Medical Humanities & Medicina Narrativa
ISBN 979-12-218-2351-6
ISSN 2724-3281
DOI 10.53136/97912218235166
pp. 73-101 (dicembre 2025)

Dal mito alla macchina: una prospettiva antropologica cibernetica sull'interazione umana con entità intelligenti

From Myth to Machine: a Cybernetic Anthropological Perspective on Human Interaction with Intelligent Entities[1]

Stefano Di Tore, Michele Domenico Todino, Lucia Campitiello, Alessio Di Paolo, Umberto Bilotti, Jeff Holmes, Massimiliano Cilurzo, Maurizio Sibilio[*]

Riassunto: Lo studio presenta una collaborazione tra il Teaching and Learning Center for Education and Inclusive Technologies dell'Università di Salerno e la Convai Technologies Inc. Il progetto sviluppa personaggi non giocanti (NPC) basati su Intelligenza Artificiale all'interno di ambienti tridimensionali, con l'obiettivo di promuovere il coinvolgimento educativo e l'accessibilità. Tali NPC, formati attraverso contenuti pedagogici curati, agiscono come guide interattive in grado di offrire esperienze di apprendimento personalizzate e supporto inclusivo, stimolando al con-

1. The work is the result of scientific collaboration between the authors; however, Stefano Di Tore is the author of paragraph "Conclusions"; Michele Domenico Todino il the author of paragraphs 1., 2., 8. and related subparagraphs; Lucia Campitiello is the author of paragraphs 5. and 7. and related subparagraphs; Alessio Di Paolo is the author of paragraphs 3. and subparagraphs 4.2, 4.3. and 4.4.; Jeff Holmes is scientific referent for Convai and author of paragraph 4; Umberto Bilotti is the author of paragraph 6.; Massimiliano Cilurzo is co-author of subparagraph 4.1.; Maurizio Sibilio is scientific referent of the work.

[*] Stefano Di Tore, Michele Domenico Todino, Lucia Campitiello, Alessio Di Paolo, Umberto Bilotti, Massimiliano Cilurzo*, Maurizio Sibilio: University of Salerno; Jeff Holmes: Convai Technologies Inc.

tempo una riflessione critica sull'interazione uomo–macchina e sulla trasformazione dei processi di costruzione e trasmissione del sapere.

Parole–chiave: Intelligenza Artificiale conversazionale, Personaggi Non Giocanti (NPC), Ambienti Virtuali 3D, Educazione Inclusiva.

Abstract: This study outlines a collaboration between the Teaching and Learning Center for Education and Inclusive Technologies (University of Salerno) and Convai Technologies Inc. The project develops AI-driven non-player characters (NPCs) in 3D environments to foster educational engagement and accessibility. These NPCs, trained with curated pedagogical content, act as interactive guides offering personalized learning experiences and inclusive support, while prompting reflection on human–machine interaction and knowledge transformation.

Key–words: conversational AI, Non-Player Characters (NPCs), 3D Virtual Environments, Inclusive Education.

1. An Introduction: from *Myth* to *Machine*

This introductory study unfolds across three interconnected layers/hypotheses: i) a historical-cultural exploration of humanity's longstanding fascination with intelligent machines, drawing from mythology and literature; ii) a critique of contemporary ethical discourse, particularly as articulated by scholars, focusing on the limits of reassurance-based narratives; and iii) a practical application of these reflections through an interdisciplinary educational project involving AI-driven non-player characters (NPCs). By integrating cultural artefacts with cybernetic anthropology, the paper offers a critical reappraisal of how AI is conceptualized and deployed in both discourse and practice.

1.2. *Philosophical and literary genealogy of Intelligent Machines*

Long before the rise of generative artificial intelligence, the European Parliament's Resolution of 16 February 2017 on civil law rules on robotics (European Commission, 2024) demonstrated notable foresight.

In its opening clauses, it acknowledged how Western culture has, for centuries, envisioned the creation of human-like intelligent entities, from Mary Shelley's *Frankenstein* and the legend of Prague's Golem to Karel Čapek's coining of the term "robot" and the myth of Pygmalion.

These narratives, far from mere curiosities, form a cultural substrate that continues to shape our collective imagination regarding machines endowed with cognitive or emotional faculties. But one can go much further back in time, go to the roots of Western thought and thus to the Greek myths. Whether it is Talos, the bronze giant of Greek mythology, Pinocchio the puppet who longs to become human (children's fiction from 1883), or Hoffmann's Olimpia, the automaton mistaken for a woman, literature and folklore reveal an enduring human desire to animate the inanimate. Modern reinterpretations of these themes abound, from the Tin Man in *The Wizard of Oz* and Maria in *Metropolis*, to the sentient beings of Asimov's robot stories, Shirow's *Ghost in the Shell*, and Philip K. Dick's explorations of synthetic identity. Works like *Blade Runner*, *Do Androids Dream of Electric Sheep?* and *The Last Question* offer speculative but deeply relevant frameworks for grappling with the philosophical implications of AI. Already, *Neon Genesis Evangelion* is more frightening because its *Magi System* is a biological-artificial computer, and here we go beyond even quantum computers, where the computable sublimates what is possible to purely digital and mathematical machines by having hybrid human-machine elements within it. But for now, it is better to stay grounded and see, beyond these mystifications, what the state of the art is.

1.3. *Ethical Narratives and the Reassurance Paradigm*

Against this mythological and literary backdrop, contemporary ethical discourse (particularly as represented by Luciano Floridi (2022) appears to adopt a predominantly reassuring tone. He famously asserted that modern AI systems are no more intelligent than a zucchini, a metaphor intended to emphasize their lack of genuine understanding or agency.

However, this comparison merits scrutiny. Unlike a zucchini, devices such as Alexa perform tasks, respond to queries, and integrate into daily life through sophisticated engineering and programming. On other points, Floridi says that our devices are as stupid as toasters, but the ar-

gument does not vary. The deeper issue lies in the ethical implications of downplaying the transformative and potentially disruptive capacities of AI. His work, although rigorous and influential, often emphasizes the importance of managing AI within existing human-centered frameworks, advocating responsible governance and sustainability. Yet in this narrative of control and reassurance, there are no monsters, risks, underestimating both the pace and the scope of technological change.

The ecological footprint of AI systems, especially large language models (LLMs), remains largely unregulated; energy consumption data is not yet a legal requirement for market approval. This lack of transparency and regulatory inertia underscores the need for more critical perspectives that go beyond optimistic self-regulation.

1.4. *From theory to practice: NPCs in Virtual Learning Environments*

To bridge theoretical reflection with applied research, the Teaching and Learning Center for Education and Inclusive Technologies, *Elisa Frauenfelder* (University of Salerno), in collaboration with Convai Technologies Inc., has developed a system leveraging AI-driven NPCs in 3D virtual environments. These NPCs, designed to engage users through responsive dialogue, demonstrate the practical potential of conversational AI for educational purposes. When trained using curated materials provided by subject-matter experts, NPCs become effective communicators capable of delivering customized, contextually relevant content to diverse audiences, ranging from children to specialists. As interactive agents, they assume roles such as virtual guides, narrators, and facilitators of active learning. In addition to presenting historical or thematic information, they can lead quizzes and interactive activities that enhance user engagement.

Moreover, these systems are built with accessibility in mind. Features like adjustable brightness, contrast, and viewing angles accommodate users with diverse needs, including individuals with learning disabilities, physical impairments, and cognitive challenges. Thus, NPCs serve not only as pedagogical tools but also as inclusive enablers, aligning with contemporary goals of educational equity and universal design. These NPCs will be presented in detail later in this work.

2. How to avoid apocalyptic reasoning?

It is better to devote more time to those who hold an apocalyptic view of AI. Consider the scenario in which an individual walks through a dark, uncanny valley, unfamiliar and shrouded in shadow. The person might panic, imagining malevolent forces lurking behind every bend. Alternatively, they might proceed with caution, guided by a light or trusted path, avoiding danger along the way. This dark valley serves as a metaphor for the future of artificial intelligence. Unfortunately, some individuals believe that, upon entering this metaphorical space, humanity may immediately encounter malevolent and superintelligent machines. Unfortunately, some individuals believe that, upon entering this metaphorical space, humanity may immediately encounter malevolent and superintelligent machines. The fear of such entities, monstrous figures, has deep roots in human cultural memory. In the digital age, this fear took a more concrete form as early as the 1960s, when British mathematician Irving John Good, who had worked as a cryptologist at Bletchley Park alongside Alan Turing, offered the following insight: an ultra-intelligent machine is defined as a machine that can far surpass all the intellectual activities of any man, however clever (Floridi, 2022). Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an intelligence explosion, and the intelligence of man would be left far behind. Thus, the first ultra-intelligent machine is the last invention that man needs ever make, provided that the machine is docile enough to tell us how to keep it under control. It is noteworthy that this issue is so rarely discussed outside the realm of science fiction. Yet, at times, science fiction deserves to be taken seriously. Floridi spent more than a few words describing the creed of the Singularity, not because it deserves to be taken seriously, but because the AI unbelievers, the atheists, can be better understood as people overreacting to all this nonsense about the Singularity. Deeply irritated by those who worship the wrong digital gods and their failed prophecies about technological transcendence, the unbelievers (the AI atheists) take it upon themselves to prove, once and for all, that any kind of faith in true AI is mistaken, completely mistaken. AI is just computation, computers are *merely Turing machines*, Turing machines are nothing but syntactic engines, and syntactic engines cannot think, cannot know, cannot

be conscious. End of story. Today, Floridi, and he is certainly not the only one, considers the creed of the Singularity to be not only unamusing or logically irritating, but irresponsibly misleading. It is a concern typical of the wealthy world, likely to trouble those in affluent societies who seem to forget the real evils that oppress humanity and our planet. Unfortunately, if this obvious thing is still being written about, it is because it is much more newsworthy to talk about AI destroying the labour market and other such things by the media. Sadly, the COVID-19 pandemic reminded the doomsday prophets that we face tragic problems that are both serious and urgent. Climate change, regardless of what Musk might think, is our greatest existential threat.

Speculating about Hollywood-style scenarios while millions of people suffer from hunger, lack access to basic education, and are deprived of even the most affordable yet essential medical treatments necessary for a dignified life is not just nave, it is ethically negligent. They live without access to clean drinking water, the ability to wash their hands with dignity, provide a daily meal for their children, or exercise the fundamental freedoms to express their ideas and live according to their values. For some, however, artificial intelligence is seen as a phenomenon that, much like writing, diminishes cognitive abilities (Ferri, 2025), a concern already raised by Plato, as noted by Di Blas (2025 p. 36). In the knowledge that it is almost impossible to dismantle the thesis of the ia apocalyptics, we proceed below to imagine how a good ia should be designed by giving suggestions to programmers and all stakeholders who are confronted with the subject of the design and programming of these tools.

3. How to test AI software, in the light of the above?

It is not fully possible to determine with certainty whether an ethical AI application is safe unless it can be tested across all possible contexts. In such a case, the testing map would merely coincide with the full scope of implementation. What is certain is that we can gather empirical evidence of operation using feedback (such as thumbs up from the users themselves who are happy with the operation). Mathematically speaking, as this reductio ad absurdum illustrates, absolute certainty is unattainable.

What is, however, within reach, particularly in an uncertain and complex world marked by unforeseen situations, is the ability to detect when a given critical requirement has not been implemented or may not function correctly. Therefore, if critical requirements are falsifiable, it becomes possible to know when an AI4SG, *AI for Social Good*, application is not reliable, though not to ascertain when it is reliable (Floridi, 2022 pp. 221-258).

Critical requirements should be evaluated through an incremental implementation cycle. Unintended harmful effects may only emerge through such testing. Simultaneously, the software should be tested in real-world environments, provided it is safe to do so. This necessitates the adoption of an implementation cycle in which developers following these working hypotheses: i) ensure that the most critical assumptions or requirements of the application are falsifiable; ii) conduct hypothesis-testing of these assumptions and requirements within safe and controlled contexts; and, if such hypotheses are not refuted within a small set of appropriate scenarios; iii) proceed to conduct tests in increasingly broad contexts and/or evaluate a wider range of less critical requirements; all while iv) remaining prepared to halt or modify implementation as soon as harmful or undesirable effects may emerge. If we want to see into the future, let us remember that even quantum computers are subject to the limits of theoretical computer science and the rules of actual computability theory dealing with the existence or non-existence of problem-solving algorithms. It is imperative that AI be explainable, as explainability constitutes a key mechanism for fostering public trust in technology and facilitating its broader understanding. The development of a society shaped by beneficial AI necessitates a multi-stakeholder approach, one that engages all relevant actors, as this is the most effective means of ensuring that AI systems respond to societal needs. Such an approach enables developers, users, and policymakers to collaborate from the earliest stages of design and deployment (Besio *et al.*, 2025). Unavoidably, diverse cultural frameworks influence how societies perceive and interact with emerging technologies. A final observation is warranted before the presentation of the recommendations. The European approach set forth herein is intended to complement, rather than supplant, other international models, and should not be misconstrued as an expression of Eurocentrism (*Ibidem*).

There is nothing in these recommendations that confines them exclusively to the European domain; rather, they are designed to be generalizable.

4. A White-Box approach to Artificial Intelligence in education

Ethically and educationally, to allay fears about AI, it is important to realise that you are working on Turing machines in finite states, but with extremely high performance. Of course, in a manner compatible with the fact that companies must protect their copyrights, patents and profits, a white box approach is employed in computer science when the internal functioning of a system is known and can be analyzed. In this respect, it is not necessary that the software to be open, but the way it works, like a functional block diagram, is at least accessible. This is like Convai and its graphical interface that allows the user to customize all the features of its AI. For the reasons described at the beginning of this section, it is also referred to as *clear box*, *glass box*, or *transparent box*. This stands in contrast to the black box approach, in which the internal workings remain opaque or unknown. In the context of media education, adopting a white box approach to AI offers educators deeper insight into AI systems, potentially enhancing teaching and learning processes (Figure 1). Creating an AI system begins with the selection, planning, and implementation of a strategy. By opening the box (that is, examining the internal logic of an AI system) educators and developers can identify its structural components.

In this case, we are thinking of media educators and computer science teachers of all school levels compatible with the age of the students. It is good to remember that coding and computer science can also be taught to children. Think of the code.org projects, just to give an example, and the CSTA (Computer Science Teachers Association, https://csteachers.org), which focuses on creating a supportive environment for K-12 educators. This perspective resonates with Edgar Morin's concept of ecological action and the second and third *viaticum* (Morin, 1999). Let's go into more detail about the didactic-pedagogical motivations that explain what was mentioned above. Floridi would say, changing the content slightly, that philosophy becomes a guide for design. According to Morin, a strategy differs from a simple program; it is more adaptive and dynamic. While both strategies and programs can be used to organize algorithms, a program is typically predetermined, a fixed sequence of actions designed to achieve a goal. A strategy, in contrast, adapts to uncertainty in the environment. It aggregates and verifies information, adjusting its behavior based on new inputs and contextual conditions. Morin refers to this as a

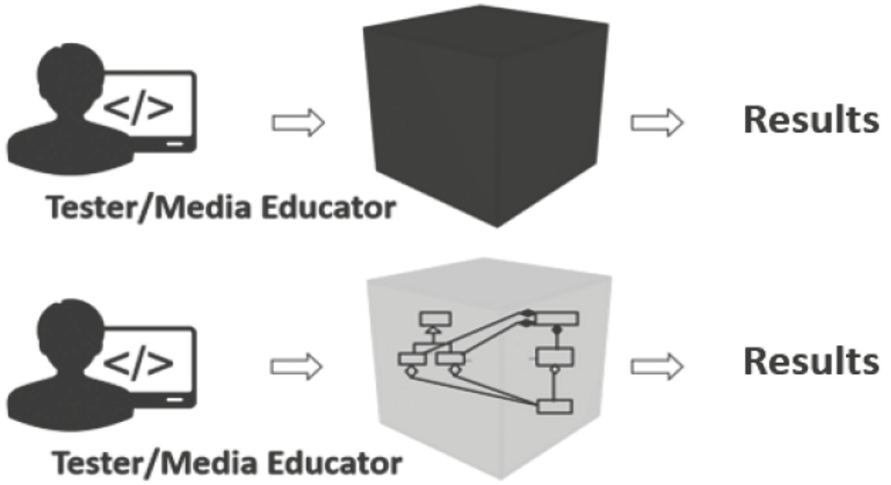

Figure 1. Comparison between the two approaches to AI.

meta, programmatic process, or the third *viaticum*, in which one commits to an uncertain path, integrating faith and hope into the decision-making process. These processes, from selecting a strategy to betting on it, mirror theories of human decision-making and are increasingly relevant in the design of intelligent machines. The more information used to train the AI, the better it can identify optimal solutions; knowledge is power.

Taking the game of chess as an example, the difficulty of determining the next optimal move is shared by both human players and programmers (Todino, 2019). The cascading nature of possible board configurations results in an explosion of potential moves, a computationally intractable problem for both humans and machines.

Instead, taking a Large Language Model (LLM) as an example, the challenge of generating the most contextually appropriate next word or phrase mirrors the complexity faced by humans when constructing coherent and meaningful discourse. Each new token generated by the model opens a branching set of linguistic possibilities, resulting in a combinatorial explosion of potential continuations. Like the cascading configurations in a chess game, this makes language generation a computationally intensive and probabilistically driven task. Although LLMs are remarkably efficient at predicting plausible sequences based on vast amounts of training data, they do not understand language in a human sense. Rather,

they operate through statistical associations, which, while powerful, can fail in cases requiring genuine reasoning, contextual judgment, or moral discernment.

The cost of a token depends on the AI model used (such as GPT-3.5, GPT-4, or GPT-4o), the method of access (API or ChatGPT subscription), and whether it applies to input (prompts) or output (completions). As of July 2025, prices via OpenAI's API range from a fraction of a cent for GPT-3.5 to several cents per 1,000 tokens for GPT-4 Turbo. Typically, 1,000 tokens equal about 750 words in English, with one token corresponding to roughly 4 characters (openai.com/it-IT/chatgpt/pricing). This obviously has an ecological impact, so making demands must be considered.

There is no need to engage in human-like etiquette with LLMs. As virtual assistants they don't require pleasantries, *thank you*, *please*, and *how are you*; concise and informative prompts are more effective because LLMs do not take offence like humans do. They are not human. Opening the AI system's black box to reveal its internal logic enables users to critically reflect on their assumptions about AI and technology to understand the things just explained. It challenges the notion that AI is inherently problematic and empowers society to define the ethical and functional framework within which AI operates. Indeed, AI software embodies the deferred intelligence of its creators. It is generally efficient at handling routine operations but often fails when confronted with irregular or novel scenarios. As recommended above, it is essential to ensure that the encoded intelligence within AI systems respects human rights and ethical principles (theme repeated several times in this paper and therefore not repeated here for brevity). In computer science, users must understand the type of hardware–software system they are interacting with and its internal logic. A system whose logic is understood is termed a white box system, while those with unknown internal mechanisms are described as black box systems. This distinction parallels educational theory: analyzing a robotic system as a black box aligns with behaviorist models, whereas adopting a white box perspective is more aligned with cognitive approaches.

4.1. *Thinking and Acting Humanly*

Russell and Norvig (2021) categorize AI systems based on various conceptual definitions found in literature. Floridi (2022) speaks instead of a defini-

tive divorce between action and intelligence, here the philosopher perhaps exaggerates, and it is better to investigate the idea of the two great veterans of AI, Russell and Norvig to be precise. Among these, AI systems that think and act humanly represent the most ambitious and futuristic category. As previously discussed, such machines remain largely in the domain of science fiction. For a system to qualify under this category, it must pass the Turing Test, defined as follows: a computer is considered to pass the test if, after submitting written questions, a human evaluator cannot determine whether the responses were generated by a human or a machine and that the Turing test is only a necessary and not a sufficient condition for a machine to appear artificially intelligent. Although this ideal has not yet been realized (and may never be, according to Rawling (1998), the goal is to program machines that are indistinguishable from humans in their cognitive and behavioral patterns. This would require a profound understanding of and ability to model the human mind, Russell & Norvig (2021) write, approximately, if we want to claim that a program thinks like a human, we first need a method for understanding how humans think. This requires exploring the inner processes of the human mind. Even with advanced technologies such as artificial neural networks and big data systems, achieving truly sentient machines remains unfeasible. Nonetheless, popular discourse often conflates AI with these hypothetical machines.

To function like a human, a machine must integrate several core capabilities, Russell & Norvig (2021) write, with great accuracy that must support the following assumptions a) natural language processing for communication; b) knowledge representation for memory and understanding; c) automated reasoning for problem solving; d) machine learning to adapt and generalize from experience. While modeling human thought remains a significant challenge, mimicking human behavior is more plausible, particularly through sophisticated hardware and biomechanical design (the *Magi System* which at the moment is science fiction for industrial ICT engineering, but that does not mean that one does not wish to go down this road, all human dreams drive scientific research the world of CEOs, programmers, current machine designs are influenced, to a certain extent, by geek culture). In everyday interactions, people often expect machines to behave in contextually appropriate ways. For instance, when using an automated toll booth or a self-service checkout, users expect an experience like that with a human operator. These systems, though simpler, are still robotic in nature.

4.2. *Acting and Thinking Rationally*

Rational action has long been considered a core aspect of human behavior. In AI, rationality provides a structured basis for decision making and is foundational to many robotics research efforts, which have achieved substantial success. Rational behavior in machines enhances user trust and fosters wider adoption in fields such as aviation (e.g., automated landing systems). This domain often includes expert systems, which emulate the decision, making processes of human specialists, and mobile agents, which are widely used in databases and operating system management (Todino, 2009). According to Russell and Norvig (2021), an agent is simply something that acts (the term comes from the Latin *agere*, meaning "to do") thus a rational agent is one that acts in a way that aims to achieve the best possible result, or, in situations of uncertainty, the most favorable expected outcome. Programming machines to act rationally, like human experts, confront the same limitations imposed by our incomplete models of cognition. For the two AI veterans, rational behavior is still considered an achievable objective, as it relies on well-established logical principles. As Aristotle noted, he was among the first to try to formalize "correct thinking", reasoning processes that are logically sound and indisputable (Russell & Norvig, 2021). However, two major challenges persist: a) translating informal human knowledge into formal logical structures, especially under uncertainty, and b) applying logic effectively in real-world scenarios where computational resources are limited.

4.3. *Weak and Strong AI in brief*

As Searle (1992) notes, AI systems can be divided into two categories: a) strong AI, i.e. Machines that *think* like humans or rational agents; b) weak AI, i.e., Machines that *simulate* human or rational behavior. Strong AI entails genuine cognitive processes, while weak AI focuses on behavioral emulation. A program functions effectively under stable and well-defined conditions. However, minimal deviations from those conditions can disrupt its operations (Morin, 1999). In contrast, strategies (although goal-directed like programs) consider various action scenarios and adjust dynamically. The point raised, now more than thirty years ago, considering Searle's distinction (1992) between strong AI and weak AI, offers the best

opportunity to clarify what Artificial Intelligence truly is today, particularly by avoiding science-fiction-like misconceptions. Clarification is not needed for insiders but seems increasingly necessary given the spread of misinformation and fears on the Internet. Current AI systems are neither thinking entities nor conscious agents. They are Turing machines, formalized symbolic systems that execute algorithms on data input to produce output. Even the most sophisticated models based on artificial neural networks (such as ChatGPT or computer vision systems) conform to this computational framework. These systems: a) do not "understand" in the human sense; b) lack intentionality (i.e., a "self" that desires or wills something); c) do not possess consciousness or phenomenological awareness of context. Although potentially their owners can achieve cultural hegemony, based on the datasets used for training in the Gramscian Chomsky's sense, and introduce unintentional bias (Besio *et al.*, 2025).

LLMs are like a confabulator friend with whom you talk; if you have read all the papers in The New York Times, you will be like an assiduous reader of this newspaper who reports to you, a little or a lot uncritically, what he has read from the first to the last issue of the newspaper. In concrete terms, modern AI does not "think" in the human sense of the term but rather performs operations on data according to predefined and trained mathematical models. Thus, speaking of strong AI as something currently existing or imminently achievable constitutes a theoretical and methodological error, an anthropomorphic projection, closer to science fiction than to science. While it is true that AI users often seek to interact with anthropomorphic agents, NPCs, avatars, and interfaces capable of voice recognition and synthesis, this desire does not imply actual intelligence or consciousness on the part of the machine. The strength of weak AI lies in its ability to simulate intelligent behavior. This simulation is, in many cases, sufficient to convince human users that the agent is understanding, deciding, or even "experiencing emotions."

However, these are cognitive illusions induced by the apparent fluidity and coherence of the responses. As also emphasized, a reading of Morin (1999), revised in an IA key program (Todino, 2019) functions effectively in stable and well-defined contexts. When the environment becomes chaotic, uncertain, or ambiguous, automation begins to falter.

This limitation is not merely technical, but epistemological: machines do not possess genuine adaptive strategies; they rely instead on imple-

mented rules and (to some extent) learned behaviors. Morin distinguishes between a program and a strategy: the former is rigid, while the latter is goal-oriented yet flexible, capable of adjusting according to contextual changes. Modern AI systems simulate strategies, but do not genuinely embody them. Even when operating in dynamic environments (e.g., in video games or mobile robotics), their adaptation is the result of predefined mathematical functions, not of intentionality or situational awareness. What appears as a "strategy" is often a conditioned response based on probabilistic or statistical patterns. From an etymological perspective, the term artificial derives from the Latin artificialism, itself rooted in *artificium*, meaning "artifice" or something crafted with skill or technique.

This notion implies a contrast with what is natural. For instance, artificial beauty refers to an appearance achieved through cosmetics or other aesthetic interventions, suggesting a form of beauty that may be real in its effects, yet perceived as lacking authenticity. Similarly, artificial intelligence, an intelligence produced by artifice, is real in its functioning but not genuine in the human sense. The same could be said for the notion of artificial consciousness (for those who wonder whether a machine will have a conscience): it may produce effects indistinguishable from those of natural consciousness, yet it lacks the ontological substance of the genuine phenomenon. While the term consciousness itself has ancient Latin roots *cum* (with) and *scire* (to know), it remains essential to underscore that AI, as previously argued, does not know in the human, epistemic sense.

4.4. *Ethical and Legal Considerations in brief*

The European Parliament (2023; 2024) has emphasized both the transformative potential and the associated risks of robotics and artificial intelligence. Ethical frameworks governing these technologies must safeguard fundamental human values, including safety, dignity, autonomy, privacy, data protection, and non-discrimination, as outlined in the European Parliament's Resolution on Civil Law Rules on Robotics previously mentioned. A key requirement stipulated in the Resolution is that it must always be possible to provide a clear explanation of any decision made with the assistance of AI, particularly when such decisions have a significant impact on one or more individuals. To this end, the implementation of a "black box" mechanism in advanced robots, capable of recording de-

cision-making processes and their underlying logic, is recommended to ensure accountability and transparency. Furthermore, the ethical governance of robotics and AI should be grounded in core principles such as beneficence, non-maleficence, autonomy, and justice. These principles are aligned with the foundational values articulated in the Treaty on European Union and the Charter of Fundamental Rights of the European Union and are reaffirmed in the EU legislative framework on robotics (HLEGAI, 2018; Arranz *et al.*, 2025). A metaphor may help to clarify the issue: suppose European countries adopt a set of regulations, based on shared recommendations, prohibiting, for instance, the dumping of waste into their neighbors' gardens. However, if neighboring non-European countries remain free to discard their waste into European gardens without similar restrictions, the effectiveness of such regulations is significantly diminished. The implication is clear: for regulatory frameworks to be truly impactful, they must be globally coordinated and consistently enforced.

But what exactly does the "waste" in this metaphor represent? It refers to a wide range of problematic externalities associated with the development and deployment of AI systems, including (but not limited to): excessive energy consumption that exacerbates climate change; data center operations disconnected from renewable energy sources; biased datasets reinforcing cultural hegemony; propaganda and misinformation that distort public understanding; applications that do not serve the common good; AI technologies designed for military or offensive purposes; systems classified as high-risk or unacceptable under ethical frameworks; the absence of ethical codes and behavioral guidelines; lack of restrictions on AI use involving minors; and the potential for widespread job displacement due to automation. These concerns underscore the need for a globally harmonized approach to AI governance, one that extends beyond regional efforts and addresses the ethical, environmental, and social impacts of AI on a planetary scale.

5. How to avoid bias in user perception? From *mythological automatons* to *cognitive inhibition*

To better understand the issue of bias, it is necessary to go deeper and start very far back. For this reason, the following section aims to trace

a historical and conceptual lineage from ancient imaginaries of autonomous machines to contemporary cognitive science and its implications for the development and use of artificial intelligence. In doing so, it examines the role of belief systems, biases, and the interplay between myth, knowledge, and rationality, further developing themes introduced earlier, such as *doxa* (δόξα), episteme (ἐπιστήμη) and the cognitive fragilities underpinning *epistemicide*. Let us take a moment to return to the myths mentioned at the beginning of this work. There has long existed only one conceptual path, already articulated in ancient Greece, for replacing slaves and manual labor: the creation of intelligent machines capable of autonomous movement and action. Mythological accounts vividly anticipated such machines. Hephaestus, the divine artisan, is said to have fashioned Talos, a bronze automaton gifted to King Minos to patrol the shores of Crete, as well as self-moving tripods that assisted during feasts (Saurdi, 2025 p. 29). These stories, found in Homeric poetry, also include figures like Triptolemus, who did not possess a traditional throne, but rather a chariot drawn by winged serpents, an artifact that, in contemporary terms, could be interpreted as a mythical precursor to robotics (Alvich *et al.*, 2025). Most strikingly, Hephaestus is described as having created golden maidens endowed with *phrenes* (mind) and *noos* (intelligence) (Saurdi, 2025 p. 31). Yet, in Homer's portrayal, these artificial beings exhibit a form of intelligence limited to *doxa*, a kind of cognition traditionally associated with servitude. This narrative encapsulates a cognitive bias of the era: slaves were perceived as capable only of *doxa*, despite their potential to also engage in *episteme* (*Ivi* p. 32). The conceptual distinction between *doxa* and *episteme* remains relevant today, particularly in understanding artificial intelligence systems and the quality of their outputs.

While *doxa* refers to mutable, uncertain knowledge based on perception and opinion, *episteme* denotes structured, causal, and stable knowledge, what might today be equated with scientific insight. In AI, high-quality data input generally leads to more reliable outputs, but an uncritical reliance on dominant narratives or model outputs risks epistemic simplification or even epistemicide, as noted by Manzetti (2025). This risk underscores Floridi's call for a new narrative, although, as he notes, this is not merely a matter of storytelling, but of cultivating knowledge, competencies, and critical faculties. Educational institutions, continuous training, and systematic monitoring of both mainstream and misleading sources

have become essential to maintaining epistemic integrity. In this context, expert systems based on physical principles and mathematical models, rather than linguistic approximations, may offer higher reliability and resistance to distortion. Still, a deeper issue remains: our capacity (or lack thereof) to inhibit erroneous beliefs when reasoning. Berthoz (2021 pp. 200–202), physiologist and honorary professor at the Collège de France, emphasizes the importance of cognitive bias and doxastic inhibition, the ability to suppress false beliefs during human problem-solving and decision-making. Humans typically blend logical reasoning with intuitive or emotionally charged beliefs, often leading to conflict. Doxastic inhibition is the cognitive mechanism that allows us to suspend misleading or deeply held convictions in favor of rational evaluation. Two strategies have been identified: one involving conscious effort to override misleading beliefs, the other excluding such beliefs early in the reasoning process.

An illustrative example from Berthoz involves a biology student at the École Polytechnique, who documented a fly's locomotion after the progressive removal of its legs. After amputating all the legs, the fly ceased to move, leading the student to conclude when all the legs are removed from a fly, it no longer understands. This ironic conclusion highlights how preconceived beliefs can distort empirical observation. At the neurocognitive level, tasks involving belief-laden but incorrect responses engage distinct brain regions, including the inferior frontal gyrus (associated with cognitive inhibition), medial frontal gyrus, cerebellum, and precuneus.

These regions are central in arbitrating between logic and belief. Importantly, people tend to believe what they *wish* to be true, rather than what is *reasoned*. For example, while a weather forecast may rationally call for carrying an umbrella, many prefer to believe in fair weather, revealing a gap between desire and logic. This gap becomes more consequential in the context of scientific communication: forecasts of extreme weather events are often dismissed as catastrophism, reflecting a broader difficulty in inhibiting comforting or ideologically convenient beliefs. This phenomenon extends to the domain of misinformation and fake news, i.e., the credibility of fake news constitutes a novel object of inquiry in cognitive neuroscience, particularly concerning individuals' ability (or inability) to inhibit belief in distorted or fabricated information. There is an important distinction between two types of misbeliefs: one adaptive and even beneficial, arising from normal belief-formation

mechanisms; the other pathological, characterized by a denial of responsibility and intentionality.

The latter has significant implications in the legal sphere.

In a classic experimental scenario, participants evaluated the intentionality of a corporate decision: when a CEO knowingly harmed the environment to boost profits, most judged the harm as intentional. However, when the same action unintentionally benefited the environment, the majority did *not* attribute intentionality. This asymmetry illustrates how moral judgment shapes our perception of intentionality, exposing a cognitive bias deeply entangled with belief management.

Having said all that was considered appropriate, in a concise manner, to motivate and correlate human and machine biases, it can be stated with a certain degree of certainty, on the technologies available on the consumer Market, that attributing human capacities, such as thought, will, consciousness, and empathy, to artificial intelligence constitutes a form of technical anthropomorphism. A realistic and critical epistemic stance is necessary: to acknowledge that AI is a powerful tool yet remains an extension of human agency rather than its ontological substitute. In this regard, the distinction between strong and weak AI must be firmly upheld while also contextualized within the scope of current technological capabilities. AI systems are advanced computational instruments, not moral agents or cognitive subjects. To assume otherwise is to conflate simulation with understanding and automatism with intelligence.

6. How can a NPC serve educational purposes and promote the common good? *Yes, maybe… it depends*

The integration of NPCs within virtual educational environments aligns closely with the principles of AI4SG and contributes significantly to achieving the targets outlined in Goal 4 of the United Nations' 2030 Agenda for Sustainable Development (sdgs.un.org/2030agenda), ensuring inclusive, equitable, and quality education for all and promoting lifelong learning opportunities. We have deliberately chosen this reference without going too far into *future literacy*, but into something that is already being implemented, as we are now almost halfway there. NPCs, By functioning as intelligent mediators between learners and multimedia content, such as videos, PowerPoint

presentations, and 3D objects, can enhance cognitive accessibility and engagement (Goal 4.1, 4.2, 4.3). These digital agents are particularly effective in addressing the needs of diverse learners, including those with disabilities or learning difficulties (Goal 4.5), by offering customized and multimodal pathways to knowledge acquisition. For example, students with dyslexia can benefit from NPCs that translate complex texts into more accessible formats or provide information orally and interactively.

Through personalized and adaptive learning facilitated by machine learning, natural language processing, and user-centered design, NPCs can identify individual knowledge gaps and adjust content delivery accordingly. This dynamic responsiveness not only increases learning effectiveness but also supports early childhood development, school readiness (Goal 4.2), and literacy and numeracy acquisition (Goal 4.6) in this regard, please note that LLMs can give explanations on mathematical topics. Furthermore, the question-time approach enabled by NPCs fosters critical thinking and active learning, with answers shaped by the learner's formulation of questions and preferred interaction style. From a technological and ethical standpoint, this model adheres to the principles of AI4SG by ensuring that AI applications are user-centric, inclusive, transparent, and aligned with societal values. Although these systems are not open-source per se, they are designed with user-friendly interfaces of Convai (Figure 2) that allow educators to create their NPCs without requiring advanced technical knowledge (convai.com/blog/building-ai-characters-knowledge-bank-with-convai).

To enhance the authenticity and engagement of an NPC, it is essential to ensure that the content within the knowledge bank is closely aligned with the character's cognitive framework and communicative style. The following guidelines should be observed: a) employ language, vocabulary, and conceptual frameworks that are congruent with the character's background, personality traits, and area of expertise; b) integrate the character's distinctive worldview, personal opinions, and lived experiences into the development of the Knowledge Bank's content; c) sustain a consistent tone, stylistic approach, and informational coherence throughout the Knowledge Bank to reinforce and maintain the integrity of the character's identity. Adhering to these principles will facilitate the creation of a nuanced and immersive conversational AI experience, thereby animating the character in a manner that is both credible and compelling

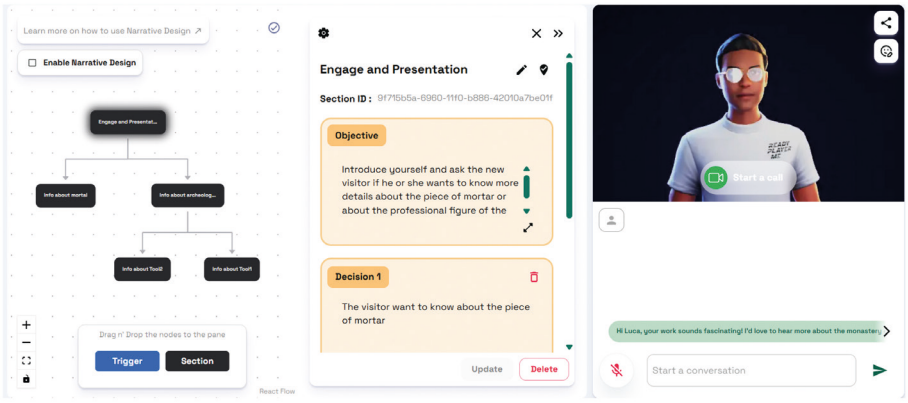

Figure 2. This is how the NPC appears in the Convai environment.

Figure 3. Imported NPC in the 3D environment (programmed with the Unity 3D engine) of the Teaching Learning Center for Education and Inclusive Technologies Elisa Frauenfelder, part of the Department of Human, Philosophical, and Educational Sciences at the University of Salerno.

(docs.convai.com/api-docs/convai-playground/character-creator-tool/knowledge-bank). This democratization of AI-based tools supports equitable access to education, particularly in underserved or remote areas, through remote connectivity and digital resource sharing. NPCs can serve as educational facilitators within virtual museums and cultural heritage platforms, deliver guided tours, interactive narratives, and contextualized learning experiences. This is a topic we have explored in depth at the

Teaching Learning Centre for Education and Inclusive Technologies – Elisa Frauenfelder (Figure 3).

This supports Goal 4.7, which emphasizes the role of education in fostering sustainable development, global citizenship, and cultural diversity, which should have a favorable impact on a culture of peace and non-violence. In this context, NPCs become not only transmitters of knowledge but also curators of values, engaging learners in conversations about human rights, sustainability, and intercultural understanding. Importantly, NPCs are not intended to replace teachers, but to act as pedagogical support, augmenting human instruction, especially in settings where educational resources are limited or unevenly distributed. They provide reliable, updatable, and context-sensitive learning assistance that can help bridge the gap between formal instruction and autonomous study.

7. What will the AIs of the future be like, if a manufacturer wants to go in the direction of simulating the human mind (if that would be useful)?

As stated by the High-Level Expert Group on Artificial Intelligence (2018), an independent expert group established by the European Commission in June 2018, the term *artificial intelligence* contains an explicit reference to the notion of intelligence. However, despite being the subject of extensive research by psychologists, biologists, and neuroscientists, intelligence, whether in machines or in humans, remains a vague and contested concept. As a result, AI researchers often prefer to work with the notion of *rationality*. Rationality is typically defined as the ability to select the best course of action to achieve a given goal, based on certain optimization criteria and within the constraints of available resources. While rationality is not the sole component of intelligence, it nevertheless constitutes a central and essential aspect of it.

But looking to the future, one should re-read Minsky more often.

One usually has two reactions when one says this: for some it is a retro thing, far removed from the useful and practical things one needs now to work in the field of ICT, for others instead it is that genius who can still give much for the future, a visionary who has imagined useful elements for the research of those who want to see beyond and link

psychology, neuroscience and technology, to understand what it means for man to reason, think, see, understand, discover in that mind of his that is so difficult to "grasp". The idea follows the hypothesis proposed by Minsky, who identified the central problem that has long challenged thinkers across ages: how can the brain, apparently so solid and concrete, be supported by something as intangible as thoughts? There seems to be no clear starting point from which researchers can work, highlighting the need for a more refined theory to understand the "building blocks" of thought and to identify the "agents of the mind." Minsky advocated for a bottom-up approach, wherein interacting agents collectively generate agencies that define behaviors. This perspective remains potentially valuable for future research, as it aligns with Gestalt theory's principle of moving from part to whole, and simultaneously from whole to part, through dynamic interactions of conflicting and compromising agencies organized in hierarchical schemas. Such frameworks partially resonate with Berthoz's concept of creative inhibition (Berthoz, 2021), characterized by processes of stop and go, which has its deepest root in the most primitive decision-making choice, hiding or flee from a predator. More advanced machines could emulate systems more complex than current multilayered artificial neural network models, evolving into software capable of "variable sense" or adaptive meaning (Berthoz, 1997; 2004). At the same time, however, our mind can simulate and emulate situations (Berthoz, 2015). Our great strength of reflective and critical perseverance arises from these opportunities given by our minds. Just to give an example, how will an exam at university go? The student simulates the exam in his head or together with a fellow student. Or I can imagine emotional states. I must meet with a person who might be angry, and I impose on myself that I will endure that emotion for the sake of peace of mind. The distinction between being, simulating, and emulating is essential for understanding the nature and limitations of artificial intelligence, and for avoiding conceptual confusion or misleading anthropomorphisms. Being refers to ontological authenticity; it denotes the actual possession of a property or state within the intrinsic structure of an entity. A human being, for example, is conscious, experiences emotions, possesses intentionality, and exhibits phenomenological subjectivity. In this sense, "being intelligent" involves not only problem-solving or generating correct responses but also understanding, autonomous learning,

and context-sensitive intentional action. Applying this notion of being to machines is philosophically and scientifically incorrect: artificial intelligence is not intelligent in the human sense, though it may appear to be according to Russell and Norvig (2021). Simulating, from the Latin *simulare* ("to make like"), means imitating a behavior, phenomenon, or state without possessing it. Simulation is concerned with outward appearance rather than internal substance. For instance, a chatbot may simulate empathy by producing emotionally appropriate responses, yet it neither experiences emotions nor understands what it is saying. Most current AI systems, including advanced language models such as ChatGPT, simulate cognitive processes, i.e. they generate coherent and relevant outputs, but without awareness or genuine understanding. Emulating, from the Latin *aemulari* ("to strive to equal"), goes beyond mere appearance. It involves faithfully replicating the functions or performance of a system, often using different underlying mechanisms, to be functionally indistinguishable from the original. In the context of artificial intelligence, a system may emulate certain aspects of human cognitive functioning, such as complex decision-making or predictive capabilities, while remaining an algorithmic machine devoid of consciousness or will. Emulation thus constitutes an advanced form of functional replication, not ontological equivalence.

These three dimensions must be clearly distinguished to avoid anthropomorphic projections onto digital technologies. Thus far, only human beings can be conscious and intentional; machines may simulate or emulate, but they cannot be what they represent. Machines that simulate and emulate human-like processes through increasingly sophisticated programs are not becoming more *human*, but rather more *useful* to humans.

Their utility lies not in replicating human nature, but in their capacity to perform reasoning processes that involve intermediate, structured steps leading to problem-solving outcomes. When these steps are transparent, that is, designed as *white-box* systems, they can offer valuable insights and support in addressing complex problems. This structured, incremental reasoning mirrors certain classical computational models, such as the simplex algorithm, which, although it typically yields sub-optimal solutions, remains fundamental in operations research for efficiently navigating solution spaces (Berthoz, 2011). These machines, however, can scale such computational strategies to unprecedented levels, amplifying their problem-solving potential, said metaphorically, to the *n*-th degree.

8. Towards a dynamic and integrated model of the mind: Embodied Mind and Adaptive Artificial Intelligence

The interaction between mind and body, in the clear knowledge that we do not mean here Cartesian dualism but simply use two commonly used terms (Damasio, 1995), suggests that mental processes cannot be understood in isolation from the body and the environmental context in which they occur. From this perspective, the mind is not a purely abstract or symbolic entity, but a dynamic and situated system that develops in close interaction with bodily experience and interpersonal relationships. In this light, the paradigm of integrated psychotherapy (Ariano, 1990; 1997; 2010) proves particularly useful for understanding how mental functioning is subject to profound reorganizations extending beyond simple cognitive changes to include emotional, bodily, and metacognitive transformations (Flavell, 1981; Flavell, 1987; Cottini, 2017; Sibilio, 2014; Sibilio, 2021); based on long-established scientific recognition that the body has its own language (Hall, 1959; Hall, 1966; Argyle, 2004). Such transformations can be triggered by significant emotional experiences, which lead to a redefinition of internal models and an evolution of affective regulation and behavioral strategies (Berthoz, 1997; 2004; 2011; 2015; 2021).

Metacognition, understood as the capacity to reflect on one's own mental states and to self-regulate psychological functioning, plays a central role in this process. It develops not in a fixed or linear fashion, but as a flexible and adaptive function responsive to experience and relationships. Based on these premises, it is possible to hypothesize the partial simulation of these processes through artificial systems. A machine endowed with adaptive capabilities and designed to modify its decision-making algorithms in response to certain conditions or "simulated emotional triggers" would not be theoretically impossible to implement. Indeed, the use of computational architectures that integrate meta-learning, adaptive dynamics, and self-monitoring mechanisms (artificial metacognition) could enable the development of artificial agents capable of modifying their strategies according to context and accumulated experience. For example, one could simulate the influence of mirror neurons from computer vision algorithms. Gallese's studies (Gallese, 2001; 2006; 2009) explore the central role of mirror neurons and embodied simulation as the neurobiological basis of empathy and intersubjectivity, and although they tread different

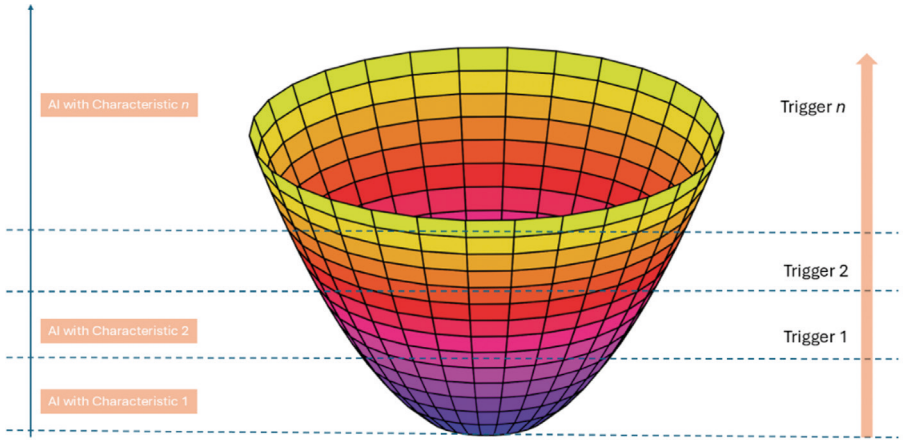

Figure 4. In Gallese's framework, the triggers are mirror neurons; whereas, in Berthoz's perspective, if one attempts a form of "reverse engineering" of the human being, the triggers correspond to dopamine levels, which reconfigure the weighting of neural networks in the human brain.

paths in some respects from Berthoz's ideas mentioned earlier, they may be one of the possible bases for moving towards algorithms that change computational levels and characteristics according to "boundary" situations. In Gallese's framework, the triggers are mirror neurons; whereas, in Berthoz's perspective, if one attempts a form of "reverse engineering" of the human being, the triggers could correspond to dopamine levels, which reconfigure the weighting of neural networks in the human brain (Figure 4). However, it is necessary to emphasize the limitations of such simulations (Eco, 1964).

Conclusions

Artificial intelligence, no matter how advanced, does not possess a biological body nor authentic subjective experience. The so-called Latin *qualia*, that is, the phenomenological aspects of conscious experience, are currently not reproducible in computational systems. Therefore, while it is possible to design machines that emulate certain aspects of human adaptive behavior, we cannot yet assert that such systems possess a mind in the full sense of the term (Elliott, 2021; Esposito, 2022). An integrated approach that views

cognitive, emotional, bodily, and relational components as interconnected offers a fruitful perspective for understanding human psychology and designing complex artificial models (Minsky, 1986; Gigenzer, 2023; Lane *et al.*, 2023; Maniello, 2023). Nonetheless, the question of consciousness and subjectivity remains open, as these elements, according to current knowledge, remain unique to human beings and are not fully replicable in artificial systems. Modern artificial intelligence has moved from logic (expert systems) to statistics (neural networks with immense amounts of input data). But human curiosity does not stop there; having reached one goal, one always tries to find a new one. This is why an attempt has been made here to provide a "suggestion" for making machines "even more intelligent" by offering a reflection that starts from the studies on the human brain carried out, separately and in some cases in competition in reasoning, by Berthoz and Gallese. While acknowledging the great potential of artificial simulations in emulating certain aspects of human mental functioning, it is essential to remain aware of their limitations. The complex dynamics that emerge from the interaction between body, mind, and relational context cannot be reduced to mere computational correlations. However, the integrated psychotherapy approach and neurophenomenological perspectives offer useful models to inspire more sophisticated artificial architectures, capable of adapting, self-reflecting, and responding to simulated "emotional" stimuli. Although the human mind remains, to this day, inimitable in its depth and plasticity, these developments open promising scenarios for the construction of artificial systems that are more sensitive to context, experience, and the intersubjective dimension of intelligence.

References

Alvich v.; cella f.; suardi l. (a cura di), *Intelligenza artificiale e scuola: guida all'uso per docenti (e genitori curiosi)*, RCS Mediagroup, Pioltello 2025.

Ariano g., *Integrazione, i Volti della Psicologia*, SIPIntegrazioni, Casoria 2010.

———, *La terapia centrata sulla persona*, Giuffrè, Milano 1990.

———, *Psicoterapia d'integrazione strutturale* – 1. Epistemologia, Armando Editore, Roma 1997.

Argyle m., *Il corpo e il suo linguaggio: studio sulla comunicazione non verbale*, Zanichelli, Bologna 2004.

Arranz d.; bianchini s.; di girolamo v.; ravet j., *Trends in the use of AI in science – a bibliometric analysis, Publications Office of the European Union*, Lussemburgo 2023, disponibile online: https://data.europa.eu/doi/10.2777/418191 (ultimo accesso: 27 luglio 2025).

Berthoz a., *L'inibizione creatrice*, Codice, Torino 2021.

———, *La scienza della decisione*, Codice, Torino 2004.

———, *La simplexité*, Odile Jacob, Parigi 2011.

———, *Le sens du mouvement*, Odile Jacob, Parigi 1997.

———, *Vicariance: le cerveau créateur de mondes*, Odile Jacob, Parigi 2015.

Besio s.; pinnelli s.; sibilio m., *Ricerche e riflessioni della pedagogia speciale sull'intelligenza artificiale*, «Ital. J. Spec. Educ. Incl.», 1 (2025), pp. 12–22.

Ciompi l.; ariano g., *Concetti di psicoterapia 1: introduzione epistemologica [video]*, disponibile online: https://www.youtube.com/watch?v=BuGYvR7Cu_Q (ultimo accesso: 27 luglio 2025).

Commissione europea, *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti*, «Gazzetta ufficiale dell'Unione Europea», Bruxelles 2024, disponibile online: https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=OJ:L_202401689 (ultimo accesso: 27 luglio 2025).

Cottini l., *Didattica speciale ed inclusione scolastica*, Carocci, Roma 2017.

Damasio a. R., *L'errore di Cartesio*, Adelphi, Milano 1995.

Di blas n., *Insegnare (e imparare) con l'IA*, in ALVICH V.; CELLA F.; SUARDI L. (a cura di), *Intelligenza artificiale e scuola: guida all'uso per docenti (e genitori curiosi),* RCS Mediagroup, Pioltello 2025, pp. 36–43.

Eco u., *Apocalittici e integrati: comunicazioni di massa e teorie della cultura di massa*, Bompiani, Milano 1964.

Elliott a., *La cultura dell'intelligenza artificiale: vita quotidiana e rivoluzione digitale*, Codice, Torino 2021.

Esposito e., *Comunicazione artificiale. Come gli algoritmi producono intelligenza sociale*, Bocconi University Press, Milano 2022.

European parliament, *Artificial Intelligence Act*, disponibile online: https://artificialintelligenceact.eu/ai-act-explorer/ (ultimo accesso: 27 luglio 2025).

———, *Resolution with recommendations to the Commission on civil law rules on robotics (2015/2103(INL))*, disponibile online: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_en.html (ultimo accesso: 27 luglio 2025).

Ferri p., *Capire l'IA "assistenti cognitivi" digitali: vantaggi e rischi,* in alvich v.; cella f.; suardi l. (a cura di), *Intelligenza artificiale e scuola: guida all'uso per docenti (e genitori curiosi),* RCS Mediagroup, Pioltello 2025, pp. 7–13.

Flavell j. H., *Cognitive monitoring, in DICKSON W. P. (a cura di), Children's oral communication*, Academic Press, New York 1981, pp. 35–60.

———, *Speculations about the nature and development of metacognition,* in WEINERT F. E.; KLUWE R. (a cura di), Metacognition, motivation, and understanding, Lawrence Erlbaum Associates, Hillsdale (NJ) 1987, pp. 21–29.

Floridi l., *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Cortina, Milano 2022.

Gallese v., *Embodied simulation: from neurons to phenomenal experience*, «Phenom. Cogn. Sci.», 4 (2005), pp. 23–48.

———, *La molteplicità condivisa. Dai neuroni mirror all'intersoggettività*, in cappuccio m. (a cura di), *Neurofenomenologia*, Bruno Mondadori, Milano 2006, pp. 293–326.

———, *The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity*, «Psychopathology», 36 (4) (2003), pp. 171–180.

———, *The "shared manifold" hypothesis: from mirror neurons to empathy*, «J. Conscious. Stud.», 8 (5–7) (2001), pp. 33–50.

———, *The two sides of mimesis: Girard's mimetic theory, embodied simulation and social identification*, «J. Conscious. Stud.», 16 (4) (2009).

Gigerenzer g., *Perché l'intelligenza umana batte ancora gli algoritmi*, Cortina, Milano 2023.

Hall e. T., *The hidden dimension, Doubleday Anchor*, New York 1966.

———, *The silent language, Doubleday Anchor*, New York 1959.

High-level expert group on artificial intelligence, *Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission in June 2018*, disponibile online: https://digital-strategy.ec.europa.eu/it/policies/expert-group-ai (ultimo accesso: 27 luglio 2025).

Lane m.; williams m.; broecke s., *The impact of AI on the workplace: main findings from the OECD AI surveys of employers and workers*, «OECD Social, Employment and Migration Working Papers», n. 288 (2023), disponibile online: https://doi.org/10.1787/ea0a0fe1-en (ultimo accesso: 27 luglio 2025).

Maniëllo d., *Augmented heritage. Dall'oggetto esposto all'oggetto narrato*. Ediz. integrale, Edizioni Le Penseur, Italia 2023.

Manzetti f., *Le narrazioni dominanti. Conta il racconto, non la realtà*, in «Corriere della Sera – La Lettura», 22 giugno 2025, p. 10.
Minsky m., *The society of mind*, Simon & Schuster, New York 1986.
Morin e., *La tête bien faite: repenser la réforme, réformer la pensée*, Éditions du Seuil, Parigi 1999.
Rawlins g. j. e., *Slaves of the machine: the quickening of computer technology*, Bradford Books, Cambridge (MA) 1998.
Russell s.; norvig p., *Intelligenza artificiale: un approccio moderno*, 4ª ed., a cura di F. Amigoni, Pearson, Milano 2022.
———, *Artificial intelligence: a modern approach*, 4th ed., Prentice Hall, Upper Saddle River (NJ) 2021.
Saurdi l., *Dei, eroi e robot – l'IA tremila anni fa*, in alvich v.; cella f.; suardi l. (a cura di), *Intelligenza artificiale e scuola: guida all'uso per docenti (e genitori curiosi)*, RCS Mediagroup, Pioltello 2025, pp. 29–35.
Searle j. r., *The rediscovery of the mind*, MIT Press, Cambridge (MA) 1992.
Sibilio m., *La didattica semplessa*, Liguori Editore, Napoli 2014.
———, *L'interazione didattica*, La Scuola, Brescia 2021.
———, *La semplessità. Proprietà e principi per agire il cambiamento*, Scholé, Brescia 2023.
Todino m. d., *Agent based distributed file system: modeling, simulation and performance evaluation*, VDM Verlag, Saarbrücken 2009.
———, *Simplexity to orient media education practices*, Aracne Editrice, Roma 2019.

Sitography

https://csteachers.org (accessed on 01/10/2025)
https://openai.com/it-IT/business/chatgpt-pricing/ (accessed on 03/10/2025)
http://sdgs.un.org/2030agenda (accessed on 01/10/2025)
https://convai.com/blog/building-ai-characters-knowledge-bank-with-convai (accessed on 02/10/2025)
https://docs.convai.com/ (accessed on 03/10/2025)