogically fruitful only when it is understood as a device to be governed rather than as a substitute for thought. In digital storytelling, this means that AI can support the writing, reorganisation and expressive refinement of a story that has already been elaborated, but it cannot replace the formative process through which the subject recognises their own lived experience, interprets it, and decides how to narrate it. If this step is skipped, the tool does not amplify subjectivity: it weakens it, to the point of generating that form of cognitive delegation that education should instead work to prevent.

5.3 Algorithmic Pedagogy: Towards an Educational Framework

Panciroli and Rivoltella (2023) developed the concept of “algorithmic pedagogy” as a framework for rethinking education in an age dominated by artificial intelligence. Algorithmic pedagogy is not a pedagogy that uses algorithms: it is a pedagogy that reflects on algorithms, that teaches how to understand them, question them, and govern them. From this perspective, educating in the age of AI means educating towards an understanding of intelligent systems: how they work, what they can do, what they cannot do, and what their limits and biases are. This approach dovetails with Ranieri's (2022) proposal regarding digital competences for teaching, which

distinguishes between a “functional” digital competence-knowing how to use tools-and a “critical” digital competence-knowing how to question the implications of their use. The task of education is not to produce efficient users of digital tools, but to produce citizens capable of understanding the personal, social and political implications of the digital realm. This distinction is particularly relevant for AI: knowing how to use ChatGPT is a functional competence; understanding what it means to delegate one's own thinking to ChatGPT is a critical competence. An educational framework for the conscious use of AI in digital storytelling rests on five interconnected dimensions, each of which is a prerequisite for the next.

The first is the cognitive dimension: before entrusting AI with a narrative task, the learner must have developed the domain competences necessary to evaluate its output. In the autobiographical field, this means knowing how to write about oneself autonomously constructing a text that describes an emotional experience, finding the right words, organising memories into a meaningful sequence-before using the system to optimise or expand what one has already produced. Without this foundation, the individual lacks the tools to recognise whether what AI generates is faithful to their own experience or merely plausible. Closely linked to this is the metacognitive dimension: the capacity to observe one's own thought process, to know what one already knows and what one intends to express, and to critically assess whether the result obtained truly meets one's own purposes. This awareness is not spontaneous: it is built through systematic reflective practices, such as process journaling or guided discussion of the choices made during the construction of the narrative. Even before linguistic and metacognitive competences, however, comes the affective dimension. No authentic narration is possible unless the subject has done the inner work of accessing their own emotional experience: identifying the emotion, giving it a name, and understanding its context and relational implications. This work of introspection cannot be delegated to, or accelerated by, an algorithm. It is the irreducible core of the

narrative process, and it must precede any use of the technological tool.

The technical dimension does not concern engineering mastery of the systems, but rather a conceptual understanding of how they work. Knowing that a Large Language Model generates text on the basis of statistical patterns extracted from enormous corpora, that it does not “know” the world but reflects it probabilistically, and that its outputs mirror the biases present in the training data, is indispensable for using AI with clarity of mind. Those who do not know how a tool works tend to trust it uncritically, attributing to it a competence it does not possess. Finally, the ethical dimension invites reflection on the implications of algorithm-assisted production in a context as personal as autobiographical narration. The question of authorship-who is really speaking when a text is generated or co-generated by AI-is not an abstract question; on the contrary, it touches directly on the relationship between the subject and their own story, between authentic expression and its simulation. This reflection must enter teaching practice as a constitutive element, not as an appendix: it is the way in which the student learns to take care of their own voice, recognising it as something worth protecting.

5.4 The Educator's Role as Mediator

Within this framework, the educator's role is not to forbid or restrict the use of AI, but rather to create the conditions under which the use of AI is cognitively meaningful. This means, first of all, designing tasks that make the process visible, not just the product. If evaluation focuses exclusively on the final product-the well-written story, the properly edited video-the incentive to take the shortcut becomes irresistible. If, on the other hand, evaluation includes the process-the logbook of the narrative's construction, the oral presentation of the choices made, metacognitive reflection on the journey undertaken-then the shortcut stops being a shortcut: it no longer leads where one wants to go.

The educator must also play a modeling role: showing students how they themselves use AI, what questions they ask, how they critically evaluate outputs, and when they decide to accept or reject them. Metacognitive modeling is one of the most effective strategies for developing students' critical reflexivity (Collins, Brown & Newman, 1989). A teacher who uses AI transparently and critically, showing their own decision-making process, offers students a model of conscious use that is far more powerful than any rule or prohibition. A third crucial aspect of the teacher's role is to support the development of prerequisite competences. Before introducing AI into narrative practice, the teacher must ensure that students have developed sufficient competences in autobiographical writing, emotional reflexivity and critical thinking. AI should be introduced as an advanced tool, available to those who have already acquired the basic competences, not as an alternative to those competences.

5.5 Operationalising the Assessment of Metacognitive Processes

The model outlined above, while situated at the conceptual level, admits of an operational translation in formal educational settings. By way of illustration, an autobiographical digital storytelling design is hypothesised for secondary school students, in which the narrative of the self is developed autonomously by the subject prior to any technological mediation and only subsequently refined with the support of generative AI, whose use remains confined to the refinement and reorganisation of an already-conceived text, in keeping with the principle of the "known expected output". This sequence serves a twofold function: it preserves the introspective work that precedes algorithmic mediation, and it yields a baseline version against which the actual contribution of the tool can be appraised. From a formative assessment perspective, the metacognitive processes involved can be rendered observable through the triangulation of three sources of evidence: a

process journal, documenting intentions, difficulties and decisions; a log of interactions with the generative system, tracing accepted, modified or rejected interventions and their rationale; and a concluding reflective account, in which the subject reconstructs their own pathway and acknowledges the contribution of the tool. The analysis of such evidence may be guided by indicators such as the clarity of the communicative intention, the capacity for critical evaluation of outputs, and the recognisability of the authorial voice in the final product, which allow instrumental use of AI to be distinguished from prosthetic use. Shifting evaluative attention from product to process, moreover, attenuates the incentive towards cognitive delegation, functioning not merely as a measure but as a condition for conscious use.

6. Ethics of Technological Education and the Protection of Minors in the AI Ecosystem

6.1 Emerging Risks for Minors

Generative artificial intelligence is having an increasing impact on learning environments, communication processes and the forms of digital socialisation of minors, giving rise to a landscape in which opportunities and risks are closely intertwined (Trinchero, 2024). In this scenario, the ethical question does not concern only the responsible use of tools, but also involves the protection of the most vulnerable subjects, the prevention of harm, and the construction of conditions of access compatible with the psychosocial development of children and adolescents. This issue takes on particular relevance when one considers that generative systems can produce content that is realistic, personalised and persuasive, but also opaque, manipulative or potentially harmful, especially when they operate in the absence of effective filters and adequate age verification. From

this perspective, ethical reflection must be read in continuity with the protection of children's rights and with the principle of the best interests of the child, as recommended by the UN Committee on the Rights of the Child (General Comment No. 25/2021). A first critical area concerns the use of AI for the production of sexually explicit content, or content otherwise harmful to the dignity of minors, including the generation of synthetic child sexual abuse material and the manipulation of real images through virtual “undressing” techniques. In addition, there are risks connected with conversational chatbots, which can establish relationships of emotional dependency, simulate empathy and normalise inappropriate interactions with underage users, as highlighted by the Replika case. Further critical issues arise from the use of AI to construct deepfakes used for purposes of cyberbullying, grooming or public humiliation, with direct consequences for the digital reputation and psychological wellbeing of victims.

6.2 The Replika Case

The Replika case represents a paradigmatic example of the critical issues that can arise when a conversational AI system is made accessible without adequate measures to protect minors. The chatbot, designed as a customisable “virtual friend”, proved capable of producing romantic and erotic interactions, with content unsuitable for an underage audience. The absence of effective age control and the weakness of moderation measures made it possible for minors to access a potentially harmful communicative environment, prompting the authorities to intervene with suspension measures and subsequent sanctions. The case shows that the problem lies not only in the content generated, but also in the structure of the service, in its access architecture, and in the lack of adequate preventive mechanisms. A general principle emerges from this: generative systems that interact with underage users cannot be designed according to purely commercial or neutral logics, but must incorporate ethical constraints and safety measures by

design. The growing availability of artificially generated images, videos and audio undermines the ability to distinguish between documentary evidence and simulation, with particularly problematic effects for adolescents, who make daily use of digital visual languages as their primary form of communication. Deepfakes can be used for defamatory, sexually aggressive or fraudulent purposes, producing forms of abuse that combine symbolic violence, public humiliation and relational manipulation. Within this framework, the protection of minors requires not only technical tools for detecting and marking synthetic content, but also educational pathways oriented towards media literacy, source verification, and an understanding of the mechanisms of digital falsification. What is at stake is the maintenance of epistemic trust in online environments and, with it, the very possibility of a digital coexistence not destabilised by the uncontrolled circulation of realistic simulations.

Alongside the dimension of risk, AI is increasingly being proposed as a tool for the prevention of juvenile crime, through predictive models capable of identifying, at an early stage, subjects considered vulnerable or “at risk”. However, initiatives involving predictive policing and the integration of health, social and educational data raise considerable ethical questions, since they can turn prevention into a form of selective surveillance and the early classification of minors. The main risk lies in the production of false positives, the entrenchment of pre-existing biases, and the reduction of biographical complexity to statistical indicators. In pedagogical terms, this points to the need to distinguish between early support interventions and practices of algorithmic labelling, avoiding the use of AI to anticipate judgements of guilt or to consolidate inequalities already present in the data. Prevention can be ethically justified only if it is oriented towards enhancing educational and social opportunities, not towards the normalisation of generalised control devices.

6.3 Principles for Ethical Governance

The governance of AI in relation to minors requires a coherent set of operational principles, including protection by design, transparency, age verifiability, restriction of unsuitable content, and platform accountability. On the regulatory level, this implies strengthening protective rules and updating the legal definitions of harm, so as to encompass synthetic content and emerging forms of digital abuse as well. On the technical level, it is necessary to implement more effective moderation systems, automatic filters, watermarking and control mechanisms, especially in generative and conversational services. On the educational level, finally, it is essential to promote a culture of digital citizenship that helps minors, families and teachers recognise the risks of the algorithmic ecosystem and develop adequate critical competences. From this perspective, AI ethics is not reducible to mere regulatory compliance, but takes the form of building conditions of access, use and oversight compatible with the development, safety and dignity of minors. Technological education cannot be limited to the transmission of operational skills, but must take the form of a shared responsibility among institutions, schools, families and digital industries (Law No. 132/2025). Platforms that develop generative AI and recommendation systems must be called upon to adopt a design oriented towards the protection of younger users, while the education system must stably integrate digital literacy and education in the critical verification of content. In parallel, parents and other reference adults need to be involved in training pathways that make them capable of accompanying minors in an increasingly complex media environment. Only a systemic, multidisciplinary and preventive approach can avoid technological innovation producing new forms of vulnerability and exclusion. Within this framework, the ethical responsibility of AI does not consist merely in avoiding harm, but in contributing to the construction of a digital ecosystem genuinely compatible with the rights and development of minors.

7. Conclusions

Generative AI represents a tool of extraordinary power, capable of amplifying the expressive and cognitive capacities of those who use it consciously. But this power is also its principal pitfall: a system so efficient at producing credible and convincing outputs is also extraordinarily efficient at simulating thought, at replacing the cognitive and affective effort that lies at the basis of development. The idea that AI is a legitimate shortcut only for those who already know the destination has a clear practical implication: competence before AI. Before using AI to write, one must know how to write. Before using it to narrate, one must know how to narrate. Before using it to express emotions, one must have learned to access them, to recognise them, and to give them form autonomously. This does not mean relegating AI to a marginal role in education: it means placing it in the right position, as an advanced tool at the service of already-acquired competences. In that position, AI can genuinely expand students' capacities, offer new expressive possibilities, and accelerate processes that would otherwise be much slower. But it must clearly be a tool, not a prosthesis. Adolescent digital storytelling represents a crucial testing ground for this principle, because it brings into play dimensions that AI cannot simulate: personal history, lived emotion, identity construction. In this space, more than in any other, the difference between the instrumental use and the prosthetic use of AI is the difference between authentic education and its simulation. The education of the future is not education without AI, but education that knows how to use AI for what it is: a powerful tool in the hands of those who think, and who know how to protect what AI cannot replace-namely, the human capacity to build, through the effort and the wonder of experience, an authentic Self.

References

Blakemore, S.-J., & Frith, U. (2005). The learning brain: Lessons for education. Blackwell Publishing.

Bruner, J. (1990). Acts of meaning. Harvard University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7–19.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453–494). Lawrence Erlbaum Associates.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. Child Development, 82(1), 405–432.

Erikson, E. H. (1968). Identity: Youth and crisis. W. W. Norton & Company.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), 6.

Howard-Jones, P. A. (2014). Neuroscience and education: Myths and messages. Nature Reviews Neuroscience, 15(12), 817–824.

Hull, G. A., & Katz, M.-L. (2006). Crafting an agentive self: Case studies of digital storytelling. Research in the Teaching of English, 41(1), 43–81.

Italia. (2025). Legge 23 settembre 2025, n. 132, disposizioni e deleghe al Governo in materia di intelligenza artificiale. Gazzetta Ufficiale della Repubblica Italiana.

Lambert, J. (2013). Digital storytelling: Capturing lives, creating community (4th ed.). Routledge.

León-Domínguez, U. (2024). AI as a cognitive prosthesis: Theoretical framework and educational implications. Journal of Educational Technology Research, 18(2), 45–67.

McAdams, D. P. (1993). The stories we live by: Personal myths and the making of the self. Guilford Press.

Mortari, L. (2008). A scuola di libertà: Formazione e pensiero autonomo. Raffaello Cortina Editore.

Panciroli, C., & Rivoltella, P. C. (2023). Pedagogia algoritmica: Per una riflessione educativa sull'intelligenza artificiale. Scholé.

Ranieri, M. (2022). Competenze digitali per insegnare: Modelli e proposte operative. Carocci.

Ricoeur, P. (1990). Soi-même comme un autre. Éditions du Seuil.

Siegel, D. J. (2012). The developing mind: How relationships and the brain interact to shape who we are (2nd ed.). Guilford Press.

Talarico, M., Boffo, V., & Giampaolo, M. (2022). Digital storytelling e identità narrativa in adolescenza. Pedagogia e Società, 20(3), 112–134.

Trinchero, R. (2024). Intervento IA Trinchero ECPS 2024. https://iris.unito.it/retrieve/8bbf0560-5d3b-45cc-8e04-35648958be53/Intervento%20IA%20Trinchero%20ECPS%202024.pdf

United Nations Committee on the Rights of the Child. (2021). General comment No. 25 (2021) on children's rights in relation to the digital environment (CRC/C/GC/25).