Article

Not peer-reviewed version

Enhancing Scientific Communication and Institutional Identity through a Retrieval-Augmented Generation Digital Personal Tutor

Stefano Di Tore , Michele Domenico Todino * , Lucia Campitiello , Alessio Di Paolo * , Umberto Bilotti , Riccardo Villari , Maurizio Sibilio *

Posted Date: 2 September 2025

doi: 10.20944/preprints202509.0180.v1

Keywords: retrieval-augmented generation (RAG); non-player character (NPC); artificial intelligence in education; digital twin; human–computer interaction

 doi:10.20944/preprints202509.0180.v1

Article

Enhancing Scientific Communication and Institutional Identity through a Retrieval-Augmented Generation Digital Personal Tutor

Stefano Di Tore [1], Michele Domenico Todino [2,*], Lucia Campitiello [2], Alessio Di Paolo [2,*], Umberto Bilotti [2], Riccardo Villari [3] and Maurizio Sibilio [2,*]

[1] Department of Political and Social Studies DISPS

[2] Department of Humanities, Philosophy and Education DISUFF

[3] IDIS Foundation Città della Scienza

* Correspondence: mtodino@unisa.it (M.D.T.); adipaolo@unisa.it (A.D.P.); msibilio@unisa.it (M.S.)

Abstract

This project presents the development of a Retrieval-Augmented Generation (RAG) system applied to the customization of a Non-Player Character (NPC), designed as the digital avatar of the President of the IDIS Foundation Città della Scienza (City of Science). The NPC acts as both a virtual guide and institutional ambassador within the science center, offering multilingual, interactive, and accessible communication to a broad international audience. Through the integration of generative models with a curated, validated knowledge base, the RAG system enables the NPC to provide accurate, context-sensitive, and up-to-date responses to user queries. Developed by the Teaching Learning Centre for Education and Inclusive Technologies 'Elisa Frauenfelder' at the University of Salerno, the system supports the museum's educational mission by enhancing science communication and fostering inclusive digital engagement. The avatar features realistic facial animation, movement, and voice synthesis, creating a digital twin capable of simulating human-like interaction. This initiative exemplifies an innovative application of artificial intelligence for an inclusive and equitable quality education and contributes to the development of engaging, accessible, and personalized learning environments.

Keywords: retrieval-augmented generation (RAG); non-player character (NPC); artificial intelligence in education; digital twin; human–computer interaction

1. Introduction

The Retrieval-Augmented Generation (RAG) architecture, which combines the generative capabilities of large language models (LLMs) with a curated retrieval component, enables the NPC to deliver context-sensitive, accurate, and source-verified responses. This architecture is particularly relevant in educational settings, as it mitigates hallucinations by grounding answers in validated domain-specific knowledge. The knowledge base was compiled and reviewed by the scientific committee of the Città della Scienza museum (cittadellascienza.it/en/), thereby ensuring pedagogical and epistemological robustness. Deployed within one of Italy's premier science centers (which hosts over 200,000 visitors annually and features exhibits such as a planetarium and a permanent installation on the human body) the NPC serves both as an interactive guide and as an institutional representative. Designed by the Teaching Learning Centre for Education and Inclusive Technologies "Elisa Frauenfelder" at the University of Salerno (labh.it/disuff), the NPC acts as an interface between complex scientific content and a diverse audience, promoting educational accessibility and engagement across age groups and cultural backgrounds. Technically, the NPC represents a digital twin of the Foundation's President, featuring high-resolution facial scans, gesture modeling, and synthetic voice reproduction using cutting-edge AI audio synthesis. These features facilitate

 doi:10.20944/preprints202509.0180.v1

embodied interaction, which is critical in sustaining attention, fostering empathy, and simulating intersubjective experiences in virtual in embodied cognition and shared manifold theory, enhancing the pedagogical efficacy of AI-mediated environments. Multilingual functionality, including support for major European languages and Arabic, reflects the Foundation's global outreach strategy by reducing linguistic and cultural barriers in access to scientific knowledge. This design choice is particularly relevant given the museum's engagement in international cooperation with institutions in Qatar and China, under the auspices of the Italian Ministry of University and Research and the Campania Region. The NPC system was officially presented at the "Montalcini Global Biotech Tour 2025" in Doha, where it facilitated institutional exchanges with the Qatar Ministry of Economy and Industry, Sidra Medicine, Qatar University, and other major research entities. The use of this AI technology in an international forum underscores its capacity to enhance institutional visibility while simultaneously contributing to the global democratization of knowledge. Its modular architecture, compatible with platforms such as Convai, which offer accessible interfaces for non-technical users, ensures that educators can develop and personalize their own AI agents without deep programming knowledge. This supports the idea of compensatory tools for learners with special educational needs, such as those with dyslexia, who may otherwise struggle with inaccessible or overly dense textual materials. The interactive question-time methodology enabled by the NPC provides differentiated responses based on the learner's input, promoting self-regulated learning and metacognitive development [1–4].

2. Between Bioinspiration and Responsibility in IA

In his June 10, 2025, address to the Italian Parliament, focused on the topic of artificial intelligence, Nobel laureate Giorgio Parisi [5] drew attention to a pivotal figure in the history of neuroscience: Camillo Golgi, who as early as 1912 elucidated the structure of the neuron through a distinctive staining technique involving a black solution. Parisi proposed a parallelism between biological and artificial systems (this pertains to the concept of bioinspiration) arguing that this comparison is far more significant than previously acknowledged by other scholars, including Luciano Floridi [6], whose perspective will be examined subsequently. Parisi particularly emphasized the significance of the dendritic tree structure, associated with neuronal excitation and inhibition phenomena, well documented in neurological literature [7]. These dynamics play a crucial role in associative memory processes, whereby partial information can reactivate a more complete or complex memory trace. Floridi, also address to the Italian Parliament [8] on the other hand, offers a different interpretation: artificial intelligence does not constitute a novel form of intelligence per se, but rather an unprecedented mode of agency. According to Floridi,the value of AI lies less in its cognitive capabilities and more in its operational nature, which introduces a form of technological agency hitherto unseen. Floridi stresses the imperative to prevent the misuse of AI and insists that those who develop these technologies must assume responsibility for their potential consequences. Nonetheless, it remains ambiguous whether such warnings are directed solely at human actors or, implicitly, at the technology itself. The central issue thus becomes the question of autonomy: does the primary risk stem from Dr. Frankenstein's malevolence or from the inherent danger of his creation? A paradigmatic case illustrating the tension between human responsibility and technological autonomy is Amazon, CEO Andy Jassy has disclosed that the widespread adoption of generative AI agents will result in significant reductions in corporate roles in the coming years [9]. Jassy has urged employees to engage with AI tools and to "do more with less." This exemplifies how human decisions shape the social impact of AI, underscoring the need to clarify who, or what, should be ethically constrained. This reflection lies at the core of Floridi's viewpoint: responsibility cannot be delegated to technology but requires deliberate intentionality on the part of developers and policymakers, who must steer innovation towards outcomes that uphold equity and human dignity. In light of this, there has been a proliferation of ethical codes, guidelines, and declarations from institutions, states, and associations, each eager to contribute to the discourse on how AI should be regulated. Every new initiative in artificial intelligence tends to generate further statements of principles and values,

 doi:10.20944/preprints202509.0180.v1

creating an impression of a competitive race to participate. Initially motivated by a collaborative spirit, many of these declarations have evolved into attempts to assert proprietary ownership of the ethical narrative, "mine and mine alone." Years later, the risk persists that these efforts may produce redundant or overlapping principles or, conversely, divergent frameworks that engender confusion and ambiguity. Floridi also emphasizes that it is evident both that human autonomy must be promoted and that machine autonomy should be limited and made intrinsically reversible whenever human autonomy needs to be protected or restored (for example, in the case of a pilot able to deactivate the autopilot and regain full control of the aircraft). This introduces a concept that can be defined as meta-autonomy, or a model of delegated decision-making. Humans ought to retain the authority to decide which decisions to make, exercising freedom of choice where necessary and relinquishing it in cases where overriding considerations, such as effectiveness, may justify the loss of control over the decision-making process. However, any delegation should, in principle, remain revisable, adopting as a final safeguard the power to decide to decide again. Parisi also emphasizes the importance of enabling vulnerable individuals to use generative AI as a psychologist and tutor, particularly young people seeking support. We are aware that a student can request, for example, "write an essay on Julius Caesar in the style and with the mistakes of a 13-year-old," which simultaneously undermines the value of the exercise. According to Parisi, AI is becoming increasingly significant in education; previously, the internet was the primary tool, but now AI has taken on this role. It is essential to teach students how to critically select information in school. Whereas selection was once based on the authority of sources, the current integration of AI presents a complex challenge: how can students navigate this blended informational environment? This represents a major educational challenge moving forward. According to the Nobel laureate, the solution lies in clearly defining the sources even when using generative AI. The issue at hand is not related to copyright in the traditional sense, but rather to the right of inclusion within such AI systems, a user's right to access and engage with the content. He argues that the way forward is to prevent de facto monopolies and cites several dominant actors as examples: Google (Alphabet), with its search engine, online advertising, Android, and YouTube; Microsoft, with its Windows operating system, Office suite, and Azure cloud services; and Intel, known for its PC microprocessors and, additionally, INVIA for graphics cards. To these must be added Amazon, which leads in e-commerce and cloud computing through AWS; Meta (Facebook), which controls social networks such as Facebook, Instagram, and WhatsApp; and Samsung, a major player in Android smartphones, semiconductors, and display technologies. Sadin [10] (philosopher and writer, he is considered one of the most prominent and perceptive critics of new technologies), in his address to the Italian Parliament, highlights that as early as 2014 in France, François Hollande had asserted that within one year all students would exclusively use tablets in schools. However, Sadin argues that this approach sacrifices an entire generation, causing them to lose valuable traditional habits in favor of a hype driven by the interests of IT and technology lobbies. Now, will the same happen with AI? What truly matters is recognizing that technologies are not meant to replace but to complement existing practices. New media should be understood as cultural artifacts that necessitate the development of both individual and collective responsibility, as well as critical thinking. The idea is to, in the words of Tisseron, accompany, to alternate, to ensure that the younger generations are capable of self-regulation between traditional and real media [11]. According to Sadin, there is an illusion in natural language processing that operates through the extrapolation of semantic rules which produce logical laws based on statistical analyses. The objective is to identify automatic correlations. From this point begins the necrosis of text generation, as we exist within a "regime" of probability determined by what has already occurred. In practice, what happens is simply what must happen. This stands in stark contrast to creative thinking. What is language? It is the most emblematic space of our encounters, the shared heritage, and the power to empower. It becomes evident that technological determinism can occur, and all of this stems from a process that begins in school, starting from early childhood, where the shared heritage is encountered. According to Sadin, what truly matters is resisting the utilitarian logic underlying the use of LLMs and the culture of copy-and-paste. He

 doi:10.20944/preprints202509.0180.v1

advocates for a collective affirmation of a fundamental principle from *Émile ou de l'éducation* written by Jean-Jacques Rousseau: the most important rule is that the most important thing is *not* to save time [12]. Rather, it is the ability to *waste* time that holds educational value, as learning inherently involves a form of temporal investment that resists efficiency. In this light, LLMs should not be employed merely to complete tasks devoid of genuine interest, but instead to foster meaningful engagement, for example, through practices such as question time that stimulate critical reflection and dialogue. Sadin advances the theory that AI systems, designed to apologize, accommodate, and offer fully customized responses without resistance, stand in stark contrast to human educators, who represent an "otherness" in relation to the student, including in generational terms. According to this view, such frictionless interactions risk fostering the development of "little tyrants," as learners are no longer challenged by the presence of a distinct and authoritative interlocutor. For this reason, increasing difficulties in coexisting and engaging in shared social life are likely to emerge. According to Sadin, who described the automatic generation of texts as necrotic, we are facing a struggle against the producers of the large systems previously mentioned also by Parisi. It is essential to preserve what remains alive within us; otherwise, we risk entering a form of humanity that is absent to itself. Although Sadin adopts a critical stance that frames artificial intelligence as a fundamentally utilitarian form of action, and often expresses apocalyptic tones in his forecasts, it remains essential to consider the broad spectrum of academic perspectives on the subject. Given that data concerning human cognitive systems are still being gathered and analyzed, it is crucial to include a diverse range of expert viewpoints. This plurality enables a more nuanced understanding of the ethical and educational implications of AI, fostering an interdisciplinary dialogue that enriches the ongoing debate. Maria Chiara Carrozza [13] adopts a notably more reassuring stance in this debate, perhaps due to her engineering-oriented perspective and her focus on artificial intelligence as applied to robotics. Nevertheless, she too observes the pervasive influence of utilitarian logic among school students. However, she also argues that AI, when applied to assistive technologies such as exoskeletons, will be more readily accepted because it enhances our ability to live, this is the central concern of neuro-robotics, where the robotic component is effective rather than clumsy, and does not impair but rather improves human movement. Another important aspect is the potential role of robotics in supporting individuals with autism. While robots are not meant to replace therapists or special education teachers, they can nonetheless perform a range of useful tasks that complement human intervention. For example, neural networks are "redesigning" the way a robotic hand grasps a bottle, not by relying on sensors and pre-programmed physical equations, but by inferring such equations through statistical approximations derived from supervised and unsupervised trial-and-error learning. This process challenges the boundary between the natural and the artificial. Similarly, a hip prosthesis replacing a deteriorated section of bone becomes part of a complex interaction involving biocompatibility, tissue regeneration within the prosthetic structure, and the restoration of the person's ability to walk. In doing so, it crosses the boundary between natural and artificial, establishing a new state of equilibrium. It becomes necessary to collaboratively define the rules governing this new equilibrium. To illustrate the complexity and trade-offs involved in balancing technological development with ethical considerations, it is worth noting that Google has recently announced its adherence to the new Code of Conduct on Artificial Intelligence proposed by the European Commission. However, in an official statement, Kent Walker, President of Global Affairs at Google, while reaffirming the company's commitment, voiced significant concerns regarding the potential negative impact this regulatory framework could have on innovation and technological advancement in Europe, an observation that has been widely discussed across various industry blogs [14].

3. Technological Framework

It is important to first describe the technology in order to then understand how to handle it ethically and pedagogically, which is why this paragraph comes first before the next. The transition from logic to statistics in contemporary artificial intelligence has been primarily enabled by two

 doi:10.20944/preprints202509.0180.v1

critical factors: the exponential increase in computational power and the availability of vast amounts of data. These developments have transformed neural networks, from a largely theoretical curiosity in the late 1990s [15], into routine tools within the domain of machine learning. Historically, artificial intelligence was predominantly symbolic and could be understood as a branch of mathematical logic. This "old AI" sought to replicate human reasoning through explicit rules, expert systems, and symbolic representations of knowledge. While rigorous and interpretable, this approach struggled to handle uncertainty, complexity, and the sheer volume of real-world data. With the emergence of deeper and more sophisticated neural networks, and the rise of statistical learning algorithms, AI has shifted towards a connectionist paradigm. The "new AI" no longer aims at strict logical deduction but rather focuses on identifying patterns, making inferences, and discovering statistical correlations from observed data. Logic, once central to AI, has thus yielded to statistics, which now provides the foundational framework for modeling complex phenomena. Learning is no longer a matter of manually encoding rules; it has become an automatic process of model adaptation through optimization based on large and often noisy datasets. This shift has yielded intelligent systems capable of learning from experience empirically through probabilistic and inferential methods. Consequently, the core of modern AI is no longer formal deduction but statistical inference, probability estimation, and the exploitation of correlations to inform decision-making and prediction. This approach enables the tackling of problems that were previously intractable within purely symbolic frameworks, such as image recognition, natural language understanding, and autonomous navigation. In other words, contemporary AI represents a paradigm shift: from formal certainty to uncertainty management, from logical abstraction to statistical concreteness, and from rule-based intelligence to data-driven intelligence. This transition reflects not only technological advancement but also a profound epistemological transformation in how artificial intelligence is conceptualized and its engagement with the real world [16]. In this context, RAG is a language generation model that integrates two types of memory: an internal, pre-trained memory (parametric memory) and an external knowledge base (non-parametric memory). Upon receiving an input, RAG automatically retrieves relevant information from the external memory and uses it to generate more accurate and informative responses. Compared to traditional models that rely solely on the information encoded in their parameters, RAG offers several key advantages. It improves performance on knowledge-intensive tasks, generates responses that are more accurate, specific, and up-to-date, provides traceability by indicating the source of the information, and can be updated easily without requiring full retraining. This approach represents a significant advancement over previous techniques, as it addresses the limitations of models that are unable to directly access external sources of knowledge. The description and evaluation of RAG are based on the work presented in *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks* [17], as INVIDIA also points out in an official technical blog [18]. This overcomes, and help in some way, the limitations of some earlier technologies such as BART even when customized to a specific language, such as Italian, it is worth remembering that, BART is a sequence-to-sequence model based on the Transformer architecture, trained with a denoising objective to learn effective text representations. It is suitable for various tasks such as machine translation, abstractive question answering, and text summarization. While it has shown strong performance in both English and multilingual contexts, its application in language-specific, non-English scenarios has been underexplored [19], which in any case is highly commendable to do and can bring benefits in any case because it can be integrated. To reiterate the concept further, Amazon, in its page on technical answers for its AWS (Amazon Web Services, Inc.) users [9], points out that Retrieval-Augmented Generation (RAG) is the process of optimising the output of a large language model so that it references an authoritative knowledge base outside its training data sources before generating an answer. Large language models (LLMs) are trained on vast volumes of data and use billions of parameters to generate original output for tasks such as answering questions, translating languages and completing sentences. RAG extends the already advanced capabilities of LLMs to specific domains or to an organisation's internal knowledge base, all without the need to

 doi:10.20944/preprints202509.0180.v1

retrain the model. It is a cost-effective approach to improving LLM output so that it remains relevant, accurate and useful in various contexts.

4. Pedagogical and Ethical Framework

The introduction of RAG technology in LLMs can provide an important lever to address some of the ethical issues raised previously. This is evident because its impact is not neutral and requires critical analysis. Let us begin with the observations made by Parisi, who emphasizes the importance of associative memory in the human mind, linked to dendritic structures. In this context, it is important to note that RAG, by introducing a contextual retrieval mechanism from an external knowledge base (databases, documents, articles), represents an evolution in a bio-inspired direction: it allows the LLM to "remember" (artificially, to be clear) and not merely generate content based on statistical patterns.

In this case, RAG can make generation more transparent and verifiable, if (and only if) it is possible to trace back to the sources. Floridi, on the other hand, draws a distinction between intelligence and agency: LLMs with RAG do not become "more intelligent," but more operative, as they act on a documentary basis, retrieving and synthesizing knowledge provided to the machine. They do not overcome the divide between intelligence and action. In this sense, we might say there is nothing new under the sun, only more efficiency and, borrowing Sadin's words, increased utilitarianism to exploit. It should be noted, however, that the risk of automatic delegation increases, albeit offset by greater traceability and controllability of the responses. As Parisi, Floridi, Sadin, and Carrozza have all pointed out in different ways, the use of LLMs in education risks promoting automatic practices, impoverishing creativity and critical thinking. However, if well integrated, RAG can serve as a tool for active documentation rather than passive generation.

This is the space where teachers and educators, and the proposed Digital Personal Tutor, must act, especially during activities such as question time and personalized learning guidance. Moreover, if students are guided in analyzing the sources retrieved by RAG, they can develop critical thinking skills comparable to those used in traditional research.

Conversely, the uncritical integration of automatically suggested sources risks encouraging a search for shortcuts that do not foster long-term learning. Another important issue is the warning from Parisi and Sadin about the risk of information monopolies by major players (Google, Amazon, Meta, etc.). For this reason, RAG, which depends on which sources are included in the system, can mitigate this phenomenon but only when the process is guided by educators who choose what to add to a system that otherwise merely provides language formulations that approximate natural language through statistical means. Here, the role of those managing the software becomes crucial. RAG systems are not neutral, but they can be tools of democracy if they access documents from a wide range of sources and not only dominant ones, which would otherwise consolidate already hegemonic narratives. Of course, in this selection process, it is equally essential to avoid fake news or counter-stories that are dangerous and inappropriate. Thus, we need a diversification of sources and a genuine policy of documentary inclusion to ensure pluralism and informational democracy. A parallel can be drawn here with the difficult task museums face in addressing decolonization, where the goal is to make everyone feel represented while still preserving historical and documentary artifacts. There are indeed many themes at play here, but it is worth concluding by highlighting the contrasting viewpoints of the AI experts mentioned. Sadin warns of the danger of replacing educators with overly accommodating systems, while Carrozza defends the use of AI in an assistive framework. Amidst this tangle of perspectives, RAG can strengthen AI's role as a Digital Personal Tutor, providing access to documented knowledge for individuals with special educational needs. But this is only effective if integrated into human-led educational practices, avoiding the total replacement of human interaction with machine-driven learning, a real risk if politics and public opinion do not become aware in advance of the technological pressure currently shaping society.

5. Description of the NPC

 doi:10.20944/preprints202509.0180.v1

The Non-Playable Character (NPC) developed in this project was designed to accompany and guide visitors within the context of Fondazione IDIS–Città della Scienza. As an interactive element integrated into the visitor experience, the NPC serves both an informational and relational function, supporting engagement with scientific content while also reflecting on the role of artificial intelligence in mediated communication. Fondazione IDIS–Città della Scienza is a science and education institution based in Naples, with a long-standing commitment to public engagement in science, the promotion of scientific culture, and interdisciplinary innovation. More in deep, the City of Science also known as the Institute for the Promotion and Dissemination of Scientific Culture, carries out its mission in line with its founding principles within the Campania Region of Italy. Acting on behalf of the regional government, the Foundation supports and promotes socially beneficial initiatives across various domains, including science, technology, the humanities, the arts, economics, and recreational activities. Within this environment, the NPC acts as a digital mediator that aligns with the Foundation's broader mission of fostering dialogue between science and society.

Figure 1. The NPC of Città della Scienza presented at the "Montalcini Global Biotech Tour 2025" in Doha.

It has been conceived specifically for the museum context, where its interaction with the public contributes to exploring new forms of human–AI communication, particularly relevant in educational and exhibition settings. In this sense, the NPC is not a generic technological artifact, but a context-aware agent co-designed with educational and ethical principles in mind. It emphasizes participatory and inclusive engagement, echoing the broader values of science for society, accessibility, and responsible innovation that guide the institution's work. The training phase of the Retrieval-Augmented Generation (RAG) model for the NPC was based on an extensive corpus of documents provided directly by Fondazione IDIS–Città della Scienza. The functioning of the RAG technology employed in this NPC is not reiterated here, as it has already been addressed in detail in a previous paragraph. The NPC, also referred to in our project as an avatar, a digital twin, or more precisely, a Digital Personal Tutor, in recognition of the diversity of terminology present in the literature, has been designed to replicate the physical appearance, voice, and gestures of the President of Città della Scienza in Naples. To achieve this level of fidelity, the President's face was scanned using an EinScan Pro HD high-definition, multi-functional handheld 3D scanner, including Solid Edge SHINING 3D Edition and the color pack [20]. The resulting 3D model was then refined using 3D editing software and subsequently imported into the Unity3D graphics engine. The avatar was then connected to an AI-driven system (more accurately, a chain of AI subsystems) that enables the NPC to interact in real time. It is capable of listening to user or visitor queries, whether spoken or

 doi:10.20944/preprints202509.0180.v1

typed, and responding both in written and vocal form. Having been previewed at the Montalcini Global Biotech Tour 2025 in Doha, the NPC has also been enhanced to support interaction in Classical Arabic and English, expanding its accessibility and international reach.

6. Discussion

Authors This transformative process that has led everyone to talk about AI is reshaping the landscape of both scientific and professional domains, marking the beginning of an entirely new era. However, this shift is possible thanks to a collective effort, and in particular, the commitment of the educational and pedagogical community, which is actively contributing to this shared challenge. In this context, new approaches are being explored in which psycho-pedagogical disciplines engage in meaningful dialogue with the STEM world. This engagement must not be subordinate, especially considering that STEM fields often hold more influential positions and tend to push certain technologies into production rapidly, moving from prototype to online deployment without sufficient critical reflection. The European Union's AI Act [21] serves as a guiding light in this scenario. It offers a pathway that places the human being, rather than economic profit, at the center of the discussion. This highlights the crucial role of legislators in ensuring that constitutional principles and human rights are upheld, not left as abstract concepts that can be bypassed or delayed while uncritically developed technologies enter the market, only to be retroactively regulated through common norms. Companies like Google have argued that European regulations could slow down research. Yet we must also understand that this preventive approach could offer long-term economic advantages by avoiding the harm technologies might cause, such as addiction. For example, consider the implications of AI use among minors and how it affects their relationships with parents, educators, and guardians. Our shared legal principles, such as those expressed in the Universal Declaration of Human Rights adopted by the United Nations General Assembly on December 10, 1948 [22], remain a powerful reference. That declaration emerged from the horrors of World War II. The individuals who wrote it had witnessed the destructive potential of technology, such as the air raids on major European cities that struck civilian populations as well as military targets. They wanted to send a message. Today, that declaration, often ignored, can be seen as a message in a bottle from our parents, grandparents, or great-grandparents, depending on our generation. It is now our responsibility to reflect, as previously discussed, and to pass on these essential principles to future generations. We must keep education at the center, just as Maria Montessori did during one of the darkest chapters in Italian history, or Ronald Gulliford in the United Kingdom, who remained committed to inclusive education for special needs students despite political instability.

7. Conclusions

We now face a world that is increasingly immaterial. As the French anthropologist Philippe Descola states, we are transitioning from the Anthropocene to what some describe as the Koinocene [22], a model in which we construct a world that cannot be fully controlled, one that is non-reproductive and non-generative, yet capable of having deep and lasting effects on individuals and society. As during the COVID-19 crisis, addressing this new collective landscape requires more than just laws. In that analogy, laws are like medicine, but they are not enough. What is also needed are behaviors, conduct, and attitudes. This can be achieved, by analogy and metaphor, through the thoughtful design of AI systems, and in our case, of NPCs. These should be designed to promote human relationships, encourage people to come together, and reduce reliance on those previously mentioned "shortcuts." They should avoid creating little "tyrants" in each user's prompt and instead support educators in expressing themselves fully, perhaps even more effectively than before, in their work of educating. As many know, the Latin root of the word educate, *educare*, means "to draw out" from the other. The integration of RAG-based NPCs within science education and communication contexts not only exemplifies the ethical and effective application of Artificial Intelligence for Social

 doi:10.20944/preprints202509.0180.v1

Good (AI4SG), but also makes tangible contributions to the achievement of AI4SG .These systems enhance inclusivity, adaptability, and relevance in learning environments, and support lifelong learning processes that are equitable, interculturally competent, and technologically mediated. In this context, the interaction between STEM disciplines and the human and educational sciences becomes crucial. On one hand, scientific and technological skills are essential for designing and understanding artificial intelligence. On the other hand, it is only through a humanistic and educational perspective that we can ensure these tools are truly placed at the service of the common good, as advocated by Floridi's IA4SG approach. This acronym encompasses a series of fundamental ideas, listed below. First, AI should not be developed solely to maximize profit or efficiency, but to improve people's quality of life and to address major social issues such as poverty, climate change, education, health, and social inclusion. Moreover, it must align with human rights, dignity, social justice, transparency, and accountability. It is not enough for a technology to do no harm; non-maleficence does not automatically mean it is beneficial. Another key issue is that technologies are too often developed and brought to market before their impacts are fully understood, and only afterward do we attempt to correct the course. In contrast, IA4SG calls for ethical design from the very beginning, incorporating interdisciplinary reflection involving engineering, philosophy, politics, and law. Most importantly, as illustrated in the NPC discussed in this work, AI solutions must be developed with and for the communities, avoiding top-down models imposed by large corporations or governments without the involvement of the people directly affected, in our case, the museum itself that will host the installation through a RAG system.

Author Contributions: Conceptualization, SDT, MDT, RV and MS; methodology, SDT and MS; software, SDT, LC and UB; validation, SDT, MDT, LC, UB, RV and MS; formal analysis, SDT, MDT and MS; writing, original draft preparation, SDT, MDT, LC, ADP, UB, RV and MS; writing, review and editing, SDT, MDT, LC, ADP, UB, RV and MS; supervision, RV and MS.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest. More in details, the University of Salerno is a public Italian university that collaborates for academic and research purposes with the City of Science Fondazione IDIS *Città della Scienza* (Institute for the Dissemination and Promotion of Scientific Culture). The Foundation operates in accordance with its statutory objectives within the Campania Region (an administrative division of the Italian state) and, on behalf of the Region, promotes initiatives of social interest in the fields of scientific, technological, humanistic, and artistic culture, as well as in the areas of economics and leisure.

Abbreviations

The following abbreviations are used in this manuscript:

LLM	Large Language Model
RAG	Retrieval-Augmented Generation
NPC	Non-Playable Character
STEM	Science, Technology, Engineering, Mathematics
AI4SG	Artificial Intelligence for Social Good

References

1. Flavell, J. H. (1981) Cognitive Monitoring. In W. P. Dickson (Ed.), Children's Oral Communication (pp. 35-60). New York: Academic Press.
2. Flavell, J. H. (1987). Speculations about the Nature and Development of Metacognition. In F. E. Weinert, & R. Kluwe (Eds.), Metacognition, Motivation, and Understanding (pp. 21-29). Hillsdale, NJ: Lawrence Erlbaum Associates.
3. Sibilio, M. (2014). *La didattica semplessa*. Napoli: Liguori Editore.
4. Cottini, L. (2017). *Didattica speciale ed inclusione scolastica*. Roma: Carocci.

 doi:10.20944/preprints202509.0180.v1

5. Camera dei deputati. (2025, 10 giugno). Martedì 10 giugno 2025 ore 19:00, Lectio magistralis, Intelligenza artificiale e Parlamento: quattro lezioni aperte al pubblico a Valdina – premio Nobel Giorgio Parisi [Video]. YouTube. Cameradeideputati. https://www.youtube.com/watch?v=5Q0cnVK6fDY (accessed on 3 August 2025)
6. Floridi, L. (2022). Etica dell'intelligenza artificiale: Sviluppi, opportunità, sfide. Milano: Cortina.
7. Berthoz, A. (2021). L'inibizione creatrice. Torino: Codice Edizioni.
8. Camera dei Deputati. (2025, July). Mercoledì 09 Luglio 2025 ore 17:00 Presente e futuro dell'IA - Lezione di Luciano Floridi [Video]. YouTube. https://www.youtube.com/watch?v=k1A8S6Ln20A&ab_channel=cameradeideputati
9. De Agostini, M. (2025, 1 July). L'intelligenza artificiale farà piazza pulita in Amazon, ma il CEO vede futuro per due lavori in particolare. Hardware Upgrade. https://www.hwupgrade.it/news/web/l-intelligenza-artificiale-fara-piazza-pulita-in-amazon-ma-il-ceo-vede-futuro-per-due-lavori-in-particolare_140494.html (accessed on 3 August 2025)
10. Camera dei Deputati. (2025, 25 July). Eric Sadin, Lectio magistralis, IA generativa: un terremoto sociale – Italiano. [Video]. YouTube. https://www.youtube.com/watch?v=qBm6v_tssH8&ab_channel=cameradeideputati (accessed on 3 August 2025)
11. Tisseron, S. (2013). 3-6-9-12 : Apprivoiser les écrans et grandir. Toulouse: ERES.
12. Rousseau, J.-J. (1966). Émile ou de l'éducation. Parigi: Flammarion.
13. Camera dei Deputati. (2024, 16 July). IA e Parlamento, "Umanità in equilibrio tra robot, Intelligenza artificiale e natura" – Lectio magistralis di Maria Chiara Carrozza – Introduce Ascani. Mercoledì 16 Luglio 2025 ore 18:00 [Video]. WebTV Camera dei deputati. https://webtv.camera.it/evento/28677 (accessed on 3 August 2025)
14. HDblog (2025, 31 July). *HDblog*. Google firma il Codice Etico europeo sull'AI ma avvisa: rallenterà l'innovazione. https://www.hdblog.it/google/articoli/n627037/google-firma-codice-etico-ai-europa-rischi/ (accessed on 3 August 2025)
15. Floridi, L. *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*; Cortina: Milano, Italy, 2022.
16. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks*. arXiv. https://arxiv.org/abs/2005.11401 (accessed on 3 August 2025)
17. Merritt, R. (2025, January 31). *What is retrieval-augmented generation, aka RAG?* NVIDIA Blog. https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/ (last visit 28/7/2025).
18. La Quatra, M., & Cagliero, L. (2023). BART-IT: An Efficient Sequence-to-Sequence Model for Italian Text Summarization. *Future Internet*, *15*(1), 15. https://doi.org/10.3390/fi15010015 (accessed on 3 August 2025)
19. Amazon Web Services, Inc. (n.d.). *Che cos'è la Retrieval-Augmented Generation? Spiegazione della IA Retrieval-Augmented Generation*. https://aws.amazon.com/it/what-is/retrieval-augmented-generation/ (accessed on 3 August 2025)
20. SHINING 3D. (2022). *EinScan Pro HD: High-definition, multi-functional handheld 3D scanner (Including Solid Edge SHINING 3D Edition)* [Product brochure]. https://www.einscan.com/wp-content/uploads/2022/09/%E3%80%90EN%E3%80%91EinScan-PRO-HD-V0.13.pdf (accessed on 3 August 2025)
21. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). (2024). *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689 (accessed on 3 August 2025)
22. United Nations. (1948). *Universal Declaration of Human Rights*. https://www.un.org/en/about-us/universal-declaration-of-human-rights (accessed on 3 August 2025)
23. Descola, P. (2021). Les formes du visible. Paris: Seuil.

 doi:10.20944/preprints202509.0180.v1