

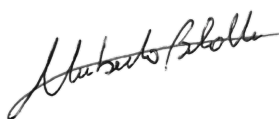
**Corso di Dottorato in
Learning Sciences and Digital Technologies
XXXVIII ciclo**

**Beyond the machine: Inclusive Technologies
for Educational Metaverse**

Tutor

Prof. Maurizio Sibilio


Dottorando

Umberto Bilotti


Co-tutor

Prof. Michele Nappi


1. Introduction

- 1.1 Concept for an educational framework
- 1.2 Learning theories and inclusion
- 1.3 Technologies for an inclusive extension

2. Educational Metaverse

- 2.1 Definition and Components
- 2.2 Historical overview of immersion and presence in Virtual Reality
- 2.3 The Internet and communication shifts
- 2.4 The Evolution of Blockchain: From digital currency to the Metaverse economy
- 2.5 Artificial Intelligence early steps for educational technology
- 2.6 The Role of AI for the development of an Educational Metaverse
- 2.7 Artificial Intelligence to building Metaverse
- 2.8 From science fiction to precursors: The birth of the conceptual Metaverse
- 2.9 Onlife existence and Educational Metaverse
- 2.10 Challenges to the Educational Metaverse: Security, Accessibility, and the Generational Gap

3. Digital for inclusive interaction

- 3.1 The role of reading in the Educational Metaverse
- 3.2 Toward an unified model for reading process
- 3.3 Dyslexia, definitions and theories on core deficit
- 3.4 Strategies to improve readability
- 3.5 GreekDyslexic: study and design for a dyslexic-friendly font

- 3.6 GreekDyslexic: experimentation of the case study
- 3.7 GreekDyslexic: results and future research for HCI
- 4. The constructivist potential of Virtual Reality
 - 4.1 Vicarious dimensions of virtual corporealities
 - 4.2 The case study of Inclusive Virtual Museum
 - 4.3 From the limits of Inclusive Virtual Museum to the Hein's constructivist model
 - 4.4 Constructivist Museum Exhibition: From Realty to Virtual
 - 4.5 Beyond description: leveraging constructivism and mediators for immersive museum education
 - 4.6 Digital asset production and virtual environment design
 - 4.7 Interdisciplinary in the virtual museum through the use of AI-based NPCs
- 5. An Inclusive VR/AI Learning Pathway
 - 5.1 Design of the Learning Unit and virtual experiences
 - 5.2 Overview of ADHD and interventions through videogames
 - 5.3 La Bottega del Vasaio – A VR experience to promote the samnite culture
 - 5.4 MUSEum Editor for a sharable peer-to-peer museum
 - 5.5 From spectators to directors: empowering students via virtual museum co-creation
- 6. Inclusion vs Intrusion: Biometrics for evaluating emotional component in the learning process

- 6.1 The emotional componentn in the learning process
- 6.2 Biometric approach to emotion recognition
- 6.3 Theory of Mind, development and intervention
- 6.4 An edugame for Howlin Progam

7. Conclusion

Abstract

This work addresses the crucial need to integrate inclusive design principles within the emerging paradigm of the Educational Metaverse. While not claiming to offer a complete prototype of the Metaverse, this work provides empirical evidence illustrating how pedagogical research, particularly in educational technology, can proactively engage with the Metaverse's potential by carving out a space for effective intervention in developing its inclusive component. The research is structured around three broad lines of inquiry, each culminating in a case study demonstrating an inclusive technological intervention.

The first line of research focuses on facilitating human-machine interaction in wearable technologies, intervening on the necessity of improving user interface accessibility through the development of a high-readability Greek alphabet font. This font is conceived as a digital tool to mitigate common reading errors encountered by dyslexic students and optimize interaction with textual content in future immersive environments.

The second line of research focuses on Virtual Reality (VR) and neurodiversity, stemming from the limitations observed during the experimentation of a previous Inclusive Virtual Museum; it proposes the creation of two new VR experiences integrated into a learning unit,

developed based on Hein's constructivist museum model and inclusive design recommendations specific for students with Attention-Deficit/Hyperactivity Disorder (ADHD), with the goal of reducing cognitive load and providing adaptive and sensory support.

Finally, the third line of research delves into Emotional Learning and AI in Autism Spectrum Disorder (ASD), exploring the importance of the emotional component in relation to the challenges that ASD students face with Theory of Mind (ToM); the intervention consists of developing an application that integrates the exercise structure of the Howlin program with the capabilities of a chatbot and a Facial Expression Recognition (FER) system. The activity focuses on the conscious emulation of primary emotions through AI-generated images, while simultaneously ensuring full ethical and normative compliance, particularly with respect to the European AI Act's prohibition on emotional inference, thanks to obtaining user consent and processing camera data exclusively *in-memory*.

Collectively, these case studies demonstrate the commitment of pedagogical research to shaping a human-centric, trustworthy, and truly inclusive Educational Metaverse, addressing critical challenges spanning accessibility, neurodiversity, and ethical compliance.

Introduction

Concept for an educational framework

A framework in Computer Science can be defined as a reusable, semi-complete structure of components that facilitates application development, significantly reducing the programmer's effort. Unlike a simple library, a framework establishes a control structure, the so-called Inversion of Control (IoC), where the framework itself calls and manages the custom code components implemented and integrated by the programmer. Some of the technical peculiarities of frameworks include:

Inversion of Control (IoC) Principle: The framework dictates the flow of execution, calling specific functions provided by the programmer.

Architecture and Methodology: The decision to adopt a framework implies following a specific development methodology and a particular code architecture.

Composition: A framework consists of a coherent set of modules, abstract classes, interfaces, and interconnected configurations that define the "hot spots" for customization.

In an analogous manner, we can conceive of an Educational Framework as a logical-operational structure designed to act as a facilitator and structured guide for the development and implementation of instructional activities by educators. The peculiarities of the educational framework can be determined through analogy, specifically:

Conceptual and Operational Foundations: Underlying an educational framework are established pedagogical theories and instructional models (which

analogously replace code modules) capable of providing a coherent lens for interpreting forms of learning.

Strategy and Design: The choice to follow a given educational framework requires the educator to adopt specific strategies and protocols for designing activities.

Structure: An educational framework consists of an interrelated set of models, protocols, and practices that offer the "hot spots" for adaptation.

The first and fundamental consideration in developing an educational framework (a logical-operational structure to support design) must concern its potential audience. To ensure the framework's validity and effectiveness, it must satisfy the needs of three distinct and interconnected levels of complexity.

The Pedagogical Community: This first level is responsibly required to engage with new perspectives capable of differently interpreting empirical data from an exponential number of diverse case studies.

The Community of Teacher and Educators: The adoption of the educational framework impacts their daily work. Therefore, the framework must support them in balancing two critical evaluations:

Efficiency: Assessing the cost-benefit ratio in terms of the time and resources dedicated to remodeling their instructional actions.

Efficacy: Analyzing the risk-benefit ratio when proposing new actions or strategies to students, evaluating the impact on the learning process.

The Students: The third and deepest level consists precisely of the students, who demand a constant update in the level of flexibility of instructional practices.

The design of our educational framework must therefore begin by starting from this last level, as it is the most oriented toward the direct stakeholders in the learning process. Consequently, the cognitive singularities of the students define one of the primary design objectives: inclusion.

Learning theories and inclusion

Although the study of the learning process has been ongoing for over a century, the perspective of inclusion has only become fundamental to instructional design in recent years. The initial learning theories were successful in explaining some of the general features of the process but failed to be exhaustive with respect to individualities. The very nature of the experiments proposed could only offer reflections on the sample rather than on the individual components.

The experiments conducted by Pavlov on conditioned stimulus are a crucial example (Pavlov, I. P. 2023). His investigations were initially focused on the digestive physiology of dogs, but they led him to understand how organisms develop the ability to associate initially neutral stimuli with biological responses that are, in themselves, involuntary. Pavlov had attached collecting devices to the cheeks of some stray dogs he had captured and was thus able to measure the progression of their saliva production. The first phase of his experiment involved taking measurements following the visual presentation of food: the increase in salivation in the presence of food was, in fact, considered an unconditioned stimulus developed by the animal during its evolutionary process. In the second phase of his experiment, the administration of food was systematically paired with a sound phenomenon, which was initially considered a neutral stimulus regarding the increase in salivation. After a period of this paired administration, an increase in salivation began to occur upon the presentation of the sound phenomenon alone. In other words, what was initially a neutral stimulus became a conditioned stimulus. Following the induced phenomenon, the dogs had consistently modified their behavior, or in other words, they had learned that the sound phenomenon anticipated an event deemed necessary for their

survival. Pavlov's experiments certainly succeeded in demonstrating the importance of context for the learning process; however, they also placed the context itself in a position of greater prominence than those who inhabit it.

The centrality of the subject regarding their own learning capabilities becomes more evident from Skinner's experiments (Skinner, B., 2016). The Skinner box was essentially a cage inside which animal subjects (rats or pigeons) were placed. The cage was equipped with a system capable of dispensing food or activating a light signal in the event of a specific interaction with the environment, and of electrifying the cage floor in the event of a different interaction. The subjects in this scenario exhibited random behaviors within the small cages until the manifestation of a certain phenomenon they perceived as positive or negative. The behaviors, which were initially random with equally random positive and negative manifestations, later began to distribute differently by frequency; in other words, a virtuous cycle was triggered with behaviors aimed at favoring the positive phenomenon. Unlike classical conditioning, where the response is a reaction to a preceding stimulus, in operant conditioning, the response is spontaneously emitted by the subject, and its consequences determine whether it will be repeated in the future.

The studies of Pavlov and Skinner, as well as those of other behaviorists, developed their theories by setting a self-imposed methodological limit: to ensure scientific rigor, they restricted themselves to evaluating the outcomes of behavioral change, refusing to investigate the organism's internal cognitive processes, the so-called "black box." Their premise was not that cognitive processes were unknown variables to be approximated, but that they were irrelevant to a science of observable behavior. The historical importance of Pavlov's and Skinner's experiments lies in having provided a scientific and, above all, universal foundation for understanding learning. However, it is precisely the search for this universal law, or generalization, that stands in sharp contrast to the current imperative of inclusion in instructional design. Behaviorist theories focus on what is shared by the majority of subjects, aiming to identify the common denominator of learning so that it can be applied in a standardized manner. The goal was the sample mean. This approach, however, inevitably ignores variability and cognitive singularities.

The limitations of early learning theories—which successfully explained general processes but failed to be exhaustive regarding individualities—necessitated a crucial shift in perspective. This began to emerge following the pivotal experiments conducted by Albert Bandura (Bandura, A., Ross, D., & Ross, S., 1961). Unlike classical behavioral studies that offered reflections primarily on the sample rather than individual components, Bandura's work focused on internal factors and the learner's capacity for choice.

The experiment involved participants who were children aged between 3 and 6 years old. The sample consisted of 72 children (36 males and 36 females), subsequently divided into three groups of 24 (12 males and 12 females). The study was structured into three phases: the first varied for each group, while the latter two were common to all three.

1. **Modeling Phase (Exposure to Behavior):** The first phase of the Bandura experiment involved dividing the sample into three distinct observation conditions: the first group was led into a room with toys and an adult operator who exhibited violent behavior towards a specific toy, the Bobo doll; the second group, conversely, was introduced into a room with toys where an operator played non-aggressively with a set of toys, showing no interest in the Bobo doll; finally, the third group was simply led into a room containing only toys, without the presence of an interacting adult model
2. **Arousal/Frustration Phase:** All groups were led to a room containing highly desirable toys. Operators initially invited the participants to play freely, but after observing a child focus their interest on a specific toy for a couple of minutes, the operator would confiscate the toy, stating that these were their best toys and were reserved for other children. This phase aimed to activate a state of potential emotional arousal.
3. **Testing Phase (Observation of Behavior):** This phase involved the observation of the children's behavior when left free to play in a room where various toys, including Bobo dolls, were present.

Observations during the third phase yielded highly significant results: the group that had witnessed the violent interactions by the adult operator subsequently reproduced, on average, more aggressive behaviors compared to the other two groups. It emerged that children tended to emulate the behaviors of the operators of the same sex; specifically, physically violent interactions were more frequent in boys, while verbally violent interactions were equal in number for both sexes. Interestingly, positive interactions tended to be emulated less than negative ones.

With these experiments, Bandura provided a critical demonstration that the learning process can occur primarily through observation or modeling. This approach, known as social or observational learning, went beyond simple behavioral reprogramming based on a stimulus-response model.

Bandura's experiments offered a crucial alternative to the behaviorist formulation of the blackbox, suggesting that learning is not a passive and mechanical process, but the result of choices and internal cognitive elaborations. The individual, especially the student, is not a simple receiver of stimuli, but the "oracle" of their own internal processes. In this sense, the biological blackbox can be associated with a cognitive "blackroom" inhabited by the student themselves. Just as a person moving from a bright environment to a dim one must gradually constrict their pupils to start seeing better and orienting themselves, the student must begin to adapt and regulate their cognitive processes. This gradual process of self-regulation is the foundation of metacognition, defined as the complex set of ideas an individual develops about the functioning of their own mind and the minds of others.

Although children, even in less sophisticated forms, are capable of monitoring their cognitive functioning, the evolution of more complex processes depends on the development of narrative thought, symbolic play, and theory of mind. Consequently, the role of the teacher becomes fundamental in promoting metacognitive development, helping students to know themselves, self-evaluate, and improve, thus positively tackling difficulties.

When a student takes control of their own learning experience, active learning is realized—a student-centered approach that inherently aligns with inclusion because it allows for the comparison of different approaches and strategies, facilitating reflection on individual

differences in methods for tackling cognitive tasks. Recognizing that every student possesses a unique form of intelligence and their own learning style is essential. Creating a learning environment that celebrates these diversities develops a sense of autonomy and competence, promoting an intrinsically inclusive and motivating learning community. This holistic vision is rooted in the pedagogy of John Dewey, who, influenced by Darwin, conceptualizes the individual in strict continuity with the surrounding environment (*transaction*), overcoming the traditional organism-context dichotomy. For Dewey, thought is an instrument of action used to interact effectively with the environment and solve problems (problem-solving), not the search for an absolute truth; knowledge is a dynamic construct derived from active interaction (the learning by doing), culminating in a transactional experience of shared meanings (Dewey, J. 2022). In a completely analogous manner, Jean Piaget also emphasizes the active and constructivist role of children as dynamic explorers, asserting that the acquisition of knowledge occurs optimally through direct interaction and concrete experience. For Piaget, knowledge and intelligence are forms of biological adaptation that evolve through continuous interaction with the environment, where the individual incessantly reformulates and adapts their cognitive schemes (schemes of assimilation), making intelligence not a discontinuous category, but an evolutionary culmination and a fundamental process for acquiring new knowledge (Piaget, J. 2013). The confluence of these perspectives—which places experience, internal cognition, and individual adaptation at the center of the process—constitutes the theoretical foundation for inclusive instructional design.

Technologies for the Inclusive Extension

Technologies as we know them today are almost always driven by a supply-and-demand logic. Companies that produce technology constantly seek to expand their potential supply, as greater demand leads to increased production and higher profits. Analogously, in the case of educational provision, the level of progress conditions an increasingly verticalized focus on specific sectors, making the perspective of a constant increase in the number of classes, courses, or programs indispensable. However, unlike the technologies produced by

companies, learning technologies should not aim to satisfy the needs of an average customer prototype. Strategies for expanding the user base for educational technologies must instead follow the paradigm of inclusion. The concept of inclusion, when applied in the educational context, extends beyond both 'insertion'—understood as providing access to the physical places of learning—and 'integration'—understood as the diversified and separate provision of instructional action for students with special educational needs. An inclusive instructional approach is, in fact, aware of both the potential the class group offers to its individual students and that every student is uniquely defined by their own cognitive singularity. Just as the principle of inclusion was reached sequentially by experimenting with the limits of previous perspectives, it is conceivable that these phases can suggest the proper development process for educational technologies.

The first phase, the insertion of a technology into an instructional practice, constitutes an awareness of its deviating potential, meaning its capacity to serve as an alternative tool for meaning-making. However, the mere introduction of technology into classrooms, as the entire teaching community attests, is insufficient. Once the initial phase of optimism regarding hypothesized scenarios passes, every technical problem leads to instructional intervention that is neither specific nor timely, and the prolonged non-use of the technology results in a consequent distrust from the student, who deems it unhelpful for their purpose. This is similar to what happens when a student, confident in their career interests or even their life plan, asks the math teacher what the trigonometry formulas will be useful for.

The second phase, integration, therefore addresses the need for a technological application that is compatible and sustainable for all parties involved in the classroom system. Efforts made over more than three decades by teachers—through specialization and refresher courses—to acquire specific competencies in both disabilities and disorders and innovative technologies converge toward achieving this goal.

Finally, the third phase, inclusion, requires structuring the technological offering in a systemic manner with respect to the entire instructional design, oriented towards the achievement of well-defined goals, and cooperative with the classroom environment.

Bibliografia (Introduction)

Pavlov, I. P. (2023). Pavlov on the conditional reflex: Papers, 1903-1936. Oxford University Press.

Skinner, B. F. (2016). The technology of teaching. BF Skinner Foundation.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology*, 63(3), 575.

Dewey, J. (2022). How we think. DigiCat.

Piaget, J. (2013). The construction of reality in the child. Routledge

Educational Metaverse

Metaverse. Definition and evolution of its components

Currently, an exhaustive, definitive, and universally accepted definition of the Metaverse has not yet been established for two primary reasons. The first reason lies in the fact that the entity itself is intrinsically composite, based on underlying technological components that are in constant and rapid evolution. The second reason is that, although several specific platforms propose themselves as models of the Metaverse, none has yet definitively

prevailed to serve as a standard. Consequently, these various proposals, despite sharing certain common characteristics, are not converging towards a unified or interoperable model.

The current Wikipedia definition proposes the metaverse as “a virtual world in which users interact while represented by avatars, typically in a 3D display, with the experience focused on social and economic connection.”

Based on this contemporary definition, we can discern not just the underlying technological components, but also the key characteristics and objectives that define the Metaverse.

The first characteristic is tied to the concept of a virtual world, implying an experience generated by a machine that offers emulations or extensions of physical spaces. The second characteristic is a direct consequence of this new immersive world, as interactions must necessarily occur through virtual representations of our identity (avatars), which is thus extended into the virtual realm. The objectives of the experience, on the other hand, focus on new possibilities in social and economic terms.

To understand how this concept was formulated and realized, it is essential to trace the fundamental milestones of several core components that enable these characteristics and objectives.

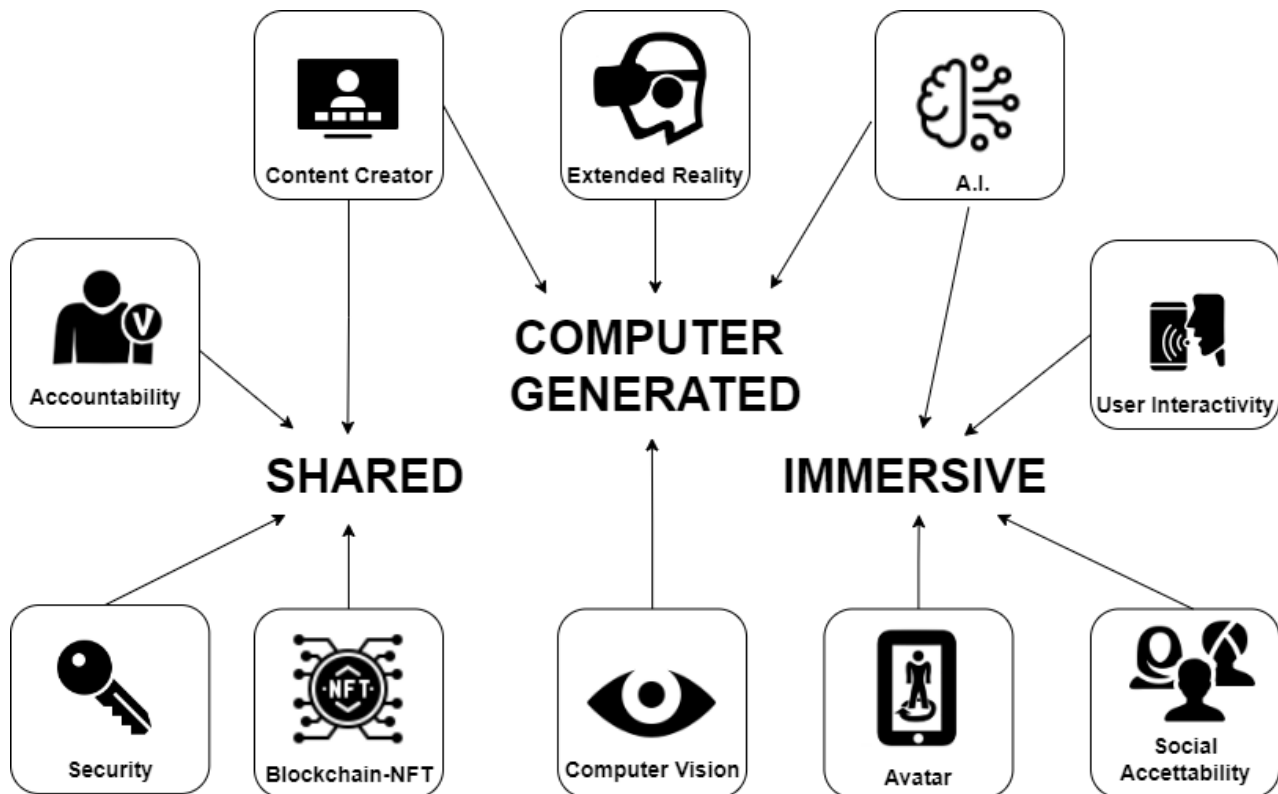


Figure 1. Some components of Metaverse and main attributes

Historical overview of Immersion and Presence in Virtual Reality

Devices used for Virtual Reality (VR) today leverage theoretical results dating back to the fifteenth century BCE with perspective geometry, and technological advancements traceable to the eighteenth century CE. However, to understand the trajectory toward the Metaverse, only key milestones relevant to the modern conception of Head Mounted Displays (HMDs) are essential.

The first significant example of a VR device, presenting many of the characteristics found in modern HMDs, was the Sword of Damocles, developed by Ivan Sutherland (1968). The system featured a visor with two optical devices simultaneously displaying images produced by a miniature cathode ray tube. Crucially, it included an external ultrasonic and mechanical

tracking system—the "Sword of Damocles" itself—designed to gather information about the user's position and send it back to the system to modify the virtual environment. Although the virtual experience was visually simple, allowing for the visualization of a cyclohexane molecule model in a cubic virtual room, it established the foundational concept of a position-aware, computer-generated visual space.

Almost simultaneously with Sutherland's work, Roberts (1966) published a paper describing the Lincoln Wand, a pen-like device whose spatial input and movements could be tracked by ultrasonic technology. A few years later, one of Sutherland's students, Vickers (1972), integrated the Lincoln Wand into the HMD system. This introduced a primary component for the interactive attribute of future VR devices, resulting in a hardware configuration highly similar to current setups.

While interactivity saw rapid hardware evolution, improving immersivity proved more complex. This challenge was linked not only to limited computational power but, critically, to the "stereo window" effect, where the user's mind perceived the screen's boundary, drastically reducing the sensation of presence, the feeling of being physically located within the virtual space. The decisive innovation arrived with the Large Expanse Enhanced Perspective (LEEP) system by Eric Howlett (1983). By utilizing a particular optical lens, the LEEP extended the field of view beyond 90°, allowing the virtual environment to fill a large portion of the user's peripheral vision. This successfully eliminated the "stereo window" effect, significantly facilitating the illusion of being truly in that place. The LEEP system continued to be used in many subsequent devices, such as the EyePhone, developed by the VPL company in 1984, which was also the first to commercialize an integrated virtual environment development system, the Reality Built For Two (Blanchard, C., et al., 1990).

Until the 1990s, the primary applications of VR were confined to high-risk contexts, such as flight simulations or military exercises. However, for a mass-market production, the virtual experience needed to be immediately pleasant by addressing virtual-sickness and reducing the cost per device. The issue of virtual-sickness was partially addressed by the Cave Automatic Virtual Environments (CAVE) system, invented in 1992 by researchers at the

Electronic Visualization Lab (Cruz-Neira et al., 1992). The CAVE, a room where walls and floor function as projection screens, created a "one-to-many" visualization tool allowing the user to move freely within the projected physical space. Despite its efficacy in high-fidelity contexts, its high investment cost, maintenance, and spatial requirements limit the CAVE's sustainability to major institutions, leaving the mass market reserved for wearable devices.

A fundamental turning point in terms of economic accessibility was 1995, when the market introduction of HMDs like the Visual I/O significantly broke the psychological threshold of \$1,000, signaling the initial democratization of VR hardware (Sherman W., Craig A., 2019). The true explosion and widespread adoption of VR in private homes, however, was ignited around 2016 by the gaming industry's release of new-generation HMDs, including iconic names like Oculus Rift, HTC VIVE, and PlayStation VR. These systems established modern VR as a complex, multi-sensorial system incorporating spatialized audio and controllers for haptic feedback (Jerald, J., 2016). The progressive trend toward simplification and autonomy culminated in the launch of fully wireless and stand-alone systems like Oculus Quest and HTC VIVE Focus (Mazhar, A. & Al Rifaaee, M., 2023). By eliminating the need for a tethered external computer, these devices drastically lowered both the logistical and financial barriers, further facilitating domestic and, by extension, educational adoption, completing the journey from a niche technology to a potential Metaverse interface.

The Internet and Communication Shifts

The roots of the Metaverse are embedded in the development of the Internet, a historical process that has progressively transformed the way we communicate, share information, and interact. The first fundamental milestone dates back to the 1960s with the creation of ARPANET and, subsequently, the standardization of the TCP/IP protocol in 1983 (Leiner, B. M., et al., 2009). This transition was crucial because it established the possibility for heterogeneous networks to communicate in a standardized way, laying the groundwork for the global connection and persistence of the Metaverse.

Direct communication between individuals took its first steps with the sending of the first email in 1971 and the spread of the BBS (Bulletin Board System) in 1978, transforming computers from mere calculation tools into instruments for asynchronous social interaction. A turning point for mass accessibility was the birth of the World Wide Web (Berners-Lee, T., 1999), enabled by Tim Berners-Lee's HTML in 1991: this not only standardized access to information via universal interfaces but also democratized the use of the Internet.

The true breakthrough toward ubiquity began with the launch of 2G (GSM) networks in 1991, which introduced instant messaging (SMS/MMS), shifting digital interaction from fixed workstations to mobile devices. The Metaverse's requirement as a shared social space was anticipated by the emergence of Social Networks, particularly Facebook in 2004, which revolutionized relational dynamics and the construction of online identity (Boyd, D. M., & Ellison, N. B., 2007).

Finally, improvements in mobile connectivity have been decisive for immersivity: the introduction of 3G in 2001, supporting data-intensive services like streaming and VoIP, represented the technological shift that enabled the extended use of 2D communication modes like video calls. The subsequent enhancement with 4G (LTE) in 2009 guaranteed the low latency and high data capacity necessary for real-time and on-the-go VR/AR experiences, effectively making the multimedia and immersive experience of the mobile Metaverse possible.

The Evolution of Blockchain: From Digital Currency to the Metaverse Economy

Even from the first realization of an immersive virtual experience, it was possible to foresee, contingent upon technological progress, an extension to an increasing number of application contexts. From this plurality of environments, the conceptual shift toward the idea of a true parallel digital universe came naturally. A similar progression has also occurred in the evolution of communication types: from websites to forums, from email to instant

messaging, and from SMS to video calls, culminating in the creation of personalized avatars tasked with emulating the native social interactions of the real world within the virtual space. These milestones have made the Metaverse possible at a sensory and connectivity level.

However, the Metaverse is defined not only by social connectivity (interaction and presence) but also by the attribute of "sharing," understood as the persistence and digital ownership of assets and identities. For this specific attribute, the fundamental milestones trace their roots well before Bitcoin, particularly in the desire to create digital systems not controlled by central entities. Foundational concepts such as public-key cryptography and the theories of the Cypherpunk movement in the 1980s and 1990s established the need for secure and anonymous digital currencies. The pioneering work of David Chaum (DigiCash) in the 1980s and 1990s demonstrated the necessity of untraceable electronic money (Chaum, D., 1983), while the publication by Haber and Stornetta on cryptographic timestamping introduced the concept of an immutable chain of blocks to ensure data integrity (Haber, S., Stornetta, W. S. 1990). These efforts conceptually culminated in Adam Back's Proof-of-Work system (1997), known as Hashcash, designed to solve the double-spending problem. If interaction was made possible by the Internet and VR, it is this long evolutionary process, culminating in the Blockchain, that provides the evidence that what is "shared" (identity, currency, property) possesses a persistent value beyond the control of a single entity.

The Blockchain is an immutable, distributed digital ledger that serves as the basis for creating a peer-to-peer electronic cash system, originally proposed in Satoshi Nakamoto's White Paper to solve the double-spending problem (Nakamoto, S. 2007). This technology achieves consensus without the need for a central authority, through a computationally expensive mechanism, the Proof-of-Work (PoW), which commits miners to validating transactions and recording an undeniable chronological order. Security is ensured by public and private key cryptography, where the private key authorizes transactions via digital signature, and the public key allows for their verification.

The genesis of the blockchain era, known as Blockchain 1.0, commenced on January 3, 2009, when Satoshi Nakamoto mined the first Bitcoin block on the Internet—the Bitcoin Genesis

Block. Initially, the technology's development and application focused almost exclusively on the transfer, exchange, and payment of currency, effectively decentralizing payments and monetary control for users. As Satoshi Nakamoto and early developers iterated on the Bitcoin blocks, the system gradually became self-sufficient, and Nakamoto ceased public communication after November 2011, remaining an anonymous legend who left behind his revolutionary creation. Key commercial milestones quickly followed: the launch of the Bitcoin exchange Mt. Gox in Tokyo in July 2010, driving a surge in users and prices, and Bitcoin reaching parity with the US dollar in February 2011. This led to the operational debut of trading platforms where Bitcoin was interchangeable with traditional currencies. By 2012, digital currencies like Ripple were emerging to facilitate cross-border currency transfers via the blockchain, and by 2013, as Bitcoin prices soared, the U.S. Department of the Treasury began issuing regulations, formally acknowledging the scope of virtual currencies.

By 2014, "Blockchain 2.0" had emerged, denoting a decentralized blockchain database that expanded far beyond currency. Enabled by open-source programming and smart contracts, its application broadened to complex financial derivatives, including futures, bonds, and crowdfunding. Crucially, this era allowed for the electronic integration of asset ownership documents, intellectual property, and notarized records with the blockchain, providing operating environments for both tangible and intangible assets. This period of expansion was further validated by institutional interest, such as the People's Bank of China's Digital Currency Symposium in 2016, which affirmed the value of digital currencies and indicated central bank exploration into their issuance. The FinTech Research Institute was formally established during this time.

The forthcoming Blockchain 3.0 era is envisioned as the period when blockchain technology achieves full application across all facets of life. It is expected to push the Internet beyond merely delivering information into an age where information itself possesses intrinsic value. Blockchain 3.0 is projected to facilitate the construction of a massive collaborative society, interconnecting finance, economics, government, health, science, culture, and the arts.

While the concept of the Metaverse often conjures images of dazzling experiences and interactive games, the most relevant scenario is that of a parallel digital world in which a functional economic ecology is inevitable. The foundational technology enabling this economic reality is the blockchain, which fulfills two irreplaceable roles within the Metaverse:

- **Data Repository and Storage:** Blockchain technology acts as a repository, allowing users to securely store essential data anywhere within the Metaverse environment.
- **Economic System Bridge:** It provides a complete economic system that seamlessly connects the Metaverse's virtual world with the real world. Specifically, Non-Fungible Tokens (NFTs), enabled by blockchain, allow virtual goods to acquire the characteristics of physical objects, enabling users to trade virtual items in the same way they would in the real world.

Furthermore, the quality of objects and experiences created within the Metaverse relies heavily on the quality of data shared by users from the real world. By providing complete audit trails of all transactions, blockchain increases data quality and ensures that individuals and organizations can validate every transaction with high integrity.

Artificial Intelligence early steps for educational technology

Due to the commitment required of the pedagogical community in recent years in reflecting on the issues advanced by the increasingly widespread use of Artificial Intelligence (AI) methods in the educational context, the discussion of the fundamental stages of AI associated with the technical and conceptual predisposition of the metaverse necessitates a distinction into three main phases. The first precedes the initial proposals for the metaverse; the second is contemporary with the emergence of the metaverse proposal for education, identifying AI as a common denominator of many of the methods underlying both

technologies for teaching and educational technologies in the strict sense; the third has simultaneously brought closer the possibility for all its users to construct the metaverse, thereby making it increasingly participatory. A fundamental starting point in this journey is certainly the formulation, by Alan Turing in 1950, of his famous Turing Test (Turing, A. M., 2007). In this context, Turing proposed replacing the metaphysical question "Can machines think?" with an empirical one: "Is there a machine that can play the Imitation Game well?". The original game involved three participants in three separate rooms: A, a human being (the interrogated); B, the machine to be tested (the other interrogated); and C, the interrogator, who is also a human being. The interrogator communicates with A and B via a teleprinter (i.e., a keyboard and a screen, comparable today to a text chat) to avoid being influenced by physical cues. The interrogator's goal is to figure out which of the two is the machine; the machine's (B) goal is to deceive the interrogator and convince them it is the human being; and the human being's (A) goal is to help the interrogator (C) correctly identify them. According to the Turing Test, a machine is considered "intelligent" if, after a reasonable period of conversation, the interrogator is unable to reliably distinguish the machine from the human being.

This test was not aimed at defining intelligence itself, but at establishing a functional criterion for judging a machine's capacity to exhibit intelligent behavior indistinguishable from human behavior during a textual conversation.

The gap between Turing's conceptual proposal and modern chatbots was bridged by a progressive evolution of machine learning techniques. These techniques initially focused on classification and pattern analysis, not text generation, but they provided the essential tools for subsequent development.

While the 1950 Turing Test set the conceptual goal for conversational AI, the true revolution of Machine Learning (ML) and Deep Learning (DL) occurred in domains seemingly unconnected to text generation, yet strategically crucial for the technical foundations of the Metaverse. The gap between Turing's goal and current chatbots was filled by a series of innovations that, by focusing on the classification and analysis of complex patterns, provided solutions to problems of biometric interaction and networking that the Metaverse would later inherit and require on a large scale. Starting as early as the first forms of ML

based on Support Vector Machines (SVM) in the 1990s and 2000s, systems were already refining the ability to distinguish and categorize data. Support Vector Machines are supervised machine learning models used for classification (the primary use) and regression analysis. They represent a powerful technique for solving complex pattern recognition problems by focusing on the boundaries between different classes of data (Vapnik, V., 2013).

The Role of AI for the development of an Educational Metaverse

This analytical power exploded with the advent of Deep Learning, which, thanks to Convolutional Neural Networks (CNNs), made possible the automatic extraction of features from raw data, such as pixels or sound waves. This visual and audio analysis capability was immediately applied to solve concrete problems of distance e-learning.

One of the most pressing problem of remote exams, the violation of academic integrity, found its most robust solution in AI, even before the advent of headsets. Biometrics, supported by AI, has been the key technology for monitoring student identity and behavior. Techniques like Facial Recognition, Voice Recognition, and Keystroke Dynamics analysis were already advanced tools for identity verification and fraud detection (person substitution, unauthorized collaboration). Furthermore, AI-based systems such as eye-tracking (even via simple webcams) were used to counter the use of prohibited materials by monitoring the gaze direction. The Metaverse inherits the entire taxonomy of fraud threats and, thanks to the richer data provided by sensors integrated into wearables (e.g., more accurate data on eye position and body movement), enhances the effectiveness of these biometric methodologies (Bilotti, U. et al., 2024). AI had already established a strong link between behavioral monitoring and learning support. Research in the field of e-learning had long developed methods for Attention Level Estimation (based on blinking and pupillary dilation) and Facial Expression Recognition to detect emotional states such as boredom, frustration, or stress. These tools, applied to online courses, allowed the teacher to obtain objective feedback on content reception and to intervene with targeted didactic modifications or interactions. AI, in this context, acted as an invisible tutor supporting

personalization. In the Metaverse, this function is elevated: biometric data acquired in an immersive environment is more faithful and detailed, allowing the AI to calibrate personalized feedback in real-time, dynamically adapting the virtual environment not only based on the student's responses (as in traditional systems) but also based on their implicit physiological and emotional reactions during the immersive experience.

Simultaneously, sequential architectures such as Recurrent Neural Networks (RNNs) and their LSTM variants, developed for machine translation and long-term context maintenance (Hochreiter, S., & Schmidhuber, J., 1997), not only laid the groundwork for the first chatbots but were also essential for the temporal analysis of behavioral data. Only after having solved these recognition and analysis challenges, parallel to the increasing effectiveness of Attention Mechanisms (Bahdanau, D., 2014) and the invention of the Transformer architecture (Vaswani, A. et al., 2017), which definitively unlocked realistic text generation and the superability of the Turing Test, did AI applications become fully mature to provide both the conversational intelligence of avatars and the systems. AI has for years provided the necessary backbone to make e-learning scalable and reliable. The use of ML and DL algorithms has been crucial for the dynamic management of networks (5G/6G), enabling automatic resource allocation, traffic optimization (traffic offloading), and the prevention of cyberattacks. These are not just simple technical requirements, but the *conditio sine qua non* for the online learning experience to take place without interruption and on a large scale (Bilotti, U. et al., 2023). The Metaverse, with its high demand for bandwidth and low latency to support immersion, merely inherits and exacerbates these infrastructural needs, relying even more intensely on the predictive and optimizing capabilities of AI to maintain the fluidity and persistence of virtual worlds.

Artificial Intelligence to building Metaverse

The future of the Metaverse is inextricably linked to the advancements and innovations in Artificial Intelligence, which is destined to play a crucial role in defining its trajectory. A fundamental objective of the Metaverse is its ambition to be shared and participatory,

aiming for a democratic construction accessible to all users. However, the acquisition of the technical skills necessary to create even the simplest digital content initially represents a barrier for novices. Although economic pressures may be lower in the educational sector, the need for a critical analysis of content to ensure inclusive usage, which is the central reason for employing the Metaverse in education, remains paramount. Fortunately, the range of tools based on AI has increased exponentially, drastically reducing not only the actual time required for the creation of a digital asset but, crucially, the time needed to acquire the necessary skills and knowledge. This phase has rapidly elevated Generative AI (GAI) to the role of a catalyst, as its capabilities for dynamic personalization and data analysis are recognized as the only means to make virtual environments persistent, realistic, and adaptive, overcoming the rigidity of static simulations.

Generative AI is, by definition, the branch of artificial intelligence that focuses on creating new and original content, such as images, text, music, or virtual environments, playing a crucial role in the evolution of the Metaverse and in enhancing user engagement, for example through the procedural generation of entire virtual worlds (Chamola, V. et al., 2023). The functionalities of Generative AI are essentially organized into four macro-domains:

1. **Text Generation:** In this domain, models like GPT power chatbot interactions with dynamic conversations and contextually relevant responses, serving as personalized virtual assistants. Furthermore, they facilitate interactive narrative where the story evolves based on user choices, as in virtual role-playing games, and sequence-to-sequence models like GNMT ensure real-time language translation, guaranteeing seamless communication across different languages. Finally, AI generates dynamic dialogues and quests for Non-Player Characters (NPCs), adapting the game progression to the player's actions in open-world environments.
2. **Image Generation:** Generative AI contributes to the visual richness of the Metaverse: models like MidJourney use GANs (Generative Adversarial Networks) to create visual effects and smooth transitions between images, enhancing immersion, and assist designers in creating virtual fashion, accessories, and wardrobes,

supporting immersive commerce. Models like DALL-E and CRAIYON enable artistic generation, filling virtual art galleries with unique masterpieces, while others like NightCafe generate high-quality landscapes and scenery, crucial for creating immersive and credible virtual worlds.

3. **Video Generation:** Leveraging the power of advanced machine learning, AI unlocks the potential for the creation of dynamic video content, with models like Flicki and Runway generating videos based on user-defined styles. This includes the synthesis of realistic virtual actors (such as with Hour One and Rephrase.ai) for simulations and films, the animation of personalized avatars using text or voice inputs for customer assistance, and the complete automation of video production processes, for instance, for personalized virtual marketing campaigns.
4. **3D Object Generation:** This domain offers significant potential for enhancing the virtual universe by facilitating the modeling and creation of complex virtual objects (such as buildings, vehicles, and furniture) by users, supporting architectural design and asset libraries. AI also improves the detection of 3D objects in augmented reality environments and allows precise control over artistic design, enabling the personalization and creation of sculptures and virtual objects with enhanced detail.

The practical use of Generative AI in these four sectors verifies the idea of a participatory and democratic Metaverse. Generative AI is making the Metaverse a more concrete reality by dramatically decreasing the hitherto high barrier of specialised technical skills necessary for content production. The theoretical foundation and technology capabilities are now well established, indicating the viability of these AI-driven opportunities. The next step in comprehending this change is to look at tangible manifestations of the Metaverse rather than just the conceptual function of AI. The following discussion will focus on the Metaverse's primary prototypes and realised platforms, comparing the original literary and design ideals to the technological and economic reality of already functioning virtual worlds.

From Science Fiction to Precursors: The Birth of the Conceptual Metaverse

The concept of the Metaverse has deep roots in science fiction, first coined by Neal Stephenson in his novel "Snow Crash" (1992). The work not only gave the name to the idea but also defined its essential distinctiveness: the shared 3D interface and the application of rigid laws governing virtual real estate and ownership, managed by private corporations. Subsequently, cinema further crystallized this vision, particularly with works like "The Matrix" in 1999, whose unique feature lies in the exploration of the philosophical and sensory implications of a virtual world indistinguishable from physical reality, emphasizing themes of total immersion and dissociation from corporeality, all themes anticipated by Baudrillard, J. (1994). More recently, Ernest Cline's novel, "Ready Player One" (2011), offered the most vivid and culturally pervasive example of the modern Metaverse (named OASIS), whose defining characteristic is the all-encompassing integration of education, economy, and entertainment within a single persistent and accessible environment, with a strong emphasis on interaction via haptic devices and personalized avatars. Such conceptual suggestions stimulated the realization of early prototypes, which we now consider direct precursors of the Metaverse: starting in the 1990s, experiments like Active Worlds (1995) explored virtual land ownership, while platforms such as Second Life (2003) offered one of the first persistent digital social environments, oriented toward user economies, despite being limited to the two-dimensional screen. Massively multiplayer role-playing games like World of Warcraft (2004) codified the formation of complex social communities and internal digital economies, while platforms such as Roblox (2006) and Minecraft (2011) demonstrated the potential of User-Generated Content (UGC) and the creation of Metaverses as genuine development infrastructures. Today, the landscape is dominated by projects like Decentraland and The Sandbox, which combine the science fiction vision of a shared, accessible 3D space with the technological capabilities enabled by VR/AR and Blockchain.

Onlife existence and Educational Metaverse

Thinking of our existence as the sum of the events we experience cannot be considered an exhaustive definition. Our condition is, in fact, constantly evolving through not only the events we live but also the contexts in which we are immersed. Both events and contexts certainly allow for a discrete interpretation, yet this remains only an approximation of a continuous process. The learning process undeniably inherits this attribute of continuity, as widely argued within pedagogical thought.

This concept of continuous and inseparable evolving existence finds profound resonance in the reflections of Luciano Floridi on the notion of *onlife* existence (Floridi, L., 2015). Floridi maintains that the traditional distinction between online life and offline life is now obsolete, as digital technologies have become so pervasive that they merge the two states into a single reality: the onlife. Our experience is no longer limited to an alternation between the digital (the context) and the physical (the event), but is rather a hybrid reality where the boundary between being connected and being disconnected vanishes. This suggests that our condition, both as individuals and as learners, can no longer be reduced to a discrete set of "real" or "virtual" moments, but develops instead in a continuous and interconnected flow, precisely as the text defines it: a condition in continuous evolution through events and contexts.

The approach of seamless learning, introduced by Chan, T. W. et al. (2006), already highlighted two strong independencies of the learning process, which can occur in places and times other than those typical of schooling, university, or training, making it potentially a non-discrete but continuous phenomenon. Unbound from the dimensions of space and time, two other attributes that directly foster continuity gain significance.

The enactive attribute allows for the relocation of the learning setting within the student-teacher relational context; instructional action must therefore be designed for a system of individuals who enter into structural coupling and grow together, mutually imparting and receiving relational and cognitive modifications (Coin, 2013). Furthermore, the emulative

attribute highlights the centrality of the act in the teaching-learning process, as it is a cognitive anticipation of an action oriented toward a specific goal (Sibilio, 2014).

However, seamless learning should not only be understood as potential learning everywhere and anytime, the use of which could already be realized by a sterile online didactics, but also, and especially, in an enactive formula, therefore coordinated with the environment in which it takes place.

The characteristics of the proposed framework open the educational offering to a multitude of scenarios, each defined by a specific technological proposal that moves beyond the traditional boundaries of the classroom. Historically, many technologies developed exclusively for educational purposes have proven to be less "resilient" than broader, multi-purpose technological tools, a limitation that clearly emerges when viewed from an extended temporal perspective. An emblematic example in Italy is the Interactive Whiteboard (IW); despite its widespread distribution via ministerial plans, its use has often remained confined within the physical and temporal boundaries of the school building, frequently devolving into a mere digital support for traditional, transmissive, and teacher-centered lectures (Kennewell, 2004; Higgins et al., 2005; Conenna, G., 2013). Conversely, the introduction of 3D printing presents a fundamentally different paradigm that aligns with the principles of Ubiquitous Learning. Although it shares the obstacles of cost and physical localization with the IW, 3D printing enables a form of didactic continuity that transcends school walls because the core of the learning process is shifted from the hardware to the digital workflow. Thanks to accessible modeling software like Tinkercad, students can engage with the preparatory and creative stages, the digital model, at any time and in any place, relegating physical printing to a final, localized phase of the process.

Transposing these reflections to our specific case, the various Extended Reality (XR) technologies facilitate an equally vast spectrum of use, whereby the classification of Mixed Reality (MR) technologies proposed by Milgram, P., & Kishino, F. (1994) extends beyond a mere taxonomy to offer the pedagogical community a critical tool for reflecting on the points of discontinuity within educational designs that fail to address the virtual components of our

existence—components that have become indispensable as they are complementary to their physical counterparts. Given the characteristics analyzed within the theoretical framework of seamless learning, the Educational Metaverse undoubtedly emerges as an environment that can naturally serve as a model; indeed, while Head-Mounted Displays (HMDs) typically remain localized tools due to the significant investment required, mobile applications and desktop edugames guarantee external accessibility. Through the multi-platform framework of the Educational Metaverse, it becomes possible to "stitch together" these points of discontinuity, thereby creating a fluid learning ecosystem that bridges the gap between school and home environments. Moreover, the Metaverse's strong technical trajectory toward immersion and shareability results in an educational offering that is coherent with both enactive and emulative attributes, as current VR and AR headsets—through iterations that are physiologically more consistent than traditional hardware components—can promote the motor involvement necessary for the construction of meaning, while simultaneously, joint participation in the content creation process can strengthen the relational structure among those sharing the context.

Challenges to the Educational Metaverse: Security, Accessibility, and the Generational Gap

Despite the enthusiasm surrounding the current Metaverse ecosystem, it presents significant accessibility shortcomings that severely limit the participation of individuals with disabilities, posing an urgent need for targeted intervention by developers and industry stakeholders. The emergence of new concepts such as computer simulation, artificial reality, virtual environments, augmented reality, and cyberspace (Biocca, Levy, 1995) defines the future of technology, simultaneously bringing to light both the potential and the problems of accessibility and design (Riva, Morganti, 2012). The problems manifest at the level of informational accessibility, where the lack of adequate alternative textual descriptions and digital markup hinders understanding and orientation for users who rely on assistive technologies, particularly for people with visual disabilities. The effects of the Metaverse on health: Unbalanced visual information might cause virtual motion sickness and physical

weariness. Headset weight, mobility issues, and hygiene problems represent further concerns. Adverse effects may include moderate motion sickness, brief convulsions, dizziness, and eye tiredness. Maintaining cognitive homeostasis and stability is essential for providing quality Metaverse services. Furthermore, the immersive experience, the foundation of the Metaverse, is compromised by profound motor and mobility barriers, as standard interaction based on virtual reality controllers often requires precise actions or a physical commitment that is impractical for those with motor limitations (such as wheelchair users). Added to these problems are the inadequacies in the user interface, which often impose rigid interaction requirements, failing to account for the variability of individual motor capabilities. To ensure an inclusive experience that respects user identity, platforms must also allow for advanced avatar personalization that permits users with disabilities to choose if and how to represent their condition. Although solution proposals exist, such as alternative input methods (based on voice, gaze, biosensors) and tactile and acoustic feedback (like canetrollers or haptic gloves), it is imperative that the industry commits concretely to implementing these solutions and extending accessibility to all types of disabilities, preventing the risk that the Metaverse becomes a new space of digital exclusion (Di Dario, D. et al., 2024). Ensuring consistency with reality and building trust are fundamental, which is essential for effective interpersonal relationships and organizational performance. Establishing standards in these areas is crucial for defining the formation, appraisal, and use of legal systems. Online bullying, abuse, adultery, and dishonesty in relationships can worsen when users mistake virtual interactions and information exchanges for genuine ones (Chamola, V. et al., 2025). Network Security is at risk because attacks against Head-Mounted Display (HMD) systems might temporarily render users powerless while overlaying or changing the pictures inside their field of view. As a result, security issues for immersive educational systems grow. Another issue with the Metaverse is its reliance on specialized hardware. The Metaverse focuses on securing external digital devices, such as VR headsets, which are vulnerable to hackers (Kaddoura, S., & Al Hussein, F., 2023). Future wearable technology, including headsets, has the potential to capture sensitive data. Protecting intellectual property may be problematic when individuals and organizations are engaged in both the real world and the Metaverse. Managing digital rights and copyrights for learning materials may be challenging in the heterogeneous world of

online education and on platforms. The current digital copyright management system utilizes private and public blockchains to facilitate the sharing of multimedia educational materials. It safeguards the privacy of educational materials, making it an advantageous choice for Metaverse platforms (Guo, J. et al., 2020). The "Netiquette" emphasizes the need for clear communication when asking inquiries and offering courteous and protective responses. Ethical online behavior is crucial in Metaverse-based instructional programs for high schools, universities, and corporations. Netiquette is fundamental in online learning because ICT have altered social conventions and communication styles. To prevent cyberbullying and fraud on social media, students and teachers should receive adequate netiquette education and guidance (Soler-Costa, R. et al., 2021). The generation difference represents a difficulty, since younger people are more open to new possibilities. Personalized design and technology-based controls can solve these challenges. Providers and software businesses may address these issues. To tackle such issues, research groups should hold constructive discussions and collaborate to find solutions. The implementation of the Metaverse for educational services presents ethical concerns, including privacy, digital citizenship, network security, copyright, netiquette, and adherence to norms. Considerations include medium selection, Metaverse sustainability, hardware and software restrictions, and bridging the generation gap. The problem of the generational gap is exacerbated in the case of the Educational Metaverse for at least two main structural reasons. The first is linked to the inheritance of a strongly hierarchical structure, typical of the scholastic conception following the industrial revolution and characterized by a one-to-many teacher-student relationship model. Despite the pedagogical proposals consolidated over time, such as mastery learning, the flipped classroom, and peer-to-peer activities, and despite the commitment shown by nations and teachers through a wide offer of specialization and refresher courses aimed at the goal of inclusion, formal training contexts maintain a physiological age asymmetry between adult educators and young learners. This generational disparity, which implies a different familiarity with digital tools, would inevitably be inherited in the Metaverse context. The second reason, however, is intrinsic to the technology itself: it can potentially be proposed to all students, but it cannot be "imposed" on teachers. Consequently, the success and effective integration of immersive learning platforms depend crucially on the willingness and preparation of the teaching staff

in adopting these innovative tools, a factor that often creates a significant disparity between the technological potential offered and the actual didactic application in the classroom. Indeed, the use of technology in education is not exempt from obstacles: external barriers often refer to the lack of adequate resources for teachers, while internal barriers concern teachers' attitudes, their resistance, and their perception in utilizing technology for educational purposes (Ertmer et al., 1999; Sugar et al. 2004). In this regard, Burch and Mohammed (2019) emphasize how the widespread use of technologies requires constant support through continuous professional development activities, particularly in the educational sector. In confirmation of this, it is reiterated that the conscious use of technologies is now one of the key competencies of the teaching professional figure. To ensure an effective teaching-learning process, teachers are called not only to use technologies critically and creatively but also to design engaging didactics that take into account the multiple learning needs of all students (Khukalenko et al. 2022).

Bibliografia (Educational Metaverse)

Back, A. (2002). Hashcash-a denial of service counter-measure.

Bahdanau, D. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Baudrillard, J. (1994). *Simulacra and simulation*. University of Michigan press.

Berners-Lee, T. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper San Francisco.

Bilotti, U., Di Dario, D., Palomba, F., Gravino, C., & Sibilio, M. (2023, January). Machine learning for educational metaverse: How far are we?. In *2023 IEEE international conference on consumer electronics (ICCE)* (pp. 01-02). IEEE.

Bilotti, U., Schiavo, F., Tore, P. A. D., & Nappi, M. (2024, January). Peek the Edu-Metaverse: From an Educational Framework to the AI Challenges for Biometrics. In *Workshop on Artificial Intelligence with and for Learning Sciences: Past, Present, and Future Horizons* (pp. 35-44). Cham: Springer Nature Switzerland.

Biocca, F., & Levy, M.R. (Eds.). (1995). Communication in the Age of Virtual Reality (1st ed.). Routledge.<https://doi.org/10.4324/9781410603128>

Blanchard, C., Burgess, S., Harvill, Y., Lanier, J., Lasko, A., Oberman, M., & Teitel, M. (1990, February). Reality built for two: a virtual reality tool. In Proceedings of the 1990 symposium on Interactive 3D graphics (pp. 35-36).

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1), 210-230.

Burch, Z. A., & Mohammed, S. (2019). Exploring faculty perceptions about classroom technology integration and acceptance: A literature review. *International Journal of Research in Education and Science (IJRES)*, 5(2), 722-729

Chamola, V., Bansal, G., Das, T. K., Hassija, V., Sai, S., Wang, J., ... & Niyato, D. (2024). Beyond Reality: The Pivotal Role of Generative AI in the Metaverse. *IEEE Internet of Things Magazine*, 7(4), 126-135.

Chamola, V., Peelan, M. S., Mittal, U., Hassija, V., Singh, A., Pareek, R., ... & Brown, D. J. (2025). Metaverse for education: Developments, challenges, and future direction. *Computer Applications in Engineering Education*, 33(3), e70018.

Chan T. W., Roschelle J., Hsi S., Kinshuk, Sharples, M., Brown, T., ... & Hoppe, U. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. *Research and Practice in Technology Enhanced Learning*, 1(01), 3-29.

Chaum, D. (1983, August). Blind signatures for untraceable payments. In *Advances in Cryptology: Proceedings of Crypto 82* (pp. 199-203). Boston, MA: Springer US.

Cline, E. (2011). *Ready player one*. Ballantine Books.

Conenna, G. (2013). La motivazione all'apprendimento e le tecnologie. Una ricerca condotta con allievi di 14-15 anni sull'uso della Lavagna Interattiva Multimediale-LIM.

Coin F. (2013). Didattica enattiva: cos'è e cosa può fare. *Formazione & insegnamento*, 11(4), 127-134.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6), 64-73.

Di Dario, D., Sellitto, G., Pentangelo, V., Fede, M. L., & Ferrucci, F. (2024). Breaking Barriers in the Metaverse: A Comprehensive Exploration of Accessibility for Users with Disabilities. In *Workshop on Artificial Intelligence with and for Learning Sciences: Past, Present, and Future Horizons* (pp. 45-55). Springer, Cham.

Ertmer, P. A., Addison, P., Lane, M., Ross, E., & Woods, D. (1999). Examining teachers' beliefs about the role of technology in the elementary classroom. *Journal of Research on Computing in Education*, 32(1/2), 54-72

Floridi, L. (2015). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Springer Briefs in Philosophy.

Guo, J., Li, C., Zhang, G., Sun, Y., & Bie, R. (2020). Blockchain-enabled digital rights management for multimedia resources of online education. *Multimedia Tools and Applications*, 79(15), 9735-9755.

Haber, S., & Stornetta, W. S. (1990, August). How to time-stamp a digital document. In *Conference on the Theory and Application of Cryptography* (pp. 437-455). Berlin, Heidelberg: Springer Berlin Heidelberg.

Higgins S., Falzon C., Hall I., Moseley D., Smith F., Smith H. e Wall K. (2005), Embedding ICT in the Literacy and Numeracy Strategies Final Report, Newcastle, University of Newcastle.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780

Howlett, E. M. (1983). U.S. Patent No. 4,406,532. Washington, DC: U.S. Patent and Trademark Office

Kaddoura, S., & Al Hussein, F. (2023). The rising trend of Metaverse in education: Challenges, opportunities, and ethical considerations. *PeerJ Computer Science*, 9, e1252.

Kennewell S. (2006), Reflections on the interactive whiteboard phenomenon: a synthesis of research from the UK, The Australian Association for Research in Education. <http://www.aare.edu.au/o6pap/ken06138.pdf>.

Khukalenko, I.S., Kaplan-Rakowski, R., An, Y. et al. Teachers' perceptions of using virtual reality technology in classrooms: A large-scale survey. *Educ Inf Technol* 27, 11591–11613 (2022). <https://doi.org/10.1007/s10639-022-11061-0>.

Jerald, J. (2016). *The VR Book: Human-centered design for virtual reality* Morgan & Claypool Publisher.

Jin, S., Huang, J., & Zhong, Z. (2024). Application of immersive technologies in primary and secondary education. *Frontiers of Digital Education*, 1(2), 142-152.

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., ... & Wolff, S. (2009). A brief history of the Internet. *ACM SIGCOMM computer communication review*, 39(5), 22-31.

Nakamoto, S., (2007). Bitcoin: A peer-to-peer electronic cash system. 2008.

Mazhar, A. A., & Al Rifaee, M. M. (2023, August). A Systematic Review of the use of Virtual Reality in Education. In 2023 International Conference on information technology (ICIT) (pp. 422-427). IEEE.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12), 1321-1329.

Mueller-Csernetzky, P., Malakhatka, E., Thuvander, L., Kragic, D., & Kabo, J. (2025). Technological Frontiers in Education: Exploring the Impact of AI and Immersive Learning. In *Human-Technology Interaction: Interdisciplinary Approaches and Perspectives* (pp. 291-328). Cham: Springer Nature Switzerland.

Riva, G. & Morganti, F. (2012). *Conoscenza, comunicazione e tecnologia: Aspetti cognitivi della realtà virtuale*. Italia: LED Edizioni Universitarie

Roberts, L. G. (1966, November). The lincoln wand. In Proceedings of the November 7-10, 1966, fall joint computer conference (pp. 223-227).

Sherman WR, Craig AB (2019) Understanding virtual reality: interface, application, and design, 2nd edn. Morgan Kaufmann, Cambridge

Sibilio M. (2014). *La didattica semplice*. Napoli: Liguori.

Soler-Costa, R., Lafarga-Ostáriz, P., Mauri-Medrano, M., & Moreno-Guerrero, A. J. (2021). Netiquette: Ethic, education, and behavior on internet—a systematic literature review. *International journal of environmental research and public health*, 18(3), 1212

Stephenson, N.: Snow Crash. Bantam Books, New York, NY, USA (1992)

Sugar, W. C., Crawley, F., & Fine, B. (2004). Examining teachers' decisions to adopt new technology. *Educational Technology & Society*, 7(4), 201–213.

Sutherland, I. E. (1968, December). A head-mounted three dimensional display. In Proceedings of the December 9-11, 1968, fall joint computer conference, part I (pp. 757-764).

Turing, A. M. (2007). Computing machinery and intelligence. In *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 23-65). Dordrecht: Springer Netherlands.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Vickers, D. L. (1972). Sorcerer's apprentice: head-mounted display and wand. The University of Utah

Digital for inclusive interaction

The role of reading in the Educational Metaverse

For a long time, many communicative technologies were primarily designed as entities toward which a user directs a message, rather than as tools facilitating messages between various users (Gunkel, 2012). Progressively, therefore, the ergonomic paradigm has combined with the communicative one, transitioning from the creation of components capable of efficiently transmitting encoded commands to the creation of tools capable of actively participating in the construction of knowledge. Artificial Intelligence, particularly through modern chatbots, could establish a bidirectional form of communication, significantly contributing to Human-Computer Interaction in multiple areas of interest. The introduction of such tools in educational design is undoubtedly a central theme of pedagogical research, and several experiments have confirmed both the potential (Zappalà et al., 2024) and the limitations (Di Tore et al., 2024).

This necessity for active, participatory tools must be integrated with the sensory reality of the learning environment. In line with the paradigm of inclusion and the goal of seamless learning, which aims to dissolve the distinction between 'real' and 'virtual' learning, the efficacy of the Educational Metaverse is intrinsically linked to its capacity to guarantee an optimal sensory and cognitive experience for every user. The shift from traditional 2D interfaces to the 3D and VR environments of the Metaverse imposes new functional

challenges, where the readability of text plays a primary role. In these spaces, where depth, viewing angle, and screen resolution (especially in headsets) can vary drastically, the effectiveness of this immersion can be easily compromised by design elements that introduce cognitive friction.

While on one hand the evolutionary pressure derived from technological progress requires the adoption of new tools to facilitate human activity, on the other hand, the cognitive load resulting from the use of these new technologies must not risk producing undesired or counterproductive effects. In this framework, Executive Functions (EFs) are positioned at an intermediate level between the exploitation of cognitive resources and the performance of the motor operations necessary to finalize a precise skill. Despite their multifactorial nature, which to this day prevents the existence of an exhaustive theoretical model universally validated by the scientific community, their involvement in the teaching-learning process is firmly established. According to the model proposed by Diamond (2013), it is possible to classify executive functions into three main groups of skills: those related to inhibitory control, the types of working memory (both verbal and visuo-spatial), and cognitive flexibility, understood as the capacity for shifting between mental sets. The development of these skills constitutes, in fact, the possibility of acquiring the elementary components that allow for the construction of complex abilities, such as reading, across an extremely high range of applications. One need only consider the highly specific micro-skills cultivated during repeated sessions of a given video game; the virtual environment of the game provides a context in which, generally speaking, performance tends to increase alongside the involvement of EFs, thereby leading to improved attentional control during gameplay (Montani et al., 2014). Based on these considerations, having identified these common constituents for the development of differentiated complex skills, one can determine transversal design elements for diverse technologies based on these shared executive function pillars (Diamond, A., & Ling, D. S., 2016).

Concerning the textual component, which remains crucial for meaning-making, this translates into the imperative to adopt high-legibility fonts. The immersive environment and the typical prolonged visual exposure of the Metaverse amplify the importance of

typographic tools that support the continuity of attention and reduce cognitive load and visual fatigue. The adoption of such fonts is not a mere aesthetic detail, but a critical design decision that ensures textual content does not become an obstacle, but a fluid support for learning, consistent with the need for universal accessibility.

Toward an unified model for reading process

One of the most widely recognized models in the literature is undoubtedly that of Coltheart (1981), where the reading process is broken down into main sub-phases, which are then explored in depth. Multiple examples are also considered, allowing the authors to postulate the existence of two distinct pathways: one for reading words and another for reading sequences of letters that do not form words, called non-words. Subsequently, an updated model successfully highlighted the two different types of reading processes (Coltheart, M. et al. 2001; Ziegler J.C. et al., 2008), further elaborating on the sub-phases that had been identified and connected in the work by Job and Sartori (1984). The dual-route model theorizes the existence of two types of processes for reading, thus attempting to decompose the act of reading into sequential cognitive sub-phases. Specifically, the reading process can occur by following two different sequences:

- Lexical-semantic-direct route: In this pathway, the word is recognized as a graphic representation stored in our orthographic input lexicon. A specific meaning is then associated with it, followed by the retrieval of its phonetic counterpart, which is finally sent to the articulatory system, enabling its pronunciation.
- Phonological-sublexical-indirect route: Here, the word is first broken down into minimal units. It is then reconstructed through grapheme-phoneme conversion according to the language's orthographic rules, independently of its meaning.

According to this model, the acquisition of new words must necessarily occur via the phonological-sublexical-indirect route. Consequently, the grapheme-phoneme conversion

process plays a crucial role during the initial stages of acquiring reading proficiency (Share, D.L., 1995; Franceschini, S. et al., 2015).

Once the primary intervention process has been identified, the next step is to understand the optimal developmental period for educational proposals. In the work of Frith (1986) three reading strategies are initially identified which give their name to three successive stages:

- **Logographic stage:** During this stage, the learner recognises words as simple images, even if composed of different graphic elements. However, they cannot discern that the letter 'c' in "cat" is the same as in "cheese."
- **Alphabetic stage:** During this stage, the words are constructed by the learner following an ordered letter-sound by letter-sound decomposition.
- **Orthographic stage:** In this stage, word reconstruction primarily occurs through the conjunction of morphemic units. Therefore, groups of letters that form a word are assigned phonemes that do not result from simply combining the individual letters within those groups.

Ehri's work (1995), on the other hand, proposes four distinct phases for learning reading proficiency. The first, called the *pre-alphabetic phase*, relies solely on a graphic-phonological correspondence for reading. In this phase, a set of letters forming a word is read correctly only if the corresponding phonological production, associated with the visual image of that letter set, is correctly remembered. Otherwise, the word is either not read at all, or errors occur when different groups of letters create similar visual images. The next phase is distinguished from the first because the reader begins to use some alphabetic skills, which is why it is called the *partial alphabetic phase*. During this stage, the reader preliminarily identifies minimal phonemic units within a word that correspond to letter patterns produced similarly at an articulatory level. In the next stage, the *fully alphabetic phase*, beginners learn more about the spelling system, especially vowel spellings, enabling

them to connect every letter in a word to its corresponding sound. This allows their sight vocabulary to grow rapidly. The final stage, the *consolidated alphabetic phase*, is characterized by the efficient use of phoneme blending strategies derived from recurring letter patterns, such as prefixes, suffixes, syllables, and subsyllabic units. As readers advance into this phase, they use these larger units to speed up their learning of new sight words. In Beech's article (Beech, J.R., 2005) the two described models are compared, establishing a correspondence between the logographic stage and the pre-alphabetic phase, and between the orthographic stage and the consolidated-alphabetic phase. The alphabetic stage is thus decomposed into the partial-alphabetic and fully-alphabetic phases. Exploring the application of this technology across various writing systems, rather than limiting the analysis to a single one, has provided several additional points for reflection. This broadened perspective has led to a more comprehensive and nuanced understanding of the reading process.

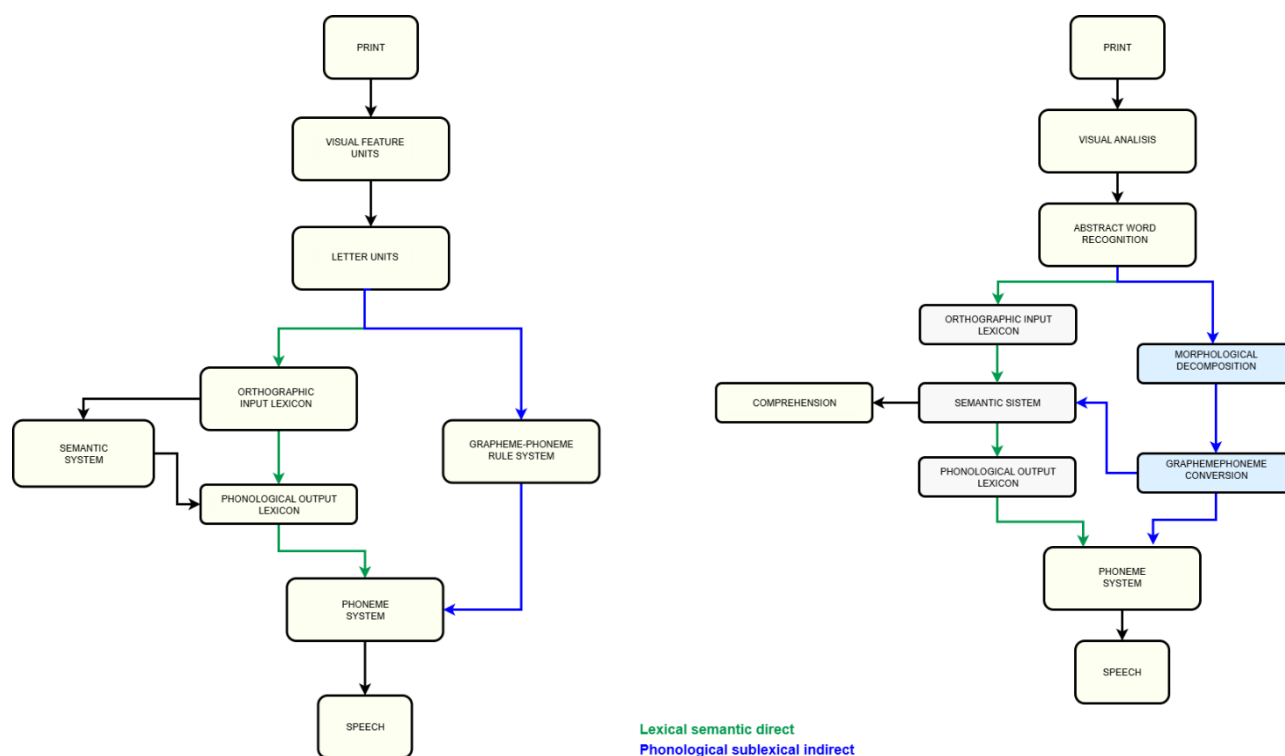


Figure 2 The model proposed by Coltheart on the left and the model with the integrations proposed by Job and Sartori on the right

A deeper analysis can be achieved by integrating the reflections on transparent and opaque languages. The correspondence between spoken and written language, which distinguishes a transparent language from an opaque one, encourages the use of strategies based on the syllabification process. In fact, in transparent languages like Italian and Spanish, there is a strong correspondence between how a word is broken down during writing and how it is read by sequentially sounding out simple phonemes. The word "casa," for example, can be read by sequentially sounding out the two syllables it's broken into in writing: "ca" and "sa." However, this process is more complex and can more easily lead to errors in opaque languages like English. For instance, the two words "record" (verb) and "record" (noun) are composed of the same syllables, "re" and "cord," but the pronunciation changes based on their function. During the fully-alphabetic phase in transparent languages, syllabification is often adopted as a support strategy for fluent reading. Later, upon reaching the consolidated alphabetic phase, the reader progressively becomes independent of this strategy. These insights are further clarified and synthesized by Di Tore (2016) who significantly expanded upon Frith's model by integrating key findings from other researchers. Specifically, Di Tore incorporated observations on the lexical stage of reading from the work of Tressoldi, Vio and Scotti (1991) and he also brought in crucial insights regarding the differences between transparent and opaque languages, as highlighted by Stella (2011). The version of the model that results from this integrated approach is presented as follows:

- **Logographic phase:** During this phase, the student recognizes words solely as well-defined graphic elements or pictures, associating them with vocal productions.
- **Alphabetic phase:** In this phase, the student uses alphabetic knowledge and elementary phonemic construction rules for simple syllables, thereby being able to read both regular words and non-words.
- **Orthographic phase:** During this phase, the student acquires the orthographic rules necessary for reading irregular words.

- **Semantic-Lexical phase:** Here, the reading process can occur without the operation of syllabification

Thanks to the research conducted, our understanding of how the reading process occurs and the related acquisition of this skill has significantly improved. However, these advancements have a "side effect": they lead to new questions whose discussion necessarily requires a more comprehensive theoretical level of abstraction. The model in Figure 3 therefore attempts to unify the consolidated findings related to both the reading process and its progressive learning. The first crucial phase involves accessing a specialized mental database known as the orthographic input lexicon. This lexicon is essentially our personal dictionary of known word spellings, holding the visual representations of words we've encountered before. The subsequent step is a rapid checking process where the brain attempts to match the visual form of the word it is currently seeing against the entries already stored in this lexicon. This quick comparison is fundamental to determining if the word, as a familiar graphic or image, has been previously committed to memory. Following this phase, the process branches, proceeding either via the direct route, which develops during the logographic stage, or the indirect route, which emerges from the alphabetic stage onwards. The word decomposition sub-phase most clearly highlights the level of reading acquisition.

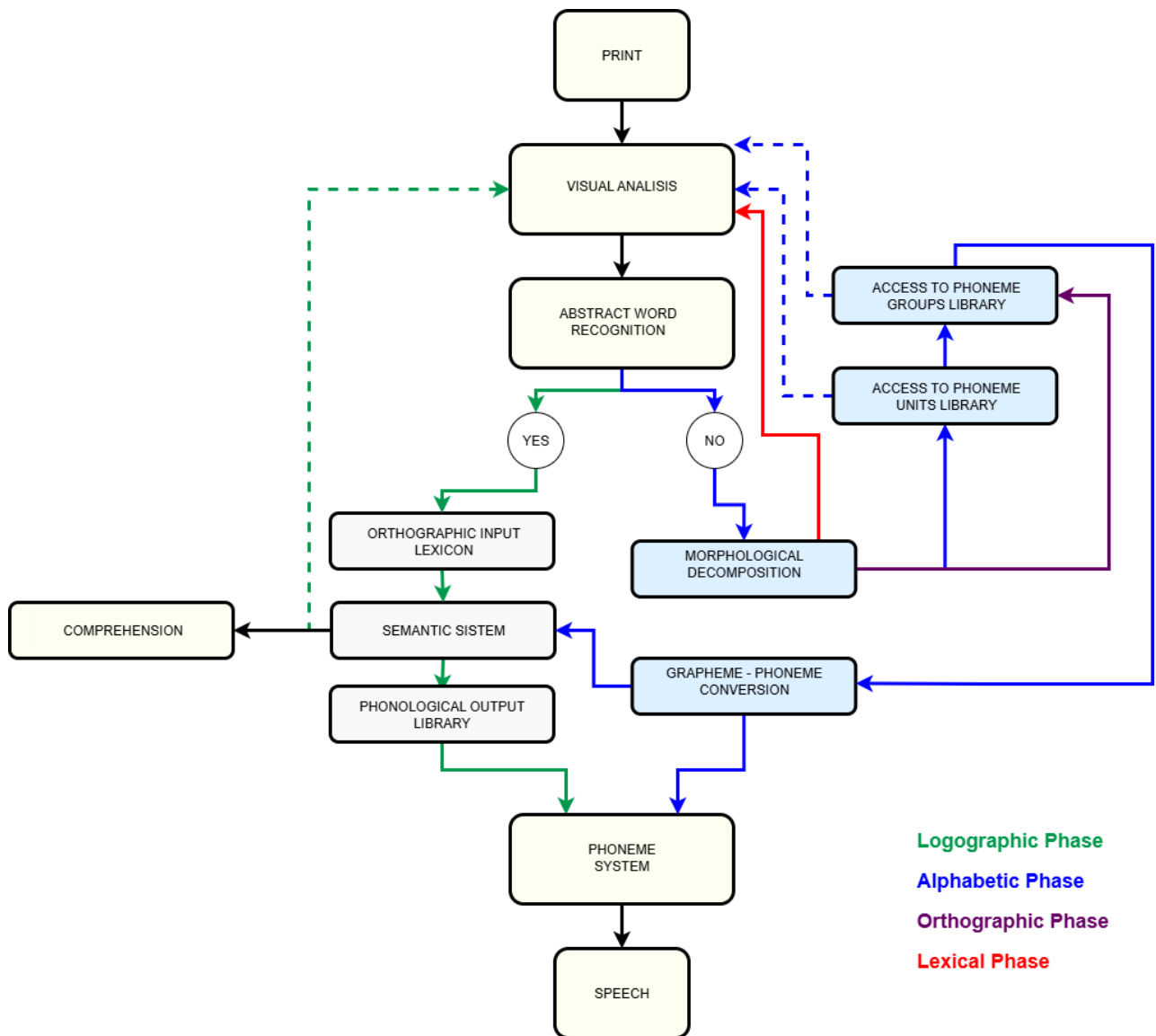


Figure 3. A proposal of a new model that organizes the cognitive processes functional to the act of reading. Specifically, it highlights the process patterns that are enabled during the different phases of learning to read.

Dyslexia, definitions and theories on core deficit

Difficulties in reading can have significant negative effects on both individuals and communities (Andresen, A., Monsrud, M.B., 2022). For example, the school dropout rate

among students with reading difficulties can reach up to 35%, nearly double the national average in many countries (Al-Lamki, L., 2012). Additionally, individuals with limited reading skills typically earn 30% to 42% less than their literate peers and have fewer opportunities for further education, limiting their economic advancement (Martinez, R. et al., 2012). Despite significant advances in understanding reading development and effective instructional strategies, many pupils globally still struggle with reading (Vaughn, S. et al., 2024). The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; APA, 2013) defines "Specific Learning Disorder with Impairment in Reading", often associated with dyslexia (Snowling, M., Hulme, C., 2024), as persistent difficulties with accurate or fluent word reading, poor decoding skills, and inadequate spelling, even after at least six months of targeted intervention. Dyslexia can only be diagnosed in individuals with a normal IQ who do not have socioeconomic issues, sensory deficits, or neurological damage that could explain their reading difficulties. A diagnosis is made when an individual's reading ability is significantly below that of their peers, a discrepancy measured by two key parameters: reading speed as the number of syllables read per second (or the seconds taken to read one syllable) and reading accuracy as the number and type of errors made during reading.

This discrepancy is identified through the use of standardized tests. In Italy, for instance, it is defined as "significant" when the subject's performance is more than two standard deviations below (or below the 5th percentile of) the mean for the reference population (Cornoldi, C., Tressoldi, P. E., & Perini, N., 2010). Because reading is so complex, building a tool to support it is equally challenging. To justify the design choices we'll present later, we must first discuss the leading theories about the core deficit of dyslexia.

Magnocellular Theory

Humans commonly tend to use their left hemisphere more than their right for complex reading tasks. However, this separation is not absolute; brain activity during less complex reading performances (which make up most daily reading) tends to be more evenly

distributed between the two hemispheres. Post-mortem studies on dyslexic brains have shown an absence of asymmetry in the planum temporale, which, based on previous observations, would be expected in non-dyslexic readers. Furthermore, researchers have observed anomalies of the cerebral cortex affecting preferentially (but not exclusively) the perisylvian regions of the left hemisphere. Additionally, the presence of cortical anomalies located asymmetrically and affecting inferior frontal and superior temporal regions predominantly on the left hemisphere has been noted (Galaburda, A.M. et al. 1985). A few years later, experiments by Eden (1996) showed that individuals with dyslexia exhibit reduced activation in specific visual areas of their brain's dorsal stream when presented with moving visual targets. Essentially, their brains don't respond to visual motion as robustly as those of non-dyslexic individuals. Thanks to these findings, Stein was able to formulate the magnocellular deficit theory (Stein, J., 2018). After leaving the occipital cortex, visual information splits into two pathways: the dorsal stream and the ventral stream (Ungerleider, L., Mishkin, M., 1982). The dorsal stream in particular, is devoted to controlling eye and limb movements and passes into the supramarginal and angular gyri in the posterior parietal cortex and this stream and is dominated by the work of magnocellular neurones. These cells are typically about 10% larger than other cells and are specifically involved in processing rapid changes, motion, and low-contrast information, also these cellules are less concerned with fine details or color.

Phonological Deficit Theory

Despite many common reading errors by dyslexics being interpretable as a consequence of a sensory-visual deficit, such as that proposed by the magnocellular deficit theory, an alternative theory was simultaneously developing that aimed to offer effective tools for early diagnosis of the disorder. One of the main issues leading to the development of other capable theories was the very regressive definition of the disorder. A dyslexic subject was considered so if their reading performance deviated significantly from a chosen standard. However, as various experiments showed, adopting a single standard couldn't be the definitive solution, as the proposed thresholds were themselves dependent on factors like different cultures,

languages, and schools systems (Snowling, M., 1998). Rather than investigating the number of errors that discriminate a dyslexic individual in a specific geographical area, it is more effective to examine the most prevalent types of errors across various geographical regions. The majority of common errors found in different experiments fall into the category of phonological encoding errors, defined as "the ability to use speech codes to represent information in the form of words and parts of words (Vellutino, F.R. et al., 2004).

Cerebellar Theory

According to certain researchers, phonological issues may occasionally be caused by magnocellular deficiencies, or distinct dyslexics may have both kinds of abnormalities at the same time. Research has looked for evidence to establish a more broad theory that can embrace the two ideas previously outlined because they are not necessarily mutually incompatible. However, recent theories suggest that phonological impairment alone does not fully explain the difficulties associated with dyslexia (Fella, A. et al. 2022; Zoccolotti, P., 2022). The Cerebellar Theory suggests that dyslexia is linked to a dysfunction in the cerebellum, a brain region primarily known for motor control and coordination. Several studies (Rae, C. et al., 2002; Ramus, F. et al., 2003) argue that an impaired cerebellum affects the automatization of skills, including those required for fluent reading, such as rapid letter-sound associations, precise eye movements, and the overall coordination of reading processes. This theory often points to co-occurring motor difficulties and problems with timing in dyslexic individuals.

Attentional Deficit Theory

The Attentional Deficit Theory suggests that dyslexia's core reading difficulties stem partly from problems with attention, including selective attention (ignoring distractions), sustained attention (maintaining focus), and efficient attention shifting between letters and

words. These weaknesses make reading harder, leading to slower speeds and reduced comprehension. Research, including a specific study mentioned, indicates that strong visuo-spatial attention is vital for phonological reading. While acknowledging the role of phonological processing, findings like the work of Facoetti (2006) that a purely phonological explanation for dyslexia is incomplete, as visuo-spatial attentional deficits independently contribute to reading difficulties in some dyslexic individuals.

Strategies to improve readability

The definition of an integrated model that try to summarize *how we read* and *how we learn to read* appears necessary to bridge the sub-phases of information retrieval with the automation strategies that learners develop throughout their growth. Analyzing the primary theories concerning the core deficit of dyslexia enables the association of specific error classes with each etiological hypothesis; however, although the phonological deficit theory remains the most widely accepted by the scientific community, the absence of an all-encompassing perspective currently precludes the design of inclusive technologies capable of responding universally to every manifestation of the disorder. Despite this complexity, the integration of the dual-route architecture with the chronological learning model suggests which decoding mechanisms can be facilitated through the use of high-readability fonts, such as OpenDyslexic. The adoption of such typographic solutions is based on the hypothesis that improved perception of individual glyphs and their respective graphemic clusters can actively support the fluent reading strategies that consolidate during the alphabetic and orthographic phases, thereby reducing the impact of visual-perceptual barriers.

The theories strengthened by years of conducted experiments have allowed the pedagogical community to formulate diverse teaching strategies to support reading skill acquisition. Consequently, teacher training has benefited from both a greater informational offering and a varied, proactive approach. In fact, once a special educational need is identified, it demands swift teaching action from educators, capable of deploying multiple solutions. To promote the well-being of dyslexic students, teaching proposals should encourage the

development of their reading proficiency by minimizing the amount of material to read and write. They should prioritize oral assessments, reduce assignments and study materials, and encourage the use of compensatory tools designed to avoid the tasks of decoding and manual writing, as well as activities directly connected to them (Cornoldi, C., Gioffrè, D., & Martini, A., 2013) In light of this, the production of digital textual material, by both teachers and students, cannot be uniform or standardized; instead, it must be adapted to the student's unique cognitive needs.

Recommended formatting

To understand how readable a text is, we have to look at several factors. This includes where the text appears, whether it is on paper or a screen, big or small (Huang, S.M., 2019) as well as the language it is written in and who the intended readers are. This is a topic that spans many scientific fields, and a lot of research shows just how vital text formatting is for making content easier to read. The goal is to make content accessible and inclusive, meaning the settings shouldn't slow down typical readers while still helping those with reading difficulties. It is crucial to understand that many formatting options (like font size, spacing, and kerning) impact the physical or digital space a text occupies. Therefore, experimenting to find the right balance is key. Overdoing any of these parameters could actually make reading harder by involving more cognitive effort. Teachers should first assess their class's needs to pick the best formatting. Here are the parameters they can adjust (Rello, L. Baeza-Yates, R., 2017):

- Font Size : Research suggests an effective range is between 12 and 18 points.
- Word Spacing: An effective way to improve spacing is by increasing the size of the "space" character. Since spaces separate words, this helps individual words stand out, making them easier to recognize.

- Line Spacing (Leading) : Generally, the most effective line spacing ranges from 1 to 2.5 cm.
- Alignment: Always use left alignment and avoid justified text. Justifying text automatically adjusts spacing to fit both margins, which can undo custom spacing changes, especially to the "space" character.
- Color Combination : The color combination of text and background is also important. While yellow text on a blue background or white text on black generally offers the highest recognition, these might be uncomfortable for typical readers.

Recommended fonts and fonts designed for dyslexic readers

Although formatting elements can often be adjusted with relative ease, selecting the most appropriate font or typeface remains a complex and largely unresolved challenge. A primary constraint in font selection is the structure of the writing system of the language in question. In logographic systems like Chinese, characters tend to be composed of intricate visual components. Interestingly, this complexity does not necessarily hinder reading; on the contrary, it can facilitate character recognition by increasing visual distinctiveness. However, studies have shown that for digital displays, particularly desktops, simpler character structures can enhance legibility (Dobres, J. et al., 2016). As a result, minimalist typefaces such as Hei are frequently adopted in experimental studies aimed at isolating the influence of stroke width and design features on reading performance (Liu, N. Yu, R., & Zhang, Y., 2016). Designing fonts for character-dense languages like Chinese or Japanese involves generating thousands of glyphs—often more than 6,000—which represents a significant design and computational burden. To address this, researchers have explored automated font creation techniques that rely on interpolation between existing typefaces (Lian, Z., Xiao, J., 2012) or on small sample sets used to synthesize new characters (Miyazaki, T. et al., 2019), offering more scalable and adaptive solutions. For writing systems that are not logographic—such as syllabic or alphabetic scripts—the challenges shift. In the

Sinhala script, for instance, character similarity can lead to confusion among readers. In response, studies have offered specific design recommendations to improve glyph distinctiveness and readability (Subasinghe, R., Samarawickrama, S., 2024). Similarly, in the case of the Devanagari script, eye-tracking techniques have been used to evaluate readability, measuring metrics such as the frequency and duration of visual fixations across different font designs (Ralekar, C. et al. 2018). For other alphabetic writing systems, such as Arabic, the matter of recommended fonts becomes more complex due to both intrinsic characteristics of the script—like its cursive nature and connective graphic elements—and cultural factors (Al Rowais, F., Wald, M., & Wills, G., 2013; Benmarrakchi, F. et al., 2017). The British Dyslexia Association recommends using fonts such as Arial, Comic Sans, or, alternatively, Verdana, Tahoma, Century Gothic, and Trebuchet (British Dyslexia Association, 2023). Various experiments conducted on languages using the Latin alphabet have confirmed the good readability of these fonts for English, French, Italian, and Spanish. Furthermore, even some alphabets with similar letters, like the Greek and Cyrillic alphabets, show positive results, particularly with Arial. Based on the literature, the choice of which font to use among available options is a problem addressed not only for languages like English but also for widely diffused linguistic variations such as Chinese or Arabic. Consequently, some common characteristics for an optimal font could be extrapolated by "averaging" preferred fonts or by trying to specifically address one or more of the error categories identified during various studies. However, despite the reflection potentially leading us to structure the problem as a case of reverse engineering, the "distillation" of effective features for a font designed as dyslexia-friendly hardly guarantees better performance than that of the currently recommended fonts. The OpenDyslexic font is a perfect example of a design crafted specifically for dyslexic readers. Its glyphs aren't focused on aesthetic appeal; instead, their shapes aim to reduce issues like reversal errors and crowding. Regardless of its actual effectiveness as a tool, OpenDyslexic highlights the critical need for such solutions to be freely available and accessible. This is not just for the end-users but also for everyone else involved in the teaching and learning process, especially teachers. OpenDyslexic's innovative approach has since inspired numerous similar fonts, not only for Neo-Latin languages sharing the same alphabet (Di Tore, S., Sibilio, M., 2015; Galiano, A.R.

et al. 2023) but also for other alphabets with visually similar letters (Bilotti, U., Todino, M.D., & Fella, A., 2023; Alexeeva, S., Dobrego, A., Zubov, V., 2020).



Figure 4. Examples of crowding effect in different language-writing systems

GreekDyslexic: Study and design for a dyslexic-friendly font

The Greek alphabet consists of 24 letters, each of which has an uppercase and lowercase variant. The Greek language can be considered a sufficiently transparent language (Protopapas and Vlahou, 2009), thus like Italian and Spanish, a language with a high frequency of grapheme-phoneme correspondence, put more simply, a language that "is read as it is written and vice-versa." The alphabet can be partitioned into four classes characterized by the vertical extension of the letters relative to the line. In particular, the letters "α", "ε", "η", "ι", "κ", "ν", "ο", "π", "ρ", "τ", "υ", "ω" belong to the standard class commonly defined by the "x-height" (Kahn and Lenk, 1998), the letters "β", "δ", "θ" belong to the ascender class, consisting of letters that extend above the standard box, the letters "γ", "λ", "μ", "ρ", "φ", "χ", "ψ" to the descender class, consisting of letters that extend below the standard box, thus below the line, and in addition to these common classes also for the Latin alphabet, it is possible to determine a fourth class consisting of the letters "ζ", "ξ"

which extend both above the box and below the line. The first characteristic chosen as the objective of the font design was to create a font with elementary shapes and without the graphic embellishments commonly called serifs in the individual glyphs. The shapes of the glyphs, referred to as *sans serifs*, on the one hand can decrease the *crowding* effect; however, even with the use of simple shapes, it is necessary to create easily distinguishable glyphs. For this reason, it was decided to start from the shapes of the Arial font, which are very simple and composed of the same basic elements (same round and stick shapes), and subsequently, the effort was made to characterize the individual glyphs. Next, an attempt was made to recover the composition structure of the Arial glyphs by extracting key lines called skeletons (Zhang and Suen, 1984) to serve as a first reference point for the design of the new glyphs (Figure 1). Subsequently, their contour was extracted from the individual glyphs, and half of the contour points were used to determine a second set of points by calculating their respective symmetrical points with respect to the skeleton points. In this way, a point cloud was created that approximates the main shapes of the Arial glyphs. The same procedure was then used to determine the point cloud of GreekDyslexic (Bilotti, U., Todino, M.D., & Fella, A., 2023), a Greek alphabet font created by attempting to emulate the glyphs of the OpenDyslexic font. This second reference point was chosen to address the problem of the reversal error, which is less studied compared to the Latin alphabet but, at least from a graphic point of view, also possible in the case of the Greek alphabet. Once the point cloud for the individual glyphs of both typefaces was identified, the Coherent Point Drift method (Myronenko and Song, 2010) was applied to locate an intermediate cloud between that of Arial and that of GreekDyslexic. The intermediate cloud was then chosen manually and used as a new starting point for the design of the new glyphs. With this method, the aim was to emulate the style of OpenDyslexic as much as possible; however, interventions were necessary in cooperation with people who use this alphabet daily, with the goal of homogenizing the thickening in the lower part between the letters and to respect the correct belonging to one of the four classes.

Translate this text:
**Η Λευκωσία είναι η πρωτεύουσα
της Κυπριακής Δημοκρατίας**

Figure 5. Examples of using OpenDyslexic and GreekDyslexic in the same sentence.

GreekDyslexic: Experimentation of the case study

The methodology of this experiment was based on the administration of an online questionnaire, designed to collect data on font preferences and perceived readability among a total of 98 Greek native-speaking participants, including subjects with and without reading difficulties. The objective was twofold: to investigate the information in an engaging manner and to analyze preferences across four different fonts—Times New Roman, Calibri, Arial, and GreekDyslexic—using a 5-point Likert scale (where 1 = "Strongly Disagree" and 5 = "Strongly Agree") for both artifact captions and mathematical formulas.

The first part of the questionnaire presented participants with captions related to three culturally significant artifacts (chosen for high ecological relevance and internal reliability): Michelangelo's *Creation of Adam*, a marble statue of Aphrodite from Soloi (1st century BC), and a copper ingot from Engomi, Cyprus (16th century BC).

The second part focused on the readability of mathematical formulas (the Pythagorean theorem, $\alpha^2 = \beta^2 + \gamma^2$, and the linear equation, $\psi = \omega\chi + \varphi$, to test the consistency of responses across a different textual format.

The third part collected essential demographic data, which revealed a female majority (81.6%, $n = 80$), and an age distribution that showed 35.1% ($n = 35$) between 18 and 22

years old, 22.7% ($n = 22$) between 23 and 25 years old, 9.3% ($n = 9$) between 26 and 29 years old, and the remaining 33% ($n = 32$) over 30 years old. Regarding dyslexia status, 73.5% ($n = 72$) of participants reported no official diagnosis, 19.4% ($n = 19$) had a diagnosis, while 7.1% ($n = 7$) were unsure of their status. Following a pilot study with four participants to evaluate clarity, all stimuli were presented at a fixed font size of 14 point, maintaining consistent line and character spacing in line with recommendations from the British Dyslexia Association (2023), although the texts and formulas were justified. Dissemination took place online via email and the University of Nicosia's LMS system, in collaboration with the Success Centre and the Department of Distance Learning, guaranteeing voluntary participation, informed consent, and ethical compliance due to the researchers' lack of direct access to sensitive data.

GreekDyslexic: Results and future research for HCI

The GreekDyslexic typeface consistently registered the lowest average scores across all evaluated tasks, as illustrated in Table 1, while still presenting standard deviation values analogous to those of the other fonts examined. This pattern of reduced scoring remained consistent even when the analysis was strictly confined to dyslexic readers (Figure 6).

The least favorable ratings emerged clearly in the first three questions, which focused on the artifact captions, showing average scores of 1.89, 2.05, and 2.11, with a prevalent modal value of 1. In contrast, the highest ratings were found in the final two questions, related to the mathematical formulas, with average scores of 2.95 and 3.05 and modal values of 5. The minimum scores attributed to GreekDyslexic were generally comparable to those of the serif font used for comparison, except in the third question, while the maximum scores were similar to those of the other typefaces, with the single exception of an outlier recorded in the first question.

	1: Adam	2: Aphrodite	3: Copper	4: Pythagorean Theorem	5: Linear equation
TimesNewRoman	3.50 (1.17) 3.11 (1.10)	3.49 (1.17) 3.00 (1.15)	3.39 (1.13) 3.36 (1.12)	4.14(1,12) 3.79 (1.51)	3.98 (1.24) 3.68 (1.53)
Calibri	3.91 (1.00) 3.52 (1.02)	3.85 (1.04) 3.53 (1.22)	3.84 (0.96) 3.57 (1.02)	4.22 (1,00) 3.94 (1.18)	4.17 (1.03) 4.00 (1.00)
Arial	4.24 (0.96) 3.84 (1.06)	4.27 (0.87) 4.16 (1.01)	4.26 (0.89) 3.94 (1.03)	4.17 (1.07) 3.58 (1.35)	4.27 (1.09) 3.95 (1.47)
GreekDyslexic	2.50 (1.36) 1.89 (1.15)	2.48 (1.39) 2.05 (1.35)	2.50 (1.42) 2.11 (1.45)	3.00 (1.52) 2.94 (1.72)	2.96 (1.41) 3.06 (1.54)

Figure 6. In the table the mean font scores across different captions (standard deviation in parentheses) calculated for the entire sample (values in black) and dyslexic readers only (values in blue)

Despite being overall the least preferred font, half of the participants still gave GreekDyslexic a total average rating of at least 2.5 out of 5. Furthermore, 25.8% of the total sample assigned it an overall score of 3.5 or higher. Focusing the analysis on the questions specifically related to the application of the typeface in mathematical expressions, the evaluations showed a significant improvement: 59.8% of participants rated it at least 2.5, and 44.3% rated it at least 3.5 out of 5.

To determine whether the choice of typeface had a statistically significant impact on the responses, two Analyses of Variance (ANOVA) were conducted on the average scores obtained by the four fonts across the five questionnaire tasks. The results demonstrated that the ratings attributed to the typefaces differed in a highly statistically significant manner in both participant groups (dyslexic and non-dyslexic) ($p < 0.001$ for both), confirming that the font used was an influencing factor. Subsequent *post-hoc* tests, performed to isolate specific preferences, revealed significant differences in all comparisons between GreekDyslexic and each of the other fonts, for both dyslexic and non-dyslexic readers. However, a key distinction emerged: while non-dyslexic participants expressed clear preferences among the standard typefaces (Arial, Calibri, and Times New Roman), dyslexic participants did not show statistically significant differences between these same fonts (as reported in Figure 7).

	TimesNewRoman	Calibri	Arial	GreekDyslexic
TimesNewRoman	- -	0.0015 0.2685	<0.0001 0.4057	< 0.0001 0.0274
Calibri	0.0015 0.2685	- -	0.0019 1.0000	< 0.0001 0.0023
Arial	< 0.0001 0.4057	0.0019 1.0000	- -	< 0.0001 0.0006
GreekDyslexic	< 0.0001 0.0274	< 0.0001 0.0023	< 0.0001 0.0006	- -

Figure 7. In the table the *p*-values of the font comparisons obtained from the post-hoc test (black values for non-dyslexics, blue values for dyslexics).

The findings of the study on typographic preference for the GreekDyslexic font raise significant questions regarding the effectiveness of current Human Computer Interaction (HCI) tools intended to support inclusion within the Educational Metaverse, particularly for readers with dyslexia. The font received consistently low preference ratings from both dyslexic and non-dyslexic participants, suggesting a misalignment with user expectations and reading comfort. This trend, which is consistent with prior research conducted in languages that use the Latin alphabet, indicates that users tend to favour familiar typefaces over designs specifically developed to support dyslexic reading.

Nevertheless, GreekDyslexic exhibited a gradual increase in the ratings assigned in the final stages of the questionnaire, indicating that extended exposure to the font may reduce the initial bias. This pattern suggests that dyslexic readers, who are typically more sensitive to readability issues, could potentially benefit from refined versions of the typeface that more closely reflect their actual preferences. Moreover, the higher ratings observed in mathematical tasks highlight that the readability of individual glyphs—when presented in isolation—may surpass that of continuous text. This finding reinforces the importance of typographic design that accounts for the varied contexts in which the font may be used.

The limitations of the study provide clear directions for future research. First, the small sample size and the imbalance between dyslexic and non-dyslexic participants restrict the generalizability of the results. Follow-up studies should include larger and more balanced

samples, as well as demographic controls such as age, educational background, and familiarity with digital reading. Additionally, the decision to compare GreekDyslexic exclusively with widely used typefaces such as Times New Roman, Calibri, and Arial may have introduced a familiarity bias. To mitigate this effect, future investigations should incorporate both less conventional fonts and a broader range of dyslexia-friendly typefaces. The structure of the questionnaire represents another critical limitation. The exclusive use of Likert-scale items, combined with the absence of open-ended questions and the lack of layouts adapted to each font, restricted the study's ability to capture subjective motivations and perceived difficulties. Integrating qualitative components—such as free-text comments, descriptions of encountered challenges, and participant-generated suggestions—would better inform the iterative development of GreekDyslexic. Expanding the variety of tested materials to include narrative passages, expository texts, and longer continuous prose would also enable a more realistic evaluation of the font's readability.

A further concern relates to the reliance on self-reported dyslexia status, which may be inaccurate. Future research should incorporate standardized screening tools to increase diagnostic reliability. Moreover, to overcome the inherent subjectivity of self-report measures, it is essential to integrate objective and neurocognitive metrics. Reading speed, accuracy, error rates, visual fatigue, eye-tracking data, and EEG-based indicators could provide meaningful insights into how the typeface influences visual processing and cognitive load, particularly for dyslexic readers.

In conclusion, the low preference reported for GreekDyslexic should not be interpreted as a failure of the typographic design itself, but rather as a starting point for refinement. The findings serve as a reminder for HCI designers working within the Metaverse: labelling a tool as “accessible” is insufficient unless its functional effectiveness and user acceptance are empirically verified through rigorous, multidimensional, and genuinely human-centred methodologies. Co-design processes with dyslexic users, cross-linguistic investigations, and evaluations conducted in real digital environments—such as e-books, educational applications, and dynamic text displays—can guide the development of future versions of GreekDyslexic that are more effective, more widely accepted, and genuinely inclusive.

Bibliography (Digital for Inclusive Interaction)

Alexeeva, S.; Dobrego, A.; Zubov, V. Towards the first dyslexic font in Russian. In Proceedings of the Language Resources and Evaluation Conference. European Language Resources Association (ELRA), 2020, pp. 9–14.

Al-Lamki, L. Dyslexia: Its impact on the individual, parents and society. Sultan Qaboos University Medical Journal 2012, 12, 269.

Al Rowais, F.; Wald, M.; Wills, G. An Arabic framework for dyslexia training tools. Journal of Research in Special Educational Needs **2013**.

Andresen, A.; Monsrud, M.B. Assessment of dyslexia—why, when, and with what? Scandinavian Journal of Educational Research 2022, 66, 1063–1075.

Beech, J.R. Ehri's model of phases of learning to read: a brief critique. Journal of research in reading 2005, 28, 50–58.

Benmarrakchi, F.; El Kafi, J.; Elhore, A.; Haie, S. Exploring the use of the ICT in supporting dyslexic students' preferred learning styles: A preliminary evaluation. Education and Information Technologies **2017**, 22, 2939–2957.

Bilotti, U.; Todino, M.D.; Fella, A. Implementation of Greek alphabet characters according to the OpenDyslexic standard and teacher's guide for font use. Journal of Inclusive Methodology and Technology in Learning and Teaching **2023**, 3.

British Dyslexia Association. Dyslexia Style Guide 2023, 2023.

Coltheart, M. Disorders of reading and their implications for models of normal reading. *Visible language* 1981, 15, 245.

Coltheart, M.; Rastle, K.; Perry, C.; Langdon, R.; Ziegler, J. DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review* 2001, 108, 204.

Cornoldi, C.; Tressoldi, P.E.; Perini, N.; et al. Valutare la rapidità e la correttezza della lettura di brani: nuove norme e alcune chiarificazioni per l'uso delle prove MT. *Dislessia* 2010, 7, 89–100.

Cornoldi, C.; Giofrè, D.; Martini, A. Problems in deriving Italian regional differences in intelligence from 2009 PISA data. *Intelligence* 2013, 41, 25–33

Di Tore, S.; Sibilio, M. L'invisibile tecnologia della parola: design e sviluppo di un font per dislessici. *Italian Journal of Educational Technology* **2015**, 23, 33–41.

Di Tore, S., Di Tore, P. A., Bilotti, U., & Sibilio, M. (2024, January). Inclusion: A Concept Too Young for Artificial Intelligence. In *Workshop on Artificial Intelligence with and for Learning 369 Sciences: Past, Present, and Future Horizons* (pp. 123-130). Cham: Springer Nature Switzerland

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*.

Dobres, J.; Chahine, N.; Reimer, B.; Gould, D.; Zhao, N. The effects of Chinese typeface design, stroke weight, and contrast polarity on glance based legibility. *Displays* 2016, 41, 42–49.

Eden, G.F.; VanMeter, J.W.; Rumsey, J.M.; Maisog, J.M.; Woods, R.P.; Zeffiro, T.A. Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. *Nature* 1996, 382, 66–69.

Ehri, L.C. Phases of development in learning to read words by sight. *Journal of research in reading* 1995.

Facoetti, A.; Zorzi, M.; Cestnick, L.; Lorusso, M.L.; Molteni, M.; Paganoni, P.; Umiltà, C.; Mascetti, G.G. The relationship between 507 visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive neuropsychology* 2006, 23, 841–855.

Fella, A.; Christoforou, C.; Loizou, M.; Papadopoulos, T.C. Investigating the relationship between phonological awareness and reading using Event-Related Potentials. *Psychology: The Journal of the Hellenic Psychological Society* 2022.

Franceschini, S.; Bertoni, S.; Ronconi, L.; Molteni, M.; Gori, S.; Facoetti, A. “Shall we play a game?”: Improving reading through action video games in developmental dyslexia. *Current Developmental disorders reports* 2015, 2, 318–329.

Frith, U. A developmental framework for developmental dyslexia. *Annals of dyslexia* 1986, 36, 67–81.

Galaburda, A.M.; Sherman, G.F.; Rosen, G.D.; Aboitiz, F.; Geschwind, N. Developmental dyslexia: four consecutive patients with cortical anomalies. *Annals of Neurology: Official Galiano, A.R.; Augereau-Depoix, V.; Baltenneck, N.; Latour, L.; Drissi, H. Luciole, a new font for people with low vision. *Acta Psychologica* **2023**, 236, 103926.*

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *communication*.

Huang, S.M. Effects of font size and font style of Traditional Chinese characters on readability on smartphones. *International Journal of Industrial Ergonomics* 2019, 69, 66–72.

Journal of the American Neurological Association and the Child Neurology Society 1985, 18, 222–233.

Job, R.; Sartori, G. Morphological decomposition: Evidence from crossed phonological dyslexia. *The Quarterly Journal of Experimental Psychology* 1984, 36, 435–458.

Kahn, P., & Lenk, K. (1998). Design: principles of typography for user interface design. *interactions*, 5(6), 15.

Lian, Z.; Xiao, J. Automatic shape morphing for chinese characters. In *SIGGRAPH Asia 2012 technical briefs*; 2012; pp. 1–4.

Liu, N.; Yu, R.; Zhang, Y. Effects of font size, stroke width, and character complexity on the legibility of Chinese characters.

Human Factors and Ergonomics in Manufacturing & Service Industries **2016**, 26, 381–392.

Martinez, R.; Fernandez, A.; et al. The social and economic impact of illiteracy: Analytical model and pilot study 2010.

Miyazaki, T.; Tsuchiya, T.; Sugaya, Y.; Omachi, S.; Iwamura, M.; Uchida, S.; Kise, K. Automatic generation of typographic font from small font subset. *IEEE computer graphics and applications* **2019**, 40, 99–111.

Montani V., De Filippo De Grazia M., Zorzi M. (2014). A new adaptive videogame for training attention and executive functions: design principles and initial validation. *Frontiers in Psychology*. 5, 409. doi: 10.3389/fpsyg.2014.00409

Myronenko, A., & Song, X. (2010). Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12), 2262-2275.

Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic 417 transparency. *Behavior research methods*, 41(4), 991–1008.
<https://doi.org/10.3758/BRM.41.4.991>

Rae, C.; Harasty, J.A.; Dzendrowskyj, T.E.; Talcott, J.B.; Simpson, J.M.; Blamire, A.M.; Dixon, R.M.; Lee, M.A.; Thompson, C.H.;

Ralekar, C.; Saha, P.; Gandhi, T.K.; Chaudhury, S. Effect of devanagari font type in reading comprehension: An eye tracking study. In *Proceedings of the Intelligent Human Computer Interaction: 10th International Conference, IHCI 2018, Allahabad, India, December 7–9, 2018, Proceedings 10*. Springer, 2018, pp. 136–147.

Ramus, F.; Rosen, S.; Dakin, S.C.; Day, B.L.; Castellote, J.M.; White, S.; Frith, U. Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain* 2003, 126, 841–865.

Rello, L.; Baeza-Yates, R. How to present more readable text for people with dyslexia. *Universal Access in the Information Society* **2017**, 16, 29–49. 514

Share, D.L. Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition* 1995, 55, 151–218

Snowling, M.; Hulme, C. Do we really need a new definition of dyslexia? A commentary. *Annals of dyslexia* 2024, 74, 355–362.

Snowling, M. Dyslexia as a phonological deficit: Evidence and implications. *Child Psychology and Psychiatry Review* 1998, 3, 4–11. 496

Stella, G.; Savelli, E. *Dislessia oggi. Prospettive di diagnosi e intervento in Italia dopo la legge 170*; Edizioni Erickson, 2011

Stein, J. The magnocellular theory of developmental dyslexia. *Reading and dyslexia: From basic functions to higher order cognition* 2018, pp. 103–134.

Styles, P.; et al. Cerebellar morphology in developmental dyslexia. *Neuropsychologia* 2002, 40, 1285–1292.

Subasinghe, R.; Samarawickrama, S. Legibility of Noto Sans Sinhala Font Features for Small-Scale Digital Device Screens. In *Proceedings of the 2024 Moratuwa Engineering Research Conference (MERCon)*. IEEE, 2024, pp. 590–594.

Tressoldi, P.E.; Vio, C.; Scotti, P. Effects of simultaneous auditory feedback on reading accuracy during reading development. *Reading and Writing* 1991, 3, 151–158

Ungerleider, L.; Mishkin, M., Two cortical visual systems. In *The Analysis of Visual Behaviour*; MIT Press: Cambridge, MA, 1982; chapter 18, pp. 549–586.

Vaughn, S.; Miciak, J.; Clemens, N.; Fletcher, J.M. The critical role of instructional response in defining and identifying students with dyslexia: A case for updating existing definitions. *Annals of Dyslexia* 2024, 74, 325–336.

Vellutino, F.R.; Fletcher, J.M.; Snowling, M.J.; Scanlon, D.M. Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of child psychology and psychiatry* 2004, 45, 2–40.

Zappalà, E., Viola, I., & Aiello, P. (2024, January). Future Support Teachers' Intentions to Adopt and Use Technology with Students with ASD. In *Workshop on Artificial Intelligence with and for Learning Sciences: Past, Present, and Future Horizons* (pp. 79-88). Cham: Springer Nature Switzerland.

Zhang, T. Y., & Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3), 236-239.

Ziegler, J.C.; Castel, C.; Pech-Georgel, C.; George, F.; Alario, F.X.; Perry, C. Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition* 2008, 107, 151–178.

Zoccolotti, P. Success is not the entire story for a scientific theory: The case of the Phonological Deficit Theory of dyslexia. *Brain Sciences* 2022, 12, 425.

The constructivist potential of Virtual Reality

Vicarious dimensions of virtual corporealties

The system of reading and writing has certainly been a crucial technology in the evolutionary process of the human being and the way we learn. Information, in this way, is not bound to human mnemonic flow but can be stabilized and made much more resistant to time compared to potential human biological capabilities. Although we continue to use it incessantly, we have always known that it has never been the definitive technology for the extraction, processing of information, or, even better, communication. The concept of the inexpressible, widely discussed by philosophers, poets, and artists throughout the centuries, is a crucial symptom that reveals the limit of every reading and writing system, transcending language, culture, and progress for every civilization. This is also one of the reasons why the process of creating neologisms never stops, and with it, the very evolution of a language. However, the human being has never bound its communication to just one medium; rather, it has expanded its offering of media to meet the ineluctable necessity of communicating. It thus makes sense to ask where the communicative capacity of text ends and that of the image begins, for instance. Should those familiar with the story of Laocoön recounted in the Aeneid then deem it unnecessary to pause before the Laocoön Group displayed in the Vatican Museums?

As we know, even if the story is the same, the two media used to tell it can offer us different information. The bodies of the suffering Trojans sculpted in marble, through mimicry and gesture, substitute the verbal repertoire with significant actions which, playing on an interpersonal refraction that produces an emotional contamination, or an embodied sharing of emotion, allow for structural coupling and co-evolution on the level of emotional meanings (Sibilio, M., 2017). In other words, a different impact is achieved on different cognitive components that characterize our individual learning process. The responsibility, naturally, should not be associated exclusively with the different medium used, but jointly with the medium, with the one who conveyed the message through it, and with the one who was disposed to receive it. These episodes, if misinterpreted, could therefore suggest to teachers that they must simultaneously become poets and sculptors as a propaedeutic condition even for their first teaching experience. This demand, which was already difficult

or unsustainable then, is perhaps even more evident today, considering the increasingly deep level of technical specialization required and the ever-more varied technological offering. A current example is certainly offered by all the AI-based technologies used by everyone but whose domain of intervention is almost exclusive to ICT experts. The role of the teacher, therefore, remains of another nature, and the example proposed must, in fact, suggest two different reflections: the importance of corporeality in the learning process and the potential for deviant response of media.

Corporeality, understood as the incorporated potential of interaction, redefines the environment in which we are immersed and which we continually attempt to adapt to. Learning, therefore, as a form of adaptation to the environment, demands the activation of both the physical and immaterial duality, as well as that of the teacher and the student. In this sense, on one hand, corporeality in didactic practice represents an identity that does not correspond exclusively to the morphological and functional peculiarities of the teacher, but to the modes of acting that translate into agency intentionally directed toward educational aims. On the other hand, the indissolubility between the learning process and the individual restores a necessary plural dimension to cognitive processes, which are expressions of the corporeality of the learning subject.

From the paradigm of corporeality, it is thus possible to define the concept of vicariance in the educational context. The human being, in fact, as Alain Berthoz (2011) argues, utilizes strategies of memory, flexibility, and adaptation to change in order to respond to the complexity of the context. This allows for a nonlinear transposition that adheres to the principles of probabilistic anticipation and deviation, giving meaning to actions that are alternative to the linearity of traditional transmissive methods. Analogous to what physiologically occurs in our body, which predisposes itself to a synergistic activation of muscles for the execution of a movement, the teacher should predispose the didactic action by activating those very semplessa (simplex) strategies to favor learning. The didactic action therefore becomes vicariant when the teacher relinquishes their own consolidated learning behaviors and instead facilitates the emergence of alternative—and thus deviant—behaviors for the resolution of complexity.

Deviation is, in fact, the common operation found in both functional vicariance, where the teacher deems it effective to substitute or modify didactic practices, and in use vicariance, where the teacher identifies a functional plurality in the same technology or space. In the educational context, both technologies and spaces can be distinguished into two analogous classes:

Didactic Technologies: technologies that were created specifically for educational-didactic purposes.

Technologies for Didactics: technologies that originate and are used in contexts other than the educational-didactic one but are adapted for didactic purposes.

Codified Spaces: places designed and intended for the use of didactic activities, such as classrooms, gyms, laboratories.

Uncodified Spaces: any place that can lend itself to the use of didactic activities.

In each of these cases, the teacher can implement vicariance strategies, for example, by delegating the location of the classroom to that of the ceramics laboratory for an Art Education lesson, and from the latter to the museum path; or, in the case of technologies, they could delegate the use of online tools to that of the IWB (Interactive Whiteboard) and vice versa.

The crucial case that inherits all the deviant potential of both spaces and technologies is precisely where the technology is de facto also a new space, such as Virtual Reality.

In line with our on-life condition, VR transcends the role of a mere didactic tool to configure itself as a new, indistinct, training environment. Like other technologies, VR opens up new educational scenarios, but its distinctiveness lies in its ability to allow complete sensory immersion (Biocca, 1995), actively promoting experiential learning (Kolb, 1981) and making

activities (playful or educational) highly engaging. This fusion between the physical and the digital is crucial for contemporary didactics. The educational institution faces the challenge of leveraging VR, as the use of immersive and interactive devices demonstrates a significant increase in student involvement and motivation in learning (Kaplan-Rakowski & Wojdyski, 2018; Kaplan-Rakowski, Johnson, Wojdyski, 2022). Khukalenko (2022) highlights how the effectiveness of VR stems from its capacity to offer visually rich simulations, support student-centered learning, and allow safe experimentation of otherwise impossible scenarios. In this context, the user experiences a deep sense of presence that facilitates the visualization of abstract concepts.

Precisely by virtue of this spatial and sensory fusion, the vicariant dimension of virtual corporeality emerges as the full actualization of the principles of a complex didactics based on flexibility and functional deviation. The didactic action becomes vicariant at the moment when VR, as an on-life platform, is no longer a technology to be adapted nor a space to be substituted, but a unique environment that inherits and amplifies the potential of both. By allowing immersion and motor activation, VR provides students with the possibility to exercise probabilistic anticipation and adaptation to change in otherwise inaccessible experiential scenarios. In this way, VR delegates to the infosfera the task of supporting the cognitive deviation necessary for the resolution of complexity, thus establishing itself as the privileged space for corporeally active learning in the digital age.

The case study of Inclusive Virtual Museum

The intrinsic capacity of Virtual Reality to guarantee a corporeally active and vicariant experience within the infosfera makes it a fundamental tool for responding to the educational needs of our digital identity. This potential finds a crucial application field in the context of Museum Education, understood according to the broadest definition of a museum: a permanent, accessible, and inclusive institution, serving society, which exhibits tangible and intangible heritage to promote education, reflection, and knowledge sharing. Since museum education is traditionally constrained by physical spaces and predominantly

transmissive exhibition modes, it requires tools capable of amplifying the didactic reach of artifacts, transforming learning into an experiential, customizable, and inclusive process. It is in response to these challenges that VR evolves from a simple technology into a genuine museum learning infrastructure. The pilot project, the Inclusive Virtual Museum (IVM), was conceived to identify design strategies that consider the needs of all students, rethinking the relationship between the body and adaptation to the digital environment. Developed entirely on Unity, the IVM initially proposes a single exhibition room utilizing 3D scans of archaeological artifacts made available by various museums in Campania. Each digital asset is enriched with descriptive panels featuring customizable fonts, audio-video screens, and an interactive humanoid guide (Campitiello, L. et al., 2023).

Although access to the IVM is also possible via a traditional computer screen, the highest level of immersion for the virtual experience is achieved through the use of a VR visor, which enables both ordinary actions of movement and interaction, as well as 'extraordinary' actions such as grasping or close-up observation of artifacts. Since visual comfort is crucial, the IVM allows for the runtime modification of settings such as light, contrast, and field of view.

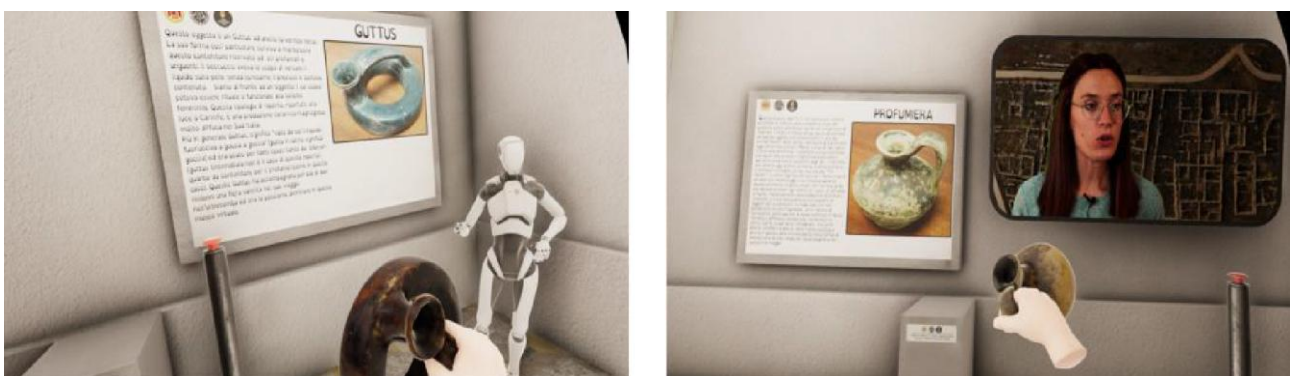


Figure 8. Examples of the VR mode of the Inclusive Virtual Museum



Figure 9. Examples of the "vignette effect", one of the possible inclusive visual configurations of the immersive experience.

This personalization made possible a preliminary qualitative analysis, conducted through the combined administration of the virtual experience and a survey on a sample of 105 primary and secondary school pupils from 5 different schools in Campania. The experience was structured into three sequential modules: the first dedicated to visor configuration, acclimatization, and preliminary questions regarding the pupils' background; the second focused on interaction with the objects and multimedia content; and the third, the final part, aimed at collecting data on user satisfaction, suggestions for improvement, and symptoms associated with virtual reality sickness.

The questionnaire, created using Google Form, followed the structure described above. Both the questions and the possible alternatives were read aloud by one operator tasked with guiding the student through the virtual experience and collected by another operator outside the user-digital environment system.

The results confirmed the students' high familiarity with digital platforms: 95.3% play video games (41.4% on consoles, 36.4% on smartphones), and 88.6% are aware of VR. Furthermore, 82.9% had already had previous VR experiences at school. The analysis highlighted that 58.1% of participants preferred the open view, while only 15.2% found the vignette effect useful—a customization feature that, although used by a minority, proved necessary to allow students with intellectual disabilities to complete the experience (approximately 15–20 minutes) without symptoms. Regarding comfort, 78.1% of

participants reported no symptoms, while only 10.5% complained of dizziness/headaches, and 6.7% experienced both symptoms.

From the limits of Inclusive Virtual Museum to the Hein's constructivist model

Analysis of the feedback collected during the final stage of the immersive experience, particularly in response to the question: "What changes do you recommend to improve the VR experience?", revealed that the students' enthusiasm for using technology similar to video games in the educational field is coupled with clear requests for expansion and greater exploratory freedom.

The majority of students suggested enriching the virtual environment with additional archaeological artifacts, paintings, and statues that can be observed and manipulated. Beyond mere quantitative expansion, a demand for greater interactivity in the museum hall emerged, highlighted by the desire to initiate conversations with avatars or to ask questions to historical figures. They also expressed interest in the creation of additional rooms and the addition of non-playable characters (NPCs) to improve the overall experience.

These critical responses indicate that, despite the initial VR experience being appreciated and demonstrating high engagement, its current configuration (focused on a single room with static artifacts and predefined descriptions) does not fully satisfy the need for self-directed learning based on personal inquiry. The students' suggestions—aimed at expanding the manipulable catalog, diversifying spaces, and enabling complex narrative interactions—signal a future direction focused on making the virtual experience not only more interactive and engaging, but also easily adaptable to the different needs of users, shifting the focus from passive consultation to autonomous exploration and the active construction of knowledge.

The theoretical framework that best addresses this need is the one proposed by George E. Hein (1999), whose ideas, rooted in constructivist learning theory and inspired by John

Dewey, profoundly influence contemporary museum education. Hein, while acknowledging the importance of interest and practical experience (hands-on), warns that engagement alone is insufficient; museum activities must promote deeper growth through structured, meaningful, and integrated experiences. His reflections on theories of learning and knowledge (the acquisition of truth versus the construction of knowledge) led to the categorization of museums into four models: the systematic, the orderly, the discovery, and the ideal constructivist museum.

Hein's proposed classification identifies a reference system where one direction is represented by two of the main theories of learning while the other direction is represented by theories of knowledge. Based then on the peculiarities of the place, the mode of information presentation, and the activities proposed, it is possible to determine 4 main types of museums.

The Constructivist Museum is strongly advocated by Hein because it adapts to the diverse educational needs of visitors—a fundamental concern raised by student feedback—rather than being constrained by the intrinsic properties of the exhibited objects. This model, which is rooted in thinkers like Jean Piaget and Lev Vygotsky, promotes personalized, socially interactive, and practical learning, allowing visitors to choose their own pathways and engage with multiple perspectives (Hein, G. 2006). Unlike the systematic museum (didactic and textual) or the discovery museum (which often predetermines outcomes), the constructivist museum transforms museums from static archives into dynamic spaces. The intrinsic nature of this model requires that the artifact has a multiperspective presentation and that information be distributed across different levels of depth, allowing the user to construct meaning by progressively accessing these levels (Todino, M.D., Campitiello, L. 2025). It follows that, for learning purposes, the manner in which the artifact is presented must have the same dedication and dignity as the artifact itself, and that a historically important collection is not in itself a guarantee of didactic effectiveness. Interactions must be diversified for a heterogeneous public, the experience must trigger the search for new experiences, and, crucially, it must yield a useful tool "outside the museum." The students' requests to manipulate and converse, therefore represent a clear indication toward the implementation of a Virtual Constructivist Museum model, where the integration of

multisensory and interactive elements, facilitated by VR, can maximize accessibility and learning outcomes.

Constructivist Museum Exhibition : From Realty to Virtual

Some of the proposed improvements for the Inclusive Virtual Museum project, which emerged from the feedback of students who participated in the experimentation, can be interpreted in light of Hein's reflections on the constructivist museum type. The Inclusive Virtual Museum originated as a virtual experience potentially integrable into real museum pathways, consistent with the artifacts and content proposed, or it can be configured as a genuine museum education activity delivered in an innovative way through immersive technologies . Future interventions on this project, or on analogous projects, should therefore be oriented toward integrating the distinctive characteristics of the constructivist model. The transposition of the standard museum environment—with its artifacts, descriptive panels, and monitors—into the virtual world can, on one hand, facilitate outreach to an extended audience, including students who benefit from both the particular technology and the customizable audiovisual settings. On the other hand, however, the way the virtual environment has been conceived seems to recall the peculiarities of a real museum of a descriptive nature. The limitations and problems raised by the students are, in fact, often the same as those encountered during the visit to a "classic" real museum; in other words, the students' feedback could represent empirical evidence that the virtualization of a descriptive museum environment or pathway tends to retain the critical issues of that model even in its digital version. The opposite question spontaneously arises: in the event that a constructivist-type museum or museum pathway could be virtualized, would it succeed in preserving the advantages highlighted by Hein's reflections?

The first step to answering this question would therefore be to find a real constructivist museum or museum pathway and transform it into a virtual experience. The choice thus fell on the temporary exhibition "Malta. History and Science in Fragments" held at the Monastery of St. John in Mústair and at Castel Grande in Bellinzona (Bilotti, U. et al. 2025).

The coherence with Hein's constructivist model is evident in the conception of the museum pathway, as George Hein articulates his constructivist theory for museums based on the idea that learning is not a process of passive reception of information, but an active process in which the visitor constructs their own meaning from their pre-existing experiences and knowledge. In full harmony with Hein's proposal, the exhibition emphasizes the pre-existing experience and the small object (bottom-up construction): the protagonist is a humble archaeological find, one of the over 5,000 fragments of mortar discovered during various archaeological excavations at the Monastery of St. John in Moustair (a UNESCO World Heritage site), which may seem "of little interest" to the untrained eye; Hein maintains that learning is situated and personal, and starting from a seemingly insignificant or marginal object challenges the visitor to actively engage in discovery, recognizing that learning occurs when the new datum (the fragment) actively connects to pre-existing knowledge schemes and is transformed into a source of information through scientific observation, with the fragment becoming an "interpretive key" that favors the construction of a personal and non-imposed meaning. Added to this is the structured process of progressive discovery (Ambrose, T., & Paine, C., 2006), since the exhibition is structured as a "journey of knowledge that leads from the fragment to the global understanding" of the construction process and its historical dimension, through "stage after stage, room after room, activity after activity"; Hein emphasizes that learning is a process, not a product, and the transition from the analysis of the mortar fragment (the elementary datum) to the reconstruction of medieval construction techniques and their temporal evolution (the global understanding) is a perfect example of how the exhibition facilitates the progressive reconstruction of the context by the visitor. The engagement and relevance of broad themes are also fundamental, as the exhibition is not limited to the artifact but creates opportunities to address much broader and more complex topics such as the archaeologist's long work, medieval construction techniques, and the necessity of interdisciplinary dialogue in research; a constructivist museum must make the object relevant by connecting it to ideas of greater scope and social/cultural significance, and transforming the study of the mortar (physical, chemical, microscopic investigation) into a mediator between science and humanism, using it to rethink the Middle Ages, elevates the object of study to a vehicle of historical and cultural understanding, stimulating the visitor to restructure their

preconceived ideas about the period. Finally, the Interdisciplinary Approach (Holistic Vision) is fully represented by the museum pathway, which is founded on a strongly interdisciplinary approach, combining archaeology, materials science, and art history within a unified interpretive framework; constructivism in a museum context encourages the integration of different perspectives and disciplines, showing how the object (the mortar) is not only technical data but a social and cultural testimony, allowing the visitor to see the links between technology, society, and culture, building a more holistic and rich understanding.

Beyond Description: Leveraging Constructivism and Mediators for Immersive Museum Education

Before proceeding with the creation of a virtual transposition of the museum pathway, it is essential to recover the key components of the constructivist museum model. Like all temporary exhibitions, this display has a defined lifecycle. Upon its conclusion, although archaeological artifacts are relocated to storage for effective conservation and multimedia materials are saved on hard drives or the cloud, the crucial interactions consistent with the constructivist approach are at greater risk of being irretrievably lost. The problem of reproducibility thus extends beyond individual artifacts or supporting materials, more complexly involving the activities proposed along the pathway. The logistical phase following the closure of the physical exhibition involves the relocation of artifacts to suitable storage environments and the disposal or reuse of equipment, making it difficult, if not impossible, to recover or update specific educational interactions for future use or for audiences with diverse needs. For this reason, the virtual museum pathway intended for creation must not be a mere digital replica, but a strategic experience aimed at preserving the central themes of the exhibition and proposing new, effective interactions for content understanding. Design principles have been defined to paraphrase the meta-objectives and serve as primary reference points for designing the virtual experience. In order to create a product consistent with current museum education proposals.

Constructivist foundation and the primacy of Interaction

The constructivist museum positions itself as the venue capable of functionally proposing its artifacts for agile and meaningful interaction with its visitors, who seek both the acquisition of new knowledge and entertainment in a non-formal context. From this perspective, the potential information intrinsic to the artifacts cannot be fully expressed through description alone. The pairing of the artifact with only descriptive material, even in various forms such as text, video, or audio, might not be sufficient for one of the two primary goals, resulting in an outcome not dissimilar to that of other sources, such as books or documentaries. However, while this type of material might be commonly considered the primary means of accessing information in the scholastic context, in the museum context, it assumes a role of information deepening subsequent to an already established interest. For instance, facing a painting, we naturally first admire it as such and then delve into all the aspects connected to the work to appreciate it from other perspectives.

Rapid Engagement and Manipulative Activities

It follows, then, that engagement must occur rapidly and certainly cannot be deferred until a subsequent comprehension of textual or multimedia content. Artifacts, by making part of their information participatory (Todino, M. D., & Campitiello, L., 2025), should manage to communicate with those who admire them from the very first moment. Once the barrier separating the artifact from the visitor is broken, it is crucial to foster the information-seeking phase by structuring activities that investigate the artifact at different levels. These activities effectively seek to reproduce the various contexts to which the artifact belongs, and these contexts, in turn, become the environments where the visitor extracts meanings and constructs knowledge. The constructivist view recognizes the existence of crucial dimensions such as the practical, sensory, and emotional, considering them indispensable for the learning process.

Leveraging Immersive Technology: Edugame and Virtual Reality

To address these issues, the decision was made to develop an edugame capable of virtually reproducing the artifacts, locations, and interactions that characterized the physical exhibition in Bellinzona. The edugame tool has been central to pedagogical research for decades due to its important potential, such as facilitating gamification strategies, coherently meeting *onlife* needs, and serving as an additional assessment tool, among others. Furthermore, Virtual Reality applications contribute to increasing motivation, engagement, and critical thinking, thus being capable, either directly or transitively, of fostering knowledge transfer.

The Teacher as Facilitator vs. The Museum Guide

In formal contexts, the responsibility for intervening on these dimensions falls almost entirely on the teacher who, knowing their students day by day, is tasked with addressing the complexity deriving from their cognitive singularities. To make these interventions effective, the teacher's role cannot be understood as a mere channel of information between two isolated entities, the environment and the student. Thus relinquishing an approach based on their own learning style and communicative modalities, the teacher must become a facilitator of the student's learning style and a communicative common denominator among students (Sibilio, M., 2023). Through this redefinition of the teacher, it is then possible to overcome the separate environment-student conception toward a context-knowledge dualism and promote the evolution of student actions from an observational to a manipulative nature. Despite the success of this model, the hypothesis that the museum guide could analogously assume the same responsibilities and facilitator functionality found in the teacher for their visitors is not sustainable for multiple reasons. Firstly, the museum guide is often an optional choice for the visitor. Furthermore, the museum guide does not know their visitors nor do they have the same opportunities to know them as the teacher, due to the limited interaction time restricted to the duration of the visit. Finally, while they are not required to assess their audience's learning—which on one hand exempts them from

one of the most complicated tasks—this simultaneously precludes them from important self-assessment and activity calibration tools. While these different initial conditions might be sustainable for a process aimed at promoting cultural heritage, entertainment, and the economy related to the museum, they are not sufficient for the acquisition of new knowledge.

Strategic Alternation of Mediators

In response to these potential gaps, the use of different types of mediators is more justified in this context. Damiano's classification of mediators (Damiano E., 2013) foresees the existence of two types of mediators: warm mediators, which are highly capable of arousing motivation and activating emotional and affective factors, and cold mediators, which are more suitable for processes of exposition, synthesis, and content organization. Once the characteristics of the two types of mediators are isolated, the choice of use becomes simpler based on the activities that have been planned in a given pathway. Given the plurality of learning styles and learning phases, the only potentially effective strategy is the structured alternation of the two types.

Digital Asset Production and Virtual Environment Design

The successful creation of a virtual environment capable of reproducing key experiences from the physical exhibition hinged on the meticulous acquisition and design of new digital assets. This process involved a multi-faceted approach to accurately digitize archaeological artifacts, design essential functional elements, and structure the resulting virtual space to maximize clarity, accessibility, and engagement.

Artifact Acquisition and High-Fidelity 3D Modeling

The first essential step was recovering and digitizing the archaeological artifacts and other objects from the original museum pathway. For this, a hybrid approach was adopted: some

virtual objects were derived via 3D scanning, while others, deemed too complex for direct scanning, were meticulously reconstructed *ex novo* using photos and videos as reference. Given the central role of the mortar fragment in the museum narrative, maximum possible fidelity was sought through high-resolution scanning followed by detailed *sculpting* and *texture* extraction (Campitiello, L. et al., 2023). The resulting object has a complex geometry and a high polygon count, accurately conveying the composite and heterogeneous nature of the material. This asset allows for close observation, preempting the detailed information to be explored later across various research fields.

Creation of functional and evocative digital assets

While the widespread availability of user-friendly 3D software and the increasing affordability of 3D printing technologies, coupled with the growing open-source philosophy, have led to a broader availability of digital assets compared to the previous decade, many necessary objects—particularly those required for hands-on, laboratory-style activities—were not found available online or for free. For these essential, activity-specific items, a rough but functional reconstruction was undertaken. This was the case for the structures replicating the scaffolding typical of construction sites and archaeological digs. These structures serve a strongly evocative function, recalling the workplace of archaeologists and the diverse manual and artisan figures across different historical periods. Furthermore, their versatility was heavily exploited in the physical exhibition, allowing the creation of varied modules—such as tables for lab activities, vertical display walls, shelving for conservation and cataloging, and truss systems for multimedia technologies. Their faithful implementation in the virtual transposition was primarily justified by the desire to inherit this evocative and flexible functionality. Moreover, these structures inherently possess characteristics tied to precise spatial placement, confirming the initial design goal of furnishing an exhibition-didactic space in the virtual realm. Finally, creating these new digital assets was immediately cost-effective in terms of 3D modeling, as a good approximation of their elementary components could be achieved by composing basic geometric figures.

Structuring the Virtual Environment and Spatial Optimization

The initial phase of asset prototyping, through both 3D scan refinement and *ex novo* reconstruction, led to process standardization and faster production. With the progressive availability of the new digital assets, the project to convert the museum pathway into a virtual experience commenced. The project team selected the Bellinzona layout as the reference due to the clearer modularity of its thematic sections. Consequently, the virtual experience was structured as a sequence of rooms, accessible to the player at any time while preserving the original sequential thematic order. The virtual environments were designed to maximize clarity and spatial order. Virtual rooms utilized textures approximating the colors and materials of corresponding physical elements (like walls and floors), but their dimensions were slightly increased. This deliberate choice aimed to optimize the perceived space and improve visitor orientation. The arrangement of the main display cases faithfully reproduces the real layout.

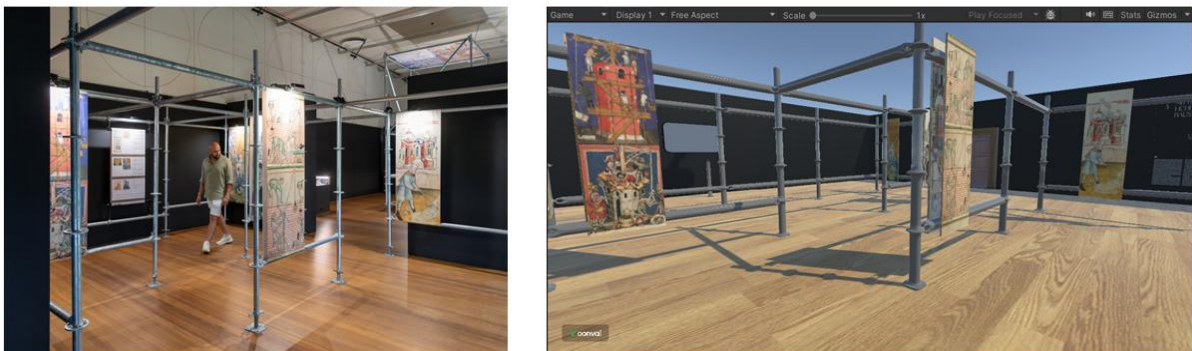


Figure 10. The third thematic room on medieval building techniques in two different versions: on the left, a photo of the temporary exhibition in Bellinzona, on the right, a screenshot of the virtual experience.

Implementing Adaptive Interactions for Accessibility

Intervention was necessary on the associated descriptive panels to ensure full legibility and accessibility within the virtual context. Since non-vertical panels proved difficult to read in the virtual environment, an adaptive interaction logic was introduced. When the user interacts with horizontally placed descriptive panels, they automatically rotate by 30° relative to the resting plane—a dynamic inclination that facilitates reading and consultation. For panels integrated into tables designed for lab activities, which are often covered by other objects, the interaction triggers the appearance of two-dimensional elements on the screen (via Canvas UI). This solution provides effortless reading of the textual content while allowing interactive objects, such as tools or artifacts, to remain on the virtual table.

The methodical process of digital asset creation, encompassing high-fidelity artifact modeling and the functional design of structural elements, alongside the strategic structuring of the virtual environment and the implementation of adaptive accessibility features, successfully laid the foundation for a virtual museum experience that conserves the thematic core and active engagement principles of the physical exhibition.

Interdisciplinarity in the virtual museum through the use of AI-based NPCs

The virtual experience significantly benefited from its constructivist origin, whose design choices regarding artifacts and activities facilitated the transposition into the new expressive paradigm. The museum guide's role as a facilitator and communicative mediator is an essential pedagogical element of the constructivist model, and it was crucial to recover this function in the virtual experience. It was decided to integrate this function through NPCs, acting as escorts and equipped with advanced functionalities based on a modern chatbot powered by Artificial Intelligence (LLM). To address the risk of generic Large Language Models producing imprecise answers in knowledge-intensive contexts—a phenomenon known as "hallucinations"—the Retrieval-Augmented Generation (RAG) strategy was adopted (Lewis, P. et al., 2020). Thanks to RAG, the NPC draws responses from an authoritative knowledge base, comprising selected references and content created by museum professionals or subject matter specialists, thereby ensuring the accuracy essential

for an educational context (Bilotti, U. et al., 2025; Campitiello, L., Beatini, V., & Di Tore, S., 2024).

To translate the constructivist requirement for a holistic vision and underscore the exhibition's intrinsic multidisciplinary nature, the initial choice of a single NPC was replaced by the creation of multiple thematic guides, each corresponding to a specific professional figure. These thematic guides are:

- *The Archaeologist*: Sets the narrative and stratigraphic contextualization in Room 1.
- *The Chemist and the Geologist*: Offer the scientific analysis of the mortar and trace "The Way of Lime" and "The Way of Sand" in Room 2.
- *The Master Carpenter*: Explains the organization of the medieval construction site and building techniques in Room 3.
- *The Historian*: Frames the artifacts and scientific data within the eight construction phases of the Müstair Monastery in Room 4.
- *The Lime Kiln Master*: Provides an experiential focus on the ancient lime production process in the Fifth Environment.

This plurality of professional voices enriches the experience, offering the visitor multiple knowledge frameworks for constructing their own meaning.

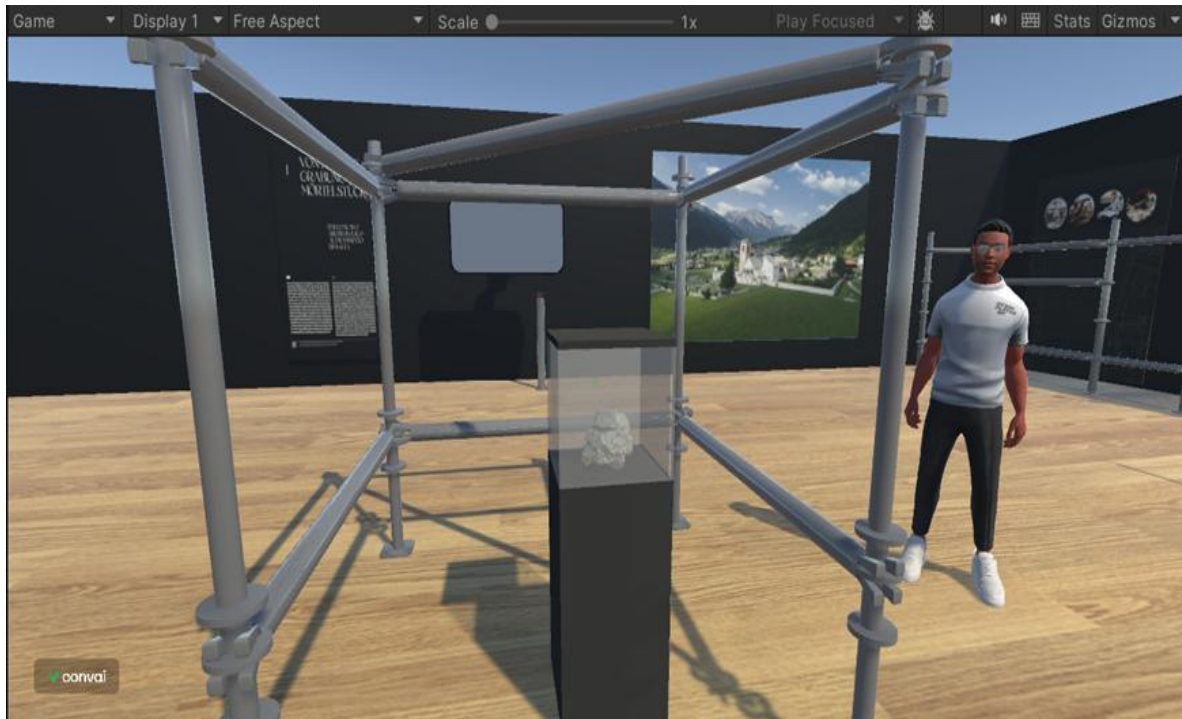


Figure 11. The first room of the museum itinerary with the piece of mortar and the archaeologist NPC at the centre of the virtual environment

Bibliography (The constructivist potential of Virtual Reality)

Ambrose, T., & Paine, C. (2006). Museum basics: The international handbook. Routledge.

Berthoz, A. (2011). La semplicità (pp. 1-224). Torino: Codice.

Bilotti, U., De Ascaniis, S., Tardini, S., Todino, M. D., Campitiello, L., Di Tore, S., & Sibilio, M. (2025). From Heritage to Gameplay: Developing an Educational 3D Experience for Museum Learning at the Monastery of St. John in Müstair. *Journal of Inclusive Methodology and Technology in Learning and Teaching*, 5(3).

Campitiello, L., Lecce, A., Bilotti, U. & Di Tore, S. (2023). INNOVATIVE EDUCATIONAL TECHNOLOGIES: A PRE-LIMINARY SURVEY TO DETECT THE ACCESSIBILITY OF THE INCLUSIVE VIRTUAL MUSEUM. *JOURNAL OF INCLUSIVE METHODOLOGY AND TECHNOLOGY IN LEARNING AND TEACHING*.

Campitiello, L., Caldarelli, A., Schiavo, F., Marras, A., & Di Tore, S. (2023). Augmented reality to promote cultural accessibility. *Journal of Inclusive Methodology and Technology in Learning and Teaching*, 3(2).

Bilotti, U., Di Paolo, A., Todino, M. D., Fella, A., & Di Tore, S. (2025). Valorizzare il materiale nell'immateriale: Il MUSEum Editor per l'accessibilità e la cultural heritage.

Campitiello, L., Beatini, V., & Di Tore, S. (2024, January). Non-player character smart in virtual learning environment: Empowering education through artificial intelligence. In *Workshop on Artificial Intelligence with and for Learning Sciences: Past, Present, and Future Horizons* (pp. 131-137). Cham: Springer Nature Switzerland.

Damiano, E. (2013). *La mediazione didattica. Per una teoria dell'insegnamento: Per una teoria dell'insegnamento*. FrancoAngeli.

Dewey, J. (1937). The educational function of a museum of decorative arts. In *the Collected Works of John Dewey, 1882–1953: Later works 11:520–524*, Electronic Edition, L. Hickman, ed. Charlottesville, VA: Intelex Corporation, 1996

Hein, G. E. (1999). The constructivist museum. *The educational role of the museum*, 2, 73-79.

Hein, G. E. (2006). Museum education. A companion to museum studies, 340-352.

Kaplan-Rakowski, R., Johnson, K. R., & Wojdynski, T. (2021). The impact of virtual reality meditation on college students' exam performance. *Smart Learning Environments*, 8(1), 21.

Kaplan-Rakowski¹, R., & Wojdynski, T. (2018). Students' attitudes toward high-immersion. *Future-proof CALL: language learning as exploration and encounters—short papers from EUROCALL 2018*, 124.

Khukalenko, I. S., Kaplan-Rakowski, R., An, Y., & Iushina, V. D. (2022). Teachers' perceptions of using virtual reality technology in classrooms: A large-scale survey. *Education and Information Technologies*, 27(8), 11591-11613.

Kolb, D. A. (1981). Learning styles and disciplinary differences. *The modern American college*, 1(January 1981), 232-235.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33, 9459-9474.

Sibilio, M. (2017). *Vicarianza e didattica. Corpo, cognizione, insegnamento*. Brescia: La Scuola.

Sibilio, M. (2023). *La semplicità: proprietà e principi per agire il cambiamento*. Morcelliana.

Todino, M. D., & Campitiello, L. (2025). Museum education. *Encyclopedia*, 5(1), 3.

An Inclusive VR/AI Learning Pathway

Design of the Learning Unit and virtual experiences

The results derived from the experimentation of the Inclusive Virtual Museum, combined with reflections stemming from Hein's constructivist model , offer a solid foundation for understanding the intrinsic constructivist capabilities of VR and, consequently, the Educational Metaverse.

While virtual experiences are increasingly integrated successfully into exhibitions and museum pathways, their full incorporation into school curriculum design remains complex, often relegated to an accessory role. This occurs because didactic attention is currently polarized by the introduction of AI-based tools. The ease of use and adaptability of these tools make their experimentation an almost irresistible priority, necessitating an urgent pedagogical reflection.

However, an exclusive focus on AI tools risks leaving underutilized those technological resources and competencies acquired through significant investment (not only by major stakeholders in the Metaverse but also by numerous smaller local educational realities). This disinvestment is evident not just in obsolete and unused equipment but, more critically, in the professional competencies that have been developed. Assuming that a well-acquired competence is a more enduring investment than the warranty of any hardware, the need arises to enhance the teacher's *vicarious use* the ability to adapt and reuse resources created for one context for a different didactic purpose.

The key reflections that follow are:

- *The Educational Metaverse as blended Learning:* The Educational Metaverse is not a monolithic environment, but a set of interconnected spaces that extend the didactic offering beyond the real dimension into the virtual, thereby facilitating blended learning.
- *Validity of VR:* The museum educational context serves as a crucial example of the constructivist capabilities of VR.
- *Priority on Interaction:* For a fully constructivist virtual experience, the simple virtualization of environments is insufficient; it is crucial to ensure the virtualization of meaningful and manipulative interaction with the content.
- *Necessary Adoption:* Teachers and students must not forgo the opportunities offered by immersive technologies.
- *AI for Vicarious Use:* The integration of Artificial Intelligence into experiential proposals is essential to provide accurate and personalized content, thereby strengthening the teacher's capacity for vicarious use of didactic resources.

Based on these observations, a Learning Unit has been proposed with the dual objective of promoting the cultural heritage of the ancient Samnite people and fostering the inclusion process for students with Attention-Deficit/Hyperactivity Disorder (ADHD).

The Learning Unit “*Journey into Ancient Samnium: a VR/AI Experience*” is conceived as an innovative and deeply inclusive educational proposal designed to actively engage first- and third-year lower secondary students through a reality-based learning task which requires them to explore and reinterpret the Samnite civilization by applying knowledge in an authentic and meaningful context. Rather than merely transmitting historical content, the unit aims to integrate History, Technology, and Art through the strategic use of emerging technologies such as VR and AI. Developed through a participatory process involving

students, subject teachers, and the research group of “Teaching Learning Center for Education and Inclusive Technologies – Elisa Frauenfelder” (University of Salerno), the project is grounded in solid scientific and methodological foundations.

The learning task is supported by a dual inclusive strategy designed in particular for students with ADHD:

- the “**Bottega del Vasaio - (Potter’s Workshop)**” offers an immersive experience tailored according to ADHD-sensitive design recommendations, fostering attention regulation, engagement, and accessible cultural understanding;
- The “**MUSeum Editor**” is conceived to promote peer collaboration and cooperative learning, enabling students to work together in a shared virtual space where structured social interaction becomes a driver of inclusion.

The combined use of VR and AI therefore represents not a mere technological addition but a deliberate pedagogical shift aimed at overcoming the limitations of traditional frontal teaching. Where complex texts or static images may hinder comprehension—especially for learners facing attentional or symbolic-processing challenges, the immersive and adaptive affordances of VR and AI create what Berthoz (2013) describes as an “inclusive detour,” enabling each student to pursue the same learning objective through differentiated pathways. Within this integrated experience, students strengthen essential competencies for active citizenship and lifelong learning, including digital literacy through the creation of multimedia content, cultural awareness through innovative representation of the Samnite heritage, social and civic responsibility in understanding the collective role of cultural preservation, and the capacity to “learn how to learn” through self-regulation, peer cooperation, and flexible use of alternative tools.

Overview of ADHD and interventions through videogames

ADHD Diagnosis and Etiology

The earliest historical trace of the disorder dates back to 1798 with the writings of Alexander Crichton, who, while not defining ADHD, described "mobile attention" (or inattention) as a condition where "Every impression seems to agitate the person and to produce a certain degree of mental restlessness." The first official clinical descriptions of a disruptive behavioral disorder, characterized by hyperactivity and a deficit in moral control (distinguished from morality), were made by George Still (1902). Still hypothesized a biological/organic cause, but the direct connection to brain lesions only became prevalent in subsequent years, particularly after the 1930s.

This led to the adoption of terms like the "brain-injured child syndrome" and its subsequent evolution toward the more neutral "Minimal Brain Dysfunction" (MBD), used to denote a neurological dysfunction in the absence of evident physical proof of damage. Beginning in the 1960s, the causal hypothesis of physical brain damage was questioned, but MBD remained a fundamental concept for differential diagnosis. The evolution of the disorder in the DSM-5 (2013) led to significant changes, including raising the age by which symptoms must be present (from 7 to 12 years) and the specification of nine criteria for inattention and nine for hyperactivity/impulsivity.

From an etiological standpoint, ADHD is considered a multifactorial disorder (genetic, neurobiological, and environmental) (Wender, P.H., 2000). One of the key neuropsychological models of reference is Russell Barkley's cognitive-behavioral hybrid model, which places Behavioral Inhibition (BI) at the center of self-regulation (Barkley, R., 1997; 2006; 2015). BI is defined as the capacity to (1) inhibit immediate prepotent responses, (2) interrupt an ongoing response, and (3) control distraction. A deficit in BI compromises the intervention of four Executive Functions (EFs) that modulate the individual's response: Working Memory (verbal and non-verbal), Internalized Speech (verbal working memory), Self-Regulation of Emotion/Motivation (the ability to modulate and delay immediate reactions), and Reconstitution/Planning (the ability to decompose and recombine behavioral units to construct new goal-directed patterns).

Edugames, Serious Games, and Inclusive Design for Students with ADHD

The effective management of ADHD, given its frequent diagnosis in school-age children and the manifestation of complex self-regulation deficits and behavioral, social, and motivational issues, requires a multimodal approach combining pharmacotherapy (stimulants) with non-pharmacological interventions (behavioral therapy and cognitive/social skills training). This clinical necessity for diverse interventions aligns perfectly with the pedagogical principle of Universal Design for Learning (UDL). New digital therapeutics, including Virtual Reality, neurofeedback, and dedicated video games, are being tested to improve engagement, ensure high scalability, and facilitate adaptive cognitive training.

The therapeutic potential of these tools is strengthened by the ubiquity of video games. Despite concerns about potential adverse effects, empirical data suggests that video games possess an intrinsic capacity for cognitive enhancement, acting through the manipulation of attention and probabilistic inference. This phenomenon has spurred the adoption of Serious Games (SGs) as a validated auxiliary mechanism to assist individuals with cognitive difficulties. Serious Games are digital applications designed to transcend mere entertainment by pursuing specific functional objectives such as education (edugames), physical activity (exergames), or professional training. Current systematic reviews strongly support the use of technology-driven SGs for managing ADHD symptoms in children, confirming their utility as a scalable digital therapeutic approach. Research focuses on the neurophysiological impact of SGs, comparing memory-based and complexity-based brain activity between subjects with ADHD and healthy controls during gameplay. Ultimately, current literature positions SGs and related digital proposals as powerful tools for personalized skill acquisition, supported by increasing empirical evaluation (Bernardelli, G. et al., 2021; De Luca, V. et al., 2025; Giannaraki, M., 2021).

La Bottega del Vasaio - A VR experience to promote the samnite culture

The Potter's Workshop (*La Bottega del Vasaio*) was conceived as an edugame set in an VR environment to pursue two primary and interdependent objectives:

1. *Cultural Heritage Promotion*: To valorize the material and procedural cultural heritage of the ancient Samnite people by tracing the production phases of specific types of pottery used between the 6th and 2nd centuries BCE.
2. *Inclusive Didactic Innovation*: To leverage the potential of VR and Artificial Intelligence (AI) to facilitate learning in an inclusive manner, with a specific focus on enhancing self-regulation skills and Executive Functions (EFs)

To ensure coherence between the didactic aims and the target population, the design phase required a systematic analysis of the Universal Design for Learning (UDL) guidelines and the criteria for inclusive Game Design tailored for ADHD. In parallel, the collection and production of scientifically accurate textual and multimedia content were essential. The resulting virtual experience aims to immerse the user in an authentic context where, upon launching the application, the learner assumes the role of an apprentice on their first day of work at the workshop of a master potter.

Development of the virtual environment

The virtual environment was realized by balancing a cartoon style with the necessary historical and geographical accuracy. The master potter's main workshop is situated in a landscape that emulates the rich vegetation of the Apennine inland of Southern-Central Italy, the area occupied by the Samnite people at the peak of their expansion. The Samnite territorial organization differed from that of the Magna Graecia city-states or the Romans of Latium, characterized by a heterogeneous and dispersed settlement system, featuring fortified centers (*oppida*) and rural structures, which makes the discovery of specific production sites complex.

Despite the lack of a unique archaeological exemplar for a potter's workshop dating back to that period, the virtual building was conceived as a hypothetical reconstruction of a functional structure of the era. For its rectangular layout, architectural inspiration was drawn from excavations described in the work of De Benedittis, particularly those relating to the Horreum, a building used for storing agricultural products found, for example, in the area of the ancient settlement of *Allifae*.

Due to the limited availability of free digital assets that were archaeologically coherent with the Samnite environment, the workshop's characterization was based on the 3D scanning of authentic archaeological finds, such as the most characteristic pottery types (Oinochoe, Olpe, Crater, Skyphos, and Amphora) preserved at the Carife Museum and the National Museum of Sannio Caudino in Montesarchio. These new digital assets, obtained through a long 3D modeling process (Figure 12), were essential not only for enriching the environment but also as positive models for the creation of derived assets, such as the stone molds required for clay forging.

Many other objects were recreated from scratch based on studies of information found in literature. Specifically, two vertical kilns with a rectangular plan were placed inside the workshop. Despite their lower efficiency compared to similar circular kilns, the rectangular model was chosen as it was simpler to construct and presumably more common. The kilns feature two chambers (a combustion chamber and a firing chamber), separated by a perforated floor, and the colors chosen for the texturing faithfully represent the exterior masonry made of stone and the interior made of brick.



Figure 12. Some of the digital assets obtained from 3D scans of the artefacts preserved in the Carife Museum.



Figure 13. Desktop mode opening scene and menu

Design of the virtual experience

The application was designed to ensure versatile accessibility, being compatible with both mid-range desktop systems and Virtual Reality (VR). For the VR mode, the use of entry-level

headsets was prioritized, selected for their affordability and growing presence as standard educational equipment in primary and secondary schools.

While the virtual experience remains consistent across both versions, the interactions vary significantly:

- **Desktop Mode:** The user selects and manipulates objects using a mouse and keyboard, utilizing directional keys for camera movement.
- **VR Mode:** Interaction occurs naturally through spatial controllers, allowing the user to physically grab and move assets. The view is managed by a head-tracking system, which synchronizes the camera orientation with the user's head movements, significantly enhancing the sense of immersion.

Upon launching from the main menu, the user is introduced into a fenced courtyard outside the workshop—a space designed to focus exploration on the adjacent work environment. Once across the threshold, the apprentice is greeted by an AI-based NPC (Non-Playable Character) representing the master potter. Interaction with the NPC can occur via voice or text input. The master potter not only guides the user through production but also provides contextual hints if the learner is unsure how to proceed and validates the successful completion of each stage.

The production cycle is structured into the following sequential phases:

1. *Clay Selection:* The user selects the raw material from three varieties historically documented for the Samnite civilization: pink, hazel, or red. This choice affects both the aesthetics and the perceptible properties of the final artifact.
2. *Wetting:* The clay is moistened to achieve the plasticity necessary for crafting, emulating real-world material preparation techniques.

3. *Laying in the Mold*: The prepared clay is pressed into matrices or molds to shape various types of pottery, such as amphorae, oinochoai, plates, or cups.
4. *Firing*: The artifact is placed inside a vertical kiln. The user must monitor the process until the object reaches the ideal color and consistency for the subsequent decoration phase.
5. *Decoration*: In this stage, the black-paint (black-gloss) technique, a hallmark of Samnite ceramic art, is applied. The user can paint freehand with virtual brushes or apply decals featuring classic iconographic motifs, such as warriors or horses.
6. *Exhibition*: Once completed, the pot is placed outside the workshop for display, ideally closing the artisanal process.

At the end of the session, the user can choose to start a new production or deepen their knowledge of Samnite culture through an open dialogue with the master potter.



Figure 14. One of the "crater" vases is about to be placed in one of the two kilns. At the bottom right, the chat with the master potter's suggestions.



Figure 15. After the firing process, the vase can be decorated using a brush or by applying decals.

Implementing recommendations for inclusive design for students with ADHD

The inclusive design of La Bottega del Vasaio incorporates a sophisticated array of strategies and technical solutions specifically engineered to support the educational needs of students with ADHD, operating in strict adherence to both the principles of Universal Design for Learning (UDL) and Cognitive Load Theory. A cornerstone of this approach is task decomposition, which effectively segments the entire production cycle into six sequential and clearly defined phases; this methodological choice significantly reduces extraneous cognitive load by transforming what could be an overwhelming process into short, manageable blocks that prevent cognitive saturation and enhance the learner's executive functioning.

The interaction between the player and the NPC (Non-Player Character) is governed by an Artificial Intelligence system that serves as the digital entity's cognitive core. Utilizing the Convai platform, the role of the NPC evolves from a mere descriptive element into an expert guide on the ancient Samnite civilization. During the design phase, it was possible to

customize the knowledge base and configure the parameters necessary to direct the generation of both natural language and its corresponding audio outputs.

Specifically, the materials populating the system's knowledge base were selected and validated by museum experts and directors involved in the project. The accuracy and priority of information are ensured by the integration of Retrieval-Augmented Generation (RAG), which allows the AI system to anchor its responses to certified documentary sources. However, communicative effectiveness does not depend solely on the quality of the baseline content, but rather on the ability of both interlocutors to establish a shared communicative ground. While the player retains the freedom to speak openly to the NPC, the entity pursues a complex pedagogical objective: rather than simply describing pottery production, it aims to adapt the virtual experience to the diverse cognitive configurations of the students.

To achieve this, the interaction architecture integrates a narrative design tool based on a state tree. This structure allows the master potter to modulate his behavior in real-time; the entity can recognize successes or failures within the game, offering rewarding feedback or targeted encouragement. Such a strategy provides the system with extreme versatility: should the target audience change, the interactions could be easily restructured by modifying the narrative nodes while keeping the underlying language generation models intact. Consequently, this tool provides a robust dimension for flexible, adaptive, and inclusive didactic action

To further maximize engagement and emotional connection, the NPC is equipped with head and eye tracking functionalities that allow it to respond visually to the presence and movements of the learner, thereby creating a more anthropomorphic and compelling interaction that grounds the user in the virtual task. From a sensory perspective, the VR environment is meticulously configured as a controlled space where external distractors are systematically minimized, ensuring that the student's limited attentional resources are focused entirely on the specific learning objectives. A key feature of the user interface is the programmatic integration of a vignette effect, which serves as a functional tool to reduce

peripheral visual noise and direct selective attention toward the central manipulative tasks, effectively acting as a visual funnel for the student's gaze. Finally, the application provides comprehensive personalization options, such as the fine-tuning of light saturation and precise audio control via a dedicated mixer, allowing students with sensory hypersensitivity to calibrate the experience to their specific tolerance thresholds, which in turn encourages the development of self-regulation skills within a protected, stimulating, and ultimately empowering virtual educational context.

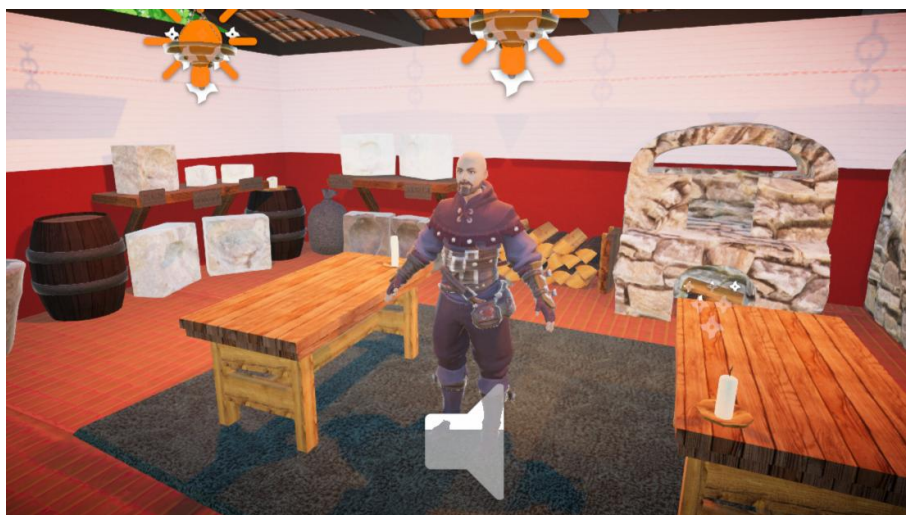


Figure 16. the NPC of the master potter who guides the production process of the vases according to the traditional samnite method.

MUSEum Editor for a sharable peer-to-peer museum

The MUSEum Editor (MUSE) application stands as a cutting-edge Virtual Reality (VR) solution designed to reshape museum education by offering an interactive and highly customizable experience. MUSE is specifically targeted at secondary school teachers and students, providing them with the tools to emulate a virtual museum classroom. This environment allows for the design of learning pathways tailored to specific didactic needs and the unique cognitive profiles of individual students. An in-depth exploration of the Vygotskian perspective allows the transition from Hein's descriptive museum model to the constructivist one to be interpreted not merely as a methodological variation, but as a

necessary evolution of the educational-didactic framework. According to Lev Vygotsky (1978; 1991), cognitive development is not an isolated process but a phenomenon deeply rooted in social interaction and cultural mediation. In this sense, the concept of the Zone of Proximal Development becomes the key to interpreting the experimentation: the virtual environment does not merely display an artifact; it acts as a scaffold (Wood, Bruner, & Ross, 1976) that enables the student to perform operations that would otherwise be out of reach. This vision finds its operational application in "La Bottega del Vasaio," where the learner's role shifts from passive observer to active maker. By interiorizing cultural meanings through the digital manipulation of artifacts, the student confirms the idea that human activity is constantly mediated by both material and psychological tools (Wertsch, 1985). The redesign of the experience through the "MUSeum Editor" further extends this principle, transforming museum curation into an activity of social meaning-making. In this context, students do not simply learn archaeological facts; they participate in a true community of practice where knowledge is negotiated and shared. The ultimate goal of this transition is the integration of educational strategies that view heritage not as a static object to be described, but as a terrain for dynamic interaction. By transforming students into creators and curators, the project fully adheres to the socio-constructivist imperative of shifting focus from an individual's internal processes to social dynamics of mutual support, making the Educational Metaverse an environment where learning emerges through collaboration and active participation.

The Editor Mode and Cultural Heritage Digitization

In its Editor Mode, the application grants teachers full control over the configuration of the virtual space. This is achieved through the selection and placement of high-fidelity 3D objects, which serve as accurate digital reproductions of artifacts and artworks from participating museums in the Campania region. This operational methodology ensures that students can interact with local cultural heritage in an immersive and manipulative manner, transcending the purely observational nature typical of physical visits and promoting a deeper, more participatory understanding of history and art.

The added value of MUSE extends beyond the mere exhibition of digital artifacts; the application offers teachers an additional layer of customization by allowing for the seamless upload and integration of self-produced didactic materials in various formats, including videos and images. Once uploaded to the system's designated subfolders, these contents are automatically recognized and made available within the Editor. This intuitive integration of personal content exponentially enriches the virtual visit, merging the teacher's specific

resources with those provided by the museum database. This creates a highly flexible and stimulating learning environment that adheres to the principle of the structured alternation of didactic mediators.

Within this self-produced virtual scenario, students are encouraged to engage in autonomous exploration and self-directed learning. They can observe museum objects from multiple angles, examine close-up details without the physical constraints of display cases, and access supplemental multimedia content, such as audio descriptions, explanatory videos, or high-resolution images, in a contextualized manner.

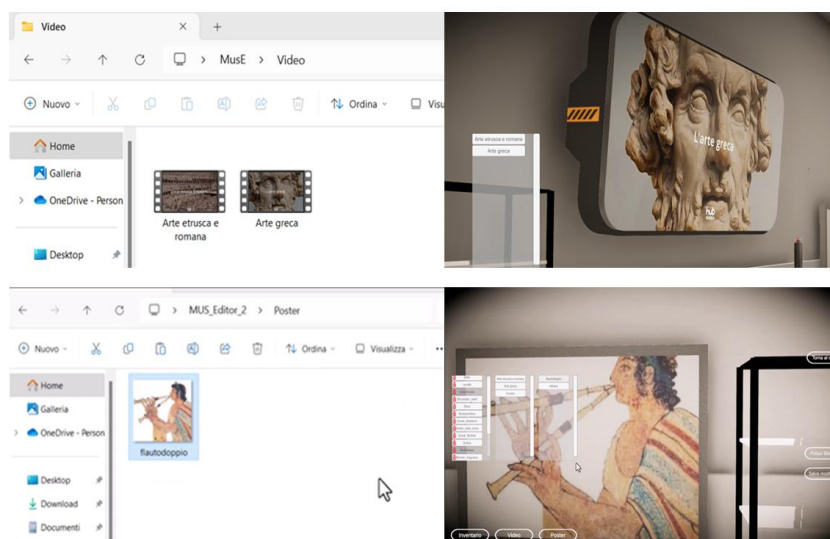


Figure 18. The files chosen by the teacher or students, once copied into the respective Video and Poster subfolders of MUSE, are automatically loaded into the application and made available in Editor mode

From Spectators to Architects: Empowering Students via Virtual Museum Co-Creation

Employing a framework rooted in Action-Research and Design-Based Research, this project aims to radically reimagine the museum-based educational journey. As a highly versatile and interactive digital space, MUSE not only ensures democratic access to cultural heritage by dismantling geographical and economic barriers, but also promotes hands-on,

experiential discovery. This process cultivates essential skills required in contemporary society. Such a pedagogical model, which prioritizes the user's needs throughout the design process, resonates with current digital education directives, boosting engagement and facilitating the construction of meaningful knowledge. Furthermore, MUSE enhances the sustainable preservation of cultural heritage by offering novel engagement methods with the past. In this light, the virtual environment serves as a groundbreaking and scalable blueprint for cultural dissemination and social inclusion.

MUSE's impact reaches learners directly, as the Editor's structure enables faculty-guided cohorts to develop bespoke thematic trajectories autonomously. Within this framework, the design of tailored digital assets and their subsequent integration into the environment serves as an effective strategy for cooperative and creative growth. This method successfully shifts participants from mere spectators to collaborative architects of museum content. The curriculum is structured around four progressive, integrated stages that guide participants from foundational insights to practical application, and from content production to shared dissemination, with the entire cycle estimated to last between six and nine weeks.

Phase 1: Foundation and Engagement (2-3 Weeks)

The initial stage, Foundation and Engagement, spans two to three weeks and focuses on establishing a solid knowledge base. It begins with an interactive lecture introducing the Samnite civilization, providing a chronological and spatial framework for their sites, customs, and social organization. To facilitate a clear grasp of the subject matter for all, instructors utilize maps, videos, and simplified texts. This is followed by mentored inquiry where participants, organized into peer-based teams, delve into specific topics such as diet, craftsmanship, or religion using a mix of print and multimedia resources. This part concludes with a classroom discussion, an essential moment for sharing discoveries and establishing a common historical context before transitioning to digital immersion.

Phase 2: Experiential Core – VR and AI (1-2 Weeks)

The subsequent stage covers activities with VR and AI over one to two weeks and represents the experiential core of the program, where these tools serve as primary instruments for discovery. The research group introduces these applications by focusing not only on

technical usage but also on the principles of accessibility that guided their design. The immersive sessions occur in shifts under close supervision to manage sensory and cognitive impact. Inclusion is a priority; therefore, individual support measures are implemented for students with specific needs, such as ADHD, to minimize sensory overload. Mentorship among peers is fostered, allowing those who have completed the task to assist their classmates, which strengthens social competence and mastery of the material while transforming the classroom into a horizontal community.

Phase 3: Synthesis and Creation (2-3 Weeks)

During the third stage, Synthesis and Creation, students transition from consumption to the generation of original material. Groups focus on crafting 3D objects, videos, and images to curate a digital gallery space. This phase is crucial for the practical verification of both technological and cultural literacy. Specific tools are employed, such as the "La Bottega del Vasaio" application for 3D modeling—which helps learners grasp the functionality of artifacts like the Greek craters discussed previously—and the MUSE platform for final assembly. Building this virtual space represents the zenith of cultural competence, as participants reflect on the communicative and exhibition value of historical legacy.

Phase 4: Dissemination and Assessment (1 Week)

The concluding stage is a one-week period focused on Dissemination, where results are shared and evaluated through a multidimensional approach. Each group formally presents its gallery room to the class, transforming the occasion into a formal showcase that may also include parents or guest visitors. This practice reinforces communication skills and validates the effort invested throughout the project. The assessment process is comprehensive, beginning with peer review where students use a simplified rubric focused on originality, clarity, and completeness to encourage critical thinking. Finally, the instructor conducts a summative evaluation of the entire process, considering individual participation, team effectiveness, the thematic quality of the output, and the ability to apply acquired knowledge across various media.

The integration of Generative Artificial Intelligence (GAI) into the Educational Metaverse, while offering unprecedented potential for dynamic content creation, raises crucial concerns regarding the technological "flip side": the risks of hallucinations and poisoning. Within the context of the proposed Learning Unit, these risks have been mitigated through a rigorous control structure designed to preserve the scientific integrity of the educational content.

Specifically, the virtual production processes in both "La Bottega del Vasaio" and the "MUSEum Editor" (MUSE) are bound to three-dimensional models derived from 3D scans of actual artifacts. This ensures historical and geographical coherence, effectively precluding the generation of incorrect information typically associated with the semantic interpolation of AI, or hallucinations. Furthermore, the decoration phase is restricted to decals that are philologically consistent with Samnite, Etruscan, and Magna Graecia cultures. A central point of this discussion involves supervision: to prevent the real-time flow from broadcasting deviant or intentionally manipulated content (poisoning), the framework incorporates a constant pedagogical oversight. Before the transition to the exhibition phase in MUSE, the teaching team evaluates the digital artifacts produced by the students. Although the current necessity for manual configuration of 3D objects by developers represents a limitation to the workflow's fluidity, it serves, at this stage, as a fundamental safety filter. This design choice directly addresses the need to monitor the output of modern *text-to-3D* or *image-to-3D* tools. Without supervision, these systems could return distorted plastic or semantic models that would compromise the validity of the educational path. Human-in-the-loop control and the pre-selection of digital assets act as a barrier against the generation of incorrect information, ensuring that the virtual environment remains a protected and scientifically reliable learning space.

Bibliography (An inclusive VR/AI learning pathway)

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing

Barkley, R. A. (1997). *ADHD and the nature of self-control*. Guilford press.

Barkley, R. A. (2006). A handbook for diagnosis and treatment. *New York: Guilford Publications*, 3-75.

Barkley, R. A. (2015). Definitional and historical aspects of attention deficit hyperactivity disorder (ADHD). In M. D. Rucklidge & A. S. Tannock (Eds.), *Attention Deficit Hyperactivity Disorder: A Handbook for Clinical Practice*. Oxford University Press.

Bernardelli, G., Flori, V., Greci, L., Scaglione, A., Zangiacomi, A. (2021). A Virtual Reality Based Application for Children with ADHD: Design and Usability Evaluation. In: De Paolis, L.T., Arpaia, P., Bourdot, P. (eds) *Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2021. Lecture Notes in Computer Science()*, vol 12980. Springer, Cham. https://doi.org/10.1007/978-3-030-87595-4_27

De Luca, V., Schena, A., Covino, A. et al. Serious Games for the Treatment of Children with ADHD: The BRAVO Project. *Inf Syst Front* 27, 841–863 (2025).
<https://doi.org/10.1007/s10796-023-10457-8>

Giannaraki, M., Moumoutzis, N., Papatzanis, Y., Kourkoutas, E., & Mania, K. (2021, April). A 3D rhythm-based serious game for collaboration improvement of children with attention deficit hyperactivity disorder (ADHD). In *2021 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1217-1225). IEEE.

Still, G. F. (1902). Some abnormal psychical conditions in children. *Lancet*.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

Vygotsky, L. S. (1991). *Pedagogicheskaya psikhologiya* [Pedagogical psychology]. Moscow, Pedagogika Publ.

Wender, P. H. (2000). *Attention-deficit hyperactivity disorder in adults*. Oxford University Press.

Wertsch, J. V. (1988). *Vygotsky and the social formation of mind*. Harvard university press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2), 89-100.

Inclusion vs. Intrusion: Biometrics for evaluating emotional component in the learning process

The emotional component in the learning process

Theories of multiple intelligences have challenged the reduction of human intelligence to a single metric such as IQ, emphasizing instead the coexistence of distinct but interrelated capacities, including interpersonal and intrapersonal intelligences that are directly linked to the understanding and regulation of emotions in oneself and others. In Gardner's perspective (Gardner, H., 1983), these forms of intelligence constitute foundational resources for interpreting social situations, guiding one's inner life, and orienting behavior in complex environments, thereby making the affective and relational dimensions structurally integral to any conception of learning. Building on this framework, the construct of emotional intelligence, as articulated by Salovey and Mayer (1990) and subsequently popularized by Goleman (Goleman, D., 1995), places empathy, self-knowledge, and self-regulation at the center of human adaptive functioning, emphasizing how the ability to recognize, label, and modulate emotions supports both personal well-being and prosocial conduct.

Within this horizon, emotions are not residual or disruptive elements but primary organizers of attention, motivation, and engagement, with direct repercussions on the quality of classroom relationships and on students' readiness to learn. As Goleman argues, emotional competencies such as impulse control, perseverance, and the capacity to manage frustration shape the conditions under which cognitive resources can be effectively mobilized, especially in situations of challenge or failure. In an inclusive school, these dimensions translate into a pedagogical responsibility to recognize and cultivate diverse profiles of intelligence and emotional competence, particularly for learners whose developmental trajectories, such as those of students with ASD, require differentiated pathways, explicit scaffolding of socio-emotional skills, and systematic attention to affective signals expressed in interaction.

From an educational standpoint, a high level of understanding of the emotional state becomes crucial both for the teacher's capacity to "hold" the class and for the success of peer-to-peer learning strategies. In Rivoltella's analysis, the ability to "hold" the class is closely linked to the teacher's sensitivity in perceiving the emotional climate, managing tensions, and transforming moments of disengagement or conflict into opportunities for learning, thus integrating instructional and relational competences within a single professional profile (Rivoltella P.C., 2014). Bonaiuti, Calvani and Picci (2014) highlight how peer-to-peer and collaborative learning arrangements work when they maintain a dynamic equilibrium between competence and recognition: students must feel both adequately challenged and acknowledged in their experiences, a balance that depends in large part on the mutual perception and negotiation of emotions within the group.

In this perspective, the analysis of emotional manifestations through multiple channels such as facial expression, voice, gesture, body posture, and other behavioral indicators, can offer additional information to support more responsive and adaptive instructional decisions. Biometric-based techniques, such as automatic emotion recognition from face and voice, exemplify how multimodal observation can be technologically mediated, providing teachers and systems with cues that might otherwise remain implicit or difficult to monitor continuously (Bilotti et al., 2024). However, the chapter explicitly rejects the idea of transforming emotion detection into a deterministic or evaluative device: rather than "measuring" or "predicting" a student's inner state in a definitive way, affective indicators are conceived as one resource among others in a broader ecology of assessment and pedagogical reflection.

Consequently, the ultimate goal is not the precise quantification of emotional states but the use of affective information to guide the design of flexible content, representations, and engagement options, in opposition to the conception of an "average student" criticized by Universal Design for Learning (Aiello, P. et al., 2014). As Savia (2016) underscores, learning processes unfold in unique ways for each learner, depending on the interplay between contextual variables and personal interests, and therefore cannot be adequately supported by a single type of content, delivered in a single form, for a single purpose. In this light, the

integration of multimodal emotion analysis, whether conducted by human observers or mediated by AI systems, serves to make more visible the diversity of learning-related emotional trajectories, enabling teachers to adjust tasks, supports, and relational strategies in ways that foster participation and inclusion, especially for students whose emotional expressions, such as those of learners with ASD, may diverge from neurotypical expectations while remaining equally meaningful.

Biometric approach to emotion recognition

The use of biometric information for automatic emotion recognition represents a key technological component in the construction of inclusive, affect-aware environments within the Educational Metaverse. In this framework, Multimodal Emotion Recognition is conceived not merely as a technical exercise in pattern classification but as a methodological contribution aimed at understanding how different biometric channels, primarily facial expression and voice, can be combined to infer emotional states that are pedagogically relevant for monitoring and supporting students' learning processes. The work discussed in this section is based on Bilotti et al. (2024), which systematically compares several deep learning strategies for multimodal emotion recognition using convolutional neural networks (CNNs) with multiple inputs. In this section we reinterpreted it in light of its potential application to inclusive educational scenarios.

Theoretical framework and problem statement

Automatic emotion recognition has traditionally been framed as a classification problem in which each instance (an image, a video segment, or an audio fragment) is assigned to one of a set of discrete emotional categories, such as happiness, sadness, anger, fear, surprise, or disgust (Bilotti et al., 2024). This formulation is closely linked to the discrete theory of

emotions, whose most influential contribution is the work of Ekman, who identified a limited set of “basic emotions” characterized by specific, cross-culturally recognizable facial expressions (Ekman P., 1992; Ekman P. & Friesen W. V., 1971). In Ekman’s model, emotions such as happiness (marked by raised cheeks and crow’s feet wrinkles), sadness (downturned mouth and inner eyebrow raise), anger (lowered brows and pressed lips), fear (raised eyebrows and wide-open eyes), surprise (raised eyebrows and dropped jaw), and disgust (wrinkled nose and raised upper lip) are considered evolutionarily grounded “families” of affective states, each associated with prototypical facial configurations that can be reliably identified across different cultures (Ekman P. & Friesen W. V., 1978). Later extensions of this taxonomy have included additional states such as contempt (unilateral lip curl), calm, or neutral baselines, broadening the set to eight or more categories while preserving the core assumption of universality (Ekman P. & Cordaro D., 2011).

Among the various theoretical models of emotions, the discrete approach, widely adopted as a standard in the literature on automatic emotion recognition, partitions the emotional space into a finite set of classes, consistent with Ekman’s taxonomy (Ekman P., 1992; Clavel, C., & Callejas, Z., 2016). From the standpoint of automatic emotion recognition, this theoretical framework provides a natural way to discretize the emotional phenomenon by mapping continuous, multifaceted affective experiences onto a finite set of class labels, typically through annotation protocols where human raters assign each sample to one primary category based on observed facial, vocal, or multimodal cues (Clavel, C., & Callejas, Z., 2016). Discretization involves defining a closed vocabulary of emotion categories, often Ekman’s basic set or a subset thereof, and establishing consistent labeling criteria that account for prototypical expressions while handling edge cases like blended emotions or low-intensity displays (Ekman P., 1992). Bilotti et al. follow this established convention by explicitly selecting a subset of basic emotions (anger, disgust, fear, happiness, and sadness) that are consistently represented across their multimodal datasets (RAVDESS and BAUM-1), ensuring each video–audio instance can be associated with a well-defined target class for supervised training of convolutional neural networks (Bilotti U. et al., 2024).

This choice, although it does not exhaust the complexity of emotional phenomena, such as mixed states, cultural display rules, or context-dependent interpretations, is particularly suited to the design of machine learning systems because it allows the definition of clear labels, facilitates dataset construction and annotation, and supports the integration of emotion recognition modules into interactive applications, such as intelligent tutors or educational games. The work also acknowledges the limitations of purely discrete models, recalling how the same facial configuration may correspond to different internal states depending on personality, context, and cultural norms, and how dimensional models (e.g., valence–arousal circumplex) offer complementary insights by representing emotions along continuous axes of pleasantness and activation rather than categorical boundaries (Posner, J., Russell, J. A., & Peterson, B. S., 2005; Russell J. A., 1980). Nonetheless, for the specific goal of building and evaluating multimodal CNN-based classifiers, the authors focus on Ekman-inspired discrete categories, as they enable direct performance benchmarking and practical deployment.

Within the Educational Metaverse perspective, this operationalization, rooted in Ekman's discretization, allows mapping emotionally salient events such as frustration (potentially akin to anger or sadness), confusion (fear-like uncertainty), or engagement (happiness or surprise) onto recognizable patterns that can trigger adaptive responses by the system or teacher, without assuming that these labels fully capture the richness of the learner's experience. By leveraging biometric cues aligned with basic emotion families, Multimodal Emotion Recognition supports inclusive pedagogies that respond to affective signals in real-time, fostering environments where emotional literacy is scaffolded through technology without reducing students, especially those with ASD, to simplistic categorical profiles.

Multimodal biometric cues

Emotion recognition in humans relies on a complex integration of visual, auditory, and bodily cues that unfold dynamically over time (Poria S. et al, 2017). In line with this observation, the biometric approach explicitly leverages multiple modalities: facial

appearance extracted from video frames, motion information encoded as optical flow between consecutive frames, and vocal characteristics represented as Mel spectrograms derived from speech signals. Each of these channels conveys complementary information: static facial features capture the configuration of key landmarks (e.g. eyebrows, eyes, mouth), optical flow provides a representation of facial motion patterns over time, and the audio channel expresses prosodic and spectral markers of emotion, such as pitch, intensity, and voice quality.

Each of these channels conveys complementary information: static facial features capture the configuration of key landmarks (e.g. eyebrows, eyes, mouth), optical flow provides a representation of facial motion patterns over time, and the audio channel expresses prosodic and spectral markers of emotion, such as pitch, intensity, and voice quality.

From a technical point of view, the transition from unimodal to multimodal emotion recognition is motivated by the limited robustness and generalizability of single-channel systems, which often suffer from sensitivity to noise, occlusions, variations in lighting, or language-specific acoustic patterns (Pan B. et al, 2023; Poria S. et al., 2017). Confusion matrix Multimodal fusion aims to mitigate these limitations by allowing the model to exploit redundancies across channels and to recover informative features when one modality is degraded or missing, as demonstrated by recent works employing attention-based fusion and residual connections to dynamically weight modalities and preserve original representations (Wu T. et al., 2025). For educational applications, this is particularly relevant: a student may avert gaze but maintain a distinctive vocal prosody, or may produce very subtle facial expressions that are difficult to capture in low-resolution settings, whereas a combination of face and voice increases the probability of correctly identifying the underlying emotional state, supporting adaptive e-learning platforms and real-time affective feedback (Bhardwaj P. et al, 2021).

Datasets: RAVDESS and BAUM-1

To investigate the effectiveness of different multimodal strategies, Bilotti et al. use two widely adopted audiovisual datasets that include both video and audio signals: the Ryerson

AudioVisual Database of Emotional Speech and Song (RAVDESS) and the BAUM-1 database.

RAVDESS (Livingstone S. R., & Russo F. A., 2018) consists of 7,356 recordings of professional actors (12 male and 12 female) who produce speech and song clips in English with acted emotional content. Each actor performs multiple trials for a set of basic emotions (calm, happy, sad, angry, fearful, surprise, and disgust) and neutral, with two intensity levels (normal and strong) for most categories, and clips are provided in three modalities: full audiovisual, video-only, and audio-only. For the purposes of Bilotti et al., a subset of emotions is selected (anger, disgust, fear, happiness, and sadness), and both facial frames and Mel spectrograms are extracted from the AV clips. The videos are characterized by controlled conditions, including uniform backgrounds and consistent framing, which simplifies face detection and segmentation but may introduce biases that limit generalization to more naturalistic scenarios. Examples of RAVDESS images are in Fig.1.

BAUM-1 (Zhalehpour S. et al., 2017), by contrast, is a multimodal database collected in Turkish, containing videos of 31 participants (17 men and 14 women) whose emotional responses are elicited by viewing carefully designed visual stimuli (images and short clips) that induce a variety of affective states. The dataset distinguishes between acted and spontaneous expressions, thereby providing a more ecologically valid representation of emotional behavior, albeit at the cost of increased variability and noise. The target emotions include the six basic categories (happiness, anger, sadness, disgust, fear, and surprise), as well as boredom and contempt, and recordings are accompanied by audio channels that reflect natural speech in Turkish. Compared to RAVDESS, BAUM-1 presents lower image resolution, differences in background color (e.g. bright green), and linguistic diversity, all of which pose additional challenges for cross-dataset generalization and highlight the importance of designing models that do not overfit to specific visual or acoustic conditions. Examples of BAUM-1 images are in Fig.2.

The joint use of RAVDESS and BAUM-1 allows the authors to test not only the performance of multimodal models within each dataset but also their capacity to generalize across languages, visual settings, and expression types (acted vs spontaneous). This is particularly meaningful in an educational context, where learners differ in cultural background,

communication style, and expressivity, and where emotion recognition systems must remain reliable despite such heterogeneity.

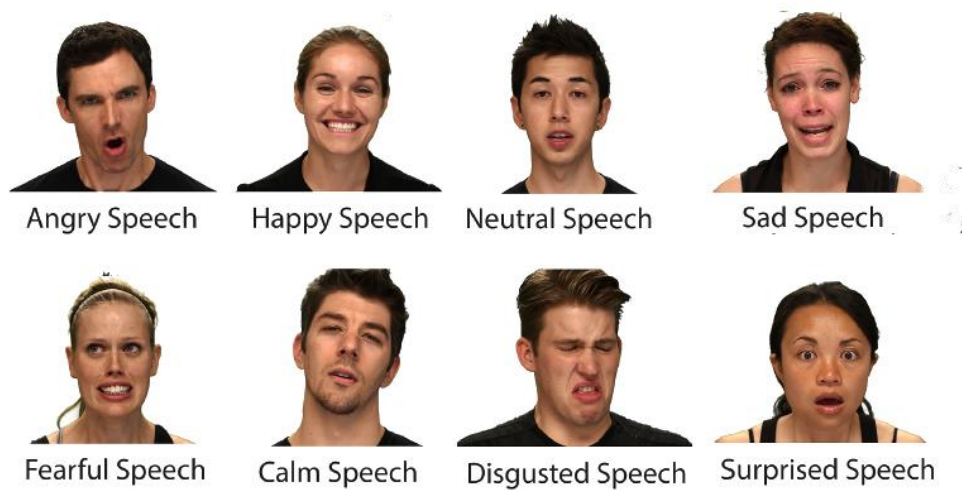


Figure 19. Examples of RAVDESS images



Figure 20. Examples of BAUM-1 images.

Model architectures: from unimodal to multi-input CNNs

The core contribution of Bilotti et al. lies in the design and comparison of several CNN-based architectures that process one or more modalities, as well as in the exploration of different strategies for fusing these modalities. The starting point is a reduced Convolutional Neural Network model built in Keras (a deep learning API), inspired by VGG-like architectures (Simonyan K. & Zisserman A. 2014) but simplified to allow faster training on conventional hardware. For the visual channel, the network processes individual frames extracted from videos, rather than stacks of frames, in order to reduce memory requirements and training time; optical flow frames, computed from pairs of consecutive frames using methods such as Lucas–Kanade (Lucas B. D. & Kanade T., 1981; Alfarano A. et al., 2024), are treated as a separate visual modality.

The unimodal visual model typically consists of a sequence of convolutional layers (Conv2D) with 3×3 kernels and increasing numbers of filters (e.g. 32, 64, 128, 256, 512), each followed by max pooling operations (MaxPooling2D) to progressively reduce spatial resolution while retaining salient features. Dropout layers (e.g. with rates 0.25–0.5) are inserted to mitigate overfitting, and the resulting feature maps are flattened and passed through fully connected layers of size 2,048 and 1,024 before reaching a final softmax layer that outputs posterior probabilities over the target emotion classes. This architecture is used to train separate models for facial frames and optical flow images, enabling a direct comparison of their discriminative power.

Building on this baseline, the authors propose multi-input architectures that implement early fusion strategies by concatenating intermediate feature representations from different modalities. In a two-input configuration, one branch of the network processes facial frames, while another branch processes optical flow frames; both branches share a similar convolutional structure up to a certain depth, after which their feature maps are concatenated along the channel dimension and fed into a shared sequence of dense layers responsible for the final classification. This design allows the network to learn joint representations that capture both spatial and temporal aspects of facial movement,

theoretically enhancing its ability to distinguish between visually similar but dynamically different expressions.

The two input CNN is shown in Fig. 3.

A third input corresponding to the audio channel, represented as Mel spectrograms (Davis S. & Mermelstein P., 1980; Chowdhury J. H., Ramanna S. & Kotecha K. 2025) computed from the speech signal associated with each video clip. In this three-input setting, a dedicated CNN branch processes the Mel spectrograms, extracting time–frequency patterns that encode prosodic and spectral markers of emotion; these features are then fused with those from the two visual branches (frames and optical flow) using intermediate fusion, so that the final classifier operates on a rich, multimodal representation of each instance. By systematically comparing the performance of one-input, two-input, and three-input models, the work quantifies the added value of each modality and provides empirical evidence for the benefits of multimodal fusion.

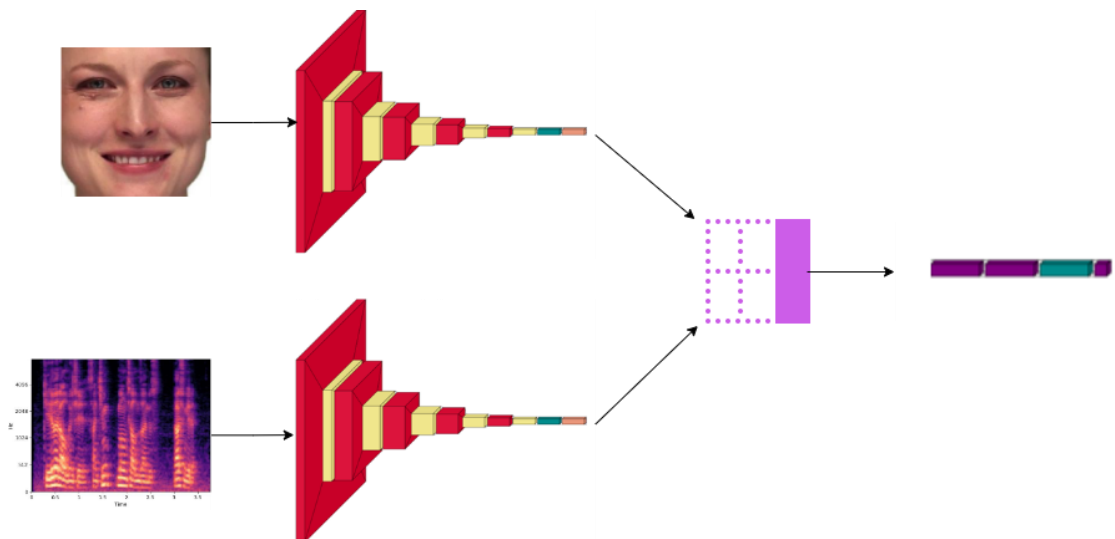


Figure 21. The 2-way CNN, for frames and optical flow

Fusion strategies and training protocol

The architectures implemented in the work realize an intermediate form of early fusion: each modality is processed by its own convolutional sub-network up to a certain depth, and the resulting feature maps or latent vectors are concatenated at an intermediate layer, allowing the shared part of the network to learn higher-level, cross-modal representations. This design benefits from the representational power of deep networks, which can abstract away from raw data peculiarities while still exploiting temporal alignment between video frames, optical flow, and the corresponding Mel spectrograms.

Training is conducted with standard supervised learning protocols: datasets are split into training, validation, and test sets, and models are optimized using gradient-based methods with categorical cross-entropy as the loss function. The simplified CNN architecture processes individual frames independently rather than video stacks, reducing memory requirements from over 20 minutes per epoch (full VGG16) to 60 seconds per epoch, with total training time of approximately 40 minutes for 40 epochs. Facial frames are cropped to 150×110 pixels and resized to $227 \times 227 \times 3$, while optical flow images, computed via Lucas-Kanade from 16-frame segments, are normalized and similarly resized; audio Mel spectrograms are extracted from speech signals and processed through dedicated CNN branches before multimodal fusion. Dropout rates of 0.25 and 0.4 were empirically optimized to prevent overfitting, with the final architecture achieving acceptable offline training times suitable for educational platform prototyping.

Experimental results and performance analysis

Results of the conducted experiments are shown in Fig. 4 and 5. The results show that unimodal CNNs trained solely on facial frames achieve relatively high accuracy, outperforming models based on optical flow alone, which indicates that static or quasi-static facial configurations already contain a significant portion of the discriminative information needed for emotion classification. For instance, at 40 training epochs on BAUM-1 acted subsets, frame-based models reach substantially higher accuracy than optical-flow-based models, highlighting the greater difficulty of extracting stable features from motion representations in this specific setting.

The two-input models that combine facial frames and optical flow via early fusion do not consistently surpass the performance of frame-only models, suggesting that, in the tested configuration, the optical flow channel either does not add sufficiently informative cues or is not exploited optimally by the concatenation mechanism. Nevertheless, experiments on spontaneous subsets and with different numbers of epochs indicate that the multimodal architecture remains competitive and, in some conditions, yields improvements over reference methods in the literature. In particular, the proposed approach achieves higher accuracy than earlier works on BAUM-1 spontaneous videos, with relative improvements on the order of tens of percentage points compared to some prior baselines using similar visual channels.

The inclusion of the audio modality through Mel spectrograms further enhances classification performance, especially when evaluating on RAVDESS, where prosodic and spectral markers of emotion are strongly present due to the controlled, acted nature of the recordings. Three-input CNNs that jointly process video frames, optical flow, and Mel spectrograms reach accuracy levels around or above 95% on certain configurations, demonstrating the effectiveness of multimodal fusion in leveraging complementary cues across face and voice. Cross-dataset experiments, in which models trained on one dataset are tested on the other, also indicate that combining audio and visual information can mitigate some of the performance degradation caused by differences in language, recording conditions, and expression style, even though generalization remains a challenging open problem.

Type of input \Epochs	10	20	30	40
F	0.933	0.9285	0.9316	0.9529
Opt	0.7038	0.7064	0.7021	0.7011
Mel	0.7423	0.9086	0.9306	0.9439
2-input : F & Mel	0.9205	0.9344	0.952	0.9595
3-input	0.9	0.9289	0.9366	0.9359

Figure 22. Results using various 2-input models and a 3-input model using RAVDESS were compared. Mel stands for Mel spectrograms, F for frames, and Opt for optical flow frames

Model type \Epochs	10	20	30	40
F & Opt	0.6187	0.8224	0.8412	0.8477
Opt & Mel	0.7558	0.8564	0.8803	0.8667
F & Mel	0.7558	0.9416	0.9667	0.95
3-input	0.8374	0.9212	0.9264	0.9392

Figure 23. Results using various 2-input models and a 3-input model using BAUM-1 were compared. Mel stands for Mel spectrograms, F for frames, and Opt for optical flow frames.

Relevance for educational and inclusive technologies

From the perspective of inclusive Educational Metaverse design, the described method contributes to a broader body of evidence demonstrating the value of combining facial expression and voice modalities for estimating basic emotional states, informing adaptive mechanisms in learning environments such as dynamic adjustment of task difficulty, pacing, or feedback in response to frustration, boredom, or engagement. This work is distinctive in systematically testing CNN-based architectures on both acted emotions (RAVDESS, with controlled professional expressions) and spontaneous emotions elicited by visual stimuli (BAUM-1, where participants react naturally to images/clips), revealing higher accuracy on acted data (~95.5%) versus more variable spontaneous performance (~83-95%), which underscores the challenge of generalizing to ecologically valid scenarios.

For students with ASD, this distinction is particularly salient: acted expressions in training materials may provide clear prototypes for initial learning, while spontaneous reactions, more representative of real peer interactions, better reflect the subtle, context-dependent emotional cues that ASD learners often struggle to decode. The demonstrated robustness across English (RAVDESS) and Turkish (BAUM-1) data further supports cross-cultural applicability, essential for diverse educational metaverses where expressivity varies by linguistic and cultural background. High benchmark performance thus serves as a proof-of-concept for deploying multimodal systems, though real-world validation remains essential given the idiosyncratic emotional profiles of ASD students.

Consequently, these findings provide a technical foundation for the Hovlin-inspired edugame presented in the next section, where Facial Expression Recognition (FER) analysis supports iterative emotion recognition and emulation exercises: acted stimuli scaffold initial matching, while real-time feedback on the student's spontaneous expressions promotes transfer to naturalistic social contexts, aligning with Universal Design for Learning principles without positioning the system as diagnostic.

Theory of Mind, Development, and Intervention

The ability to recognize and attribute mental states—such as beliefs, intentions, desires, and emotions, to oneself and others known as the Theory of Mind (ToM) is a fundamental cognitive skill with significant implications in the personal and social sphere. ToM is a multidimensional construct that includes several key components. Among these are emotional inference (the capacity to understand or predict how a person feels or will feel following an event), deception (the ability to mislead someone into believing something false), and persuasion (the capacity to influence the opinion or intention of others). A further characteristic involves self-reflection and self-awareness, which allow a child to discriminate between their own internal states, such as uncertainty, intention, or certainty regarding a given element.

Despite the potentially uncomfortable feeling of being the object of reflection by someone trying to read our mind, in many small daily moments, we take it for granted that this process occurs smoothly. For example, we expect a friend or relative to maintain the pretense when, in front of our child or grandchild, we ask if they wrote a letter to Santa Claus or if they are behaving well because otherwise, they will not receive the much-desired gift. The development of the ability to infer one's own mind and that of others manifests spontaneously in infancy, following well-defined developmental stages. When these abilities are compromised or fail to develop according to age-expected timelines, a possible impairment in social cognition is hypothesized, often associated with Autism Spectrum Disorder (ASD).

One of the earliest and most studied developmental milestones of ToM is the understanding of First-Order False Belief, an ability that typically emerges between the ages of four and five years. A classic test is to narrate, sometimes using visual aids like a comic strip, the story of Sally and Anne. Sally and Anne are two girls; Sally possesses a marble and a basket, while Anne possesses a box. In a subsequent scene, Sally leaves her marble in the basket and departs. While Sally is absent (not present in the third scene), Anne takes the marble from the basket and moves it into her box. In the final scene, Sally returns near the box and the basket, and the child is then asked where Sally will first look to find her marble. The test, in its various formulations, is used to evaluate the child's ability to deactivate their own knowledge of reality (where the object is actually located) in order to infer where Sally *believes* the object is located. While typically developing children generally pass this test around age four, various studies demonstrate that children with ASD, as well as children with learning difficulties, tend to fail it. The development of ToM proceeds progressively and hierarchically. Parallel to the capacity for perceptual perspective-taking, the inference abilities related to conative (desire) and cognitive (belief) states develop (Wellman, 1990). The ability to infer another person's desire generally develops earlier than the ability to infer their opinion (or belief). During the same developmental period, the understanding of different emotional nuances also unfolds: while simple emotions such as happiness, sadness, anger, and fear can be correctly identified as early as age three, the correct interpretation of more complex or socially mediated emotions requires a more advanced level of maturity.

An edugame for Howlin Program

The proposed program, traceable to the theoretical framework of Pat Howlin (1999), articulates exercises aimed at developing the understanding of emotional states across five different levels of cognitive depth:

Level 1: Static Facial Expression Recognition. This ability consists of correctly associating one of the four primary emotions (happiness, sadness, anger, and fear) with photographs of human faces that exhibit the prototypical facial expressions of these emotional states.

Level 2: Recognition of Emotions in Schematic Drawings. This is the ability to associate the four primary emotions with stylized or simplified face images, analogous to those used in *emoticons*.

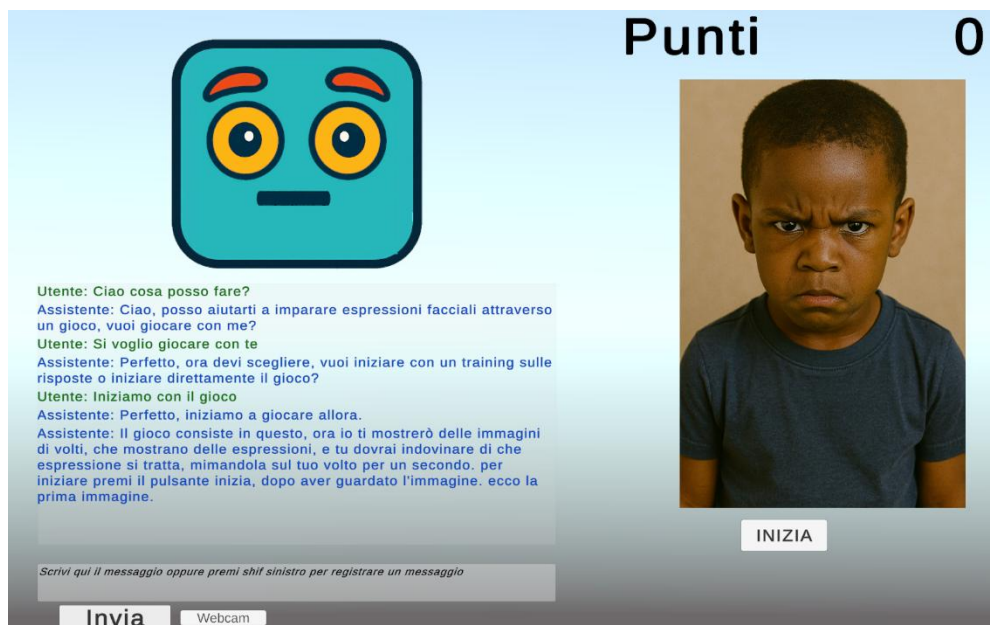
Level 3: Recognition of Emotions Caused by Situational Context. The ability to infer and associate the emotions felt by people present in a specific scene (e.g., someone witnessing an accident) based exclusively on the external and objective context.

Level 4: Recognition of Emotions Caused by Desire. This is the ability to infer an individual's emotional states in situations where a specific desire of theirs is about to be fulfilled or, conversely, frustrated.

Level 5: Recognition of Emotions Caused by Beliefs (Opinions). The ability to infer the emotions felt by an individual following a specific stream of thought, or a false belief (e.g., an emotion based on an opinion that does not correspond to objective reality).

The developed application is configured as an edugame and can be used both to evaluate the child's ability to successfully complete the first level identified by Howlin and to actively help them develop this competence.

The interface is presented as a chatbot, whose knowledge base contains references validated by the pedagogical community, including the Howlin program. Upon initiating the conversation, the chatbot proposes starting the edugame, which constitutes a modernized version of the activities planned for testing or practicing the first level of emotional understanding.



During the session, images of children's faces generated by Artificial Intelligence (AI) are created. The child is then asked to identify which emotion the person in the photo is experiencing and to attempt to imitate it.

If the emotion expressed by the child aligns with the one shown in the image, a point is awarded, and they proceed to a new image. In both outcomes of the test, the chatbot provides feedback interactions that are strongly aligned with the strategies suggested by Howlin. For example:

Positive Outcome: Compliments and encouragement are formulated.

Negative Outcome: The child is first asked which emotion was confused. Subsequently, the chatbot formulates a response suggesting possible reasons for the confusion and, finally, explains the typical context in which that specific emotion usually manifests

Punti 2



Assistente: Perfetto, ora devi scegliere, vuoi iniziare con un training sulle risposte o iniziare direttamente il gioco?
Utente: Iniziamo con il gioco
Assistente: Perfetto, iniziamo a giocare allora.
Assistente: Il gioco consiste in questo, ora io ti mostrerò delle immagini di volti, che mostrano delle espressioni, e tu dovrai indovinare di che espressione si tratta, mimandola sul tuo volto per un secondo. per iniziare premi il pulsante inizia, dopo aver guardato l'immagine. ecco la prima immagine.
Assistente: Bravo, hai mimato bene l'espressione "arrabbiato".
Assistente: Ora passiamo alla prossima immagine.
Assistente: Hai mimato "triste". Qual è, secondo te, l'emozione corretta dell'immagine?
Utente: Paura
Assistente: Esatto, l'immagine mostra paura, bravo!



Emozione: FELICITÀ
Probabilità: 63,4%

INIZIA

Webcam

Invia

Potentialities, limits and AI issues

The developed application represents a viable alternative to previous methodologies used for the implementation of the Howlin program (or related models). A significant advantage lies in resolving the issues related to the availability and copyright of photographic material, as the facial images are generated *on-demand* by Artificial Intelligence (AI). The ability to artificially generate visual material offers a potential solution to the complex problem of cultural biases in emotion perception. It is well-established that the interpretation of facial expressions is modulated by emotional display rules, which vary significantly across different cultures (e.g., Western vs. Eastern cultures). In theory, AI generation allows for the training or fine-tuning of models to produce faces with controlled ethnic and phenotypic variability, and to manipulate the saturation or intensity of the expression to align with the diverse cultural manifestations of primary emotions, thereby overcoming the limitations of standardized photographic datasets which often show reduced diversity.

Despite the broad potential of the tool, stemming from the versatility of the Artificial Intelligence methods upon which it is based, the application is currently limited to verification and practice activities related to Level 1. The implementation of Level 2 is technically the least onerous; for coherence with the technological framework, reliance on manual drawing programs is advised against in favor of using AI for procedural generation or vector stylization of emoticons, thereby ensuring the consistency of the minimalist design and the efficiency of automatic creation. Conversely, the implementation of the higher stages—namely Levels 3, 4, and 5 (which require the inference of emotions based on situational context, desire, and belief)—presents a significantly greater difficulty, as these levels necessitate the generation not merely of a single face, but of entire complex narrative scenes composed of multiple characters, objects, and an environment that suggests a conflict or a specific mental state required for cognitive inference.

A future strategy could delegate the creation of this complex material to advanced AI image generators (such as Stable Diffusion or DALL-E). However, this would impose an additional cognitive burden on the end-user (the teacher), who would require considerable mastery of Prompt Engineering to achieve archetypal and didactically valid results (i.e., analogous in clarity and validity to the materials proposed by Howlin). A preferable solution, to maintain interface simplicity, would be to develop simplified User Interfaces (UI) that translate the teacher's didactic requests (e.g., "Scene: Boy sad because the balloon popped") into complex *prompts* managed by the system in the *backend*.

A crucial and non-secondary aspect in the design of the application concerns compliance with European regulations and guidelines on the use of Artificial Intelligence (AI) based technologies in educational and didactic contexts. The main regulatory reference is the European Union's Artificial Intelligence Act (AI Act). The Regulation clearly defines the general purposes of such technologies, as indicated in Article 1: "The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI

systems in the Union and supporting innovation." Particular attention is paid to applications considered high-risk or prohibited. Article 5(1)(f) establishes that the following uses are considered prohibited: "...the placing on the market, the putting into service for this specific purpose, or the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons." The developed didactic application utilizes two distinct AI methods: a Large Language Model (LLM) for managing the chatbot and a Facial Expression Recognition (FER) system. Several technical and procedural measures have been adopted to ensure the full compliance of the FER system with the AI Act, demonstrating that the use is didactic and non-inferential: the user has total control over the function, with the webcam able to be switched off at any time by simply selecting the appropriate toggle to "off," allowing the chatbot functionalities and didactic flow to proceed unaltered; the experience is designed not to retain any data from the camera's video stream, as individual frames are analyzed exclusively in-memory (i.e., without permanent storage and preferably on-device), and real-time feedback is provided solely as a didactic aid, a kind of "smart mirror", to help the learner better simulate the required emotion. The core normative criticality revolves around inference (deducing the internal emotional state); the application actively asks the user to emulate a specific facial expression (e.g., "make a sad face") in a fully conscious manner. Furthermore, it is explicitly acknowledged that emulating a certain facial expression does not imply that the person is actually experiencing the corresponding emotion, thereby clearly distinguishing the didactic simulation from the prohibited emotional monitoring. Regarding the LLM Chatbot, the greatest ethical risk in an educational context lies in the possible profiling of the user's cognitive or emotional states based on language analysis. To mitigate this risk, it is essential that the textual interaction also takes place under a regime of anonymization and that conversation logs are not used for targeted profiling, thereby preserving the fundamental rights of the learner in line with the provisions of the Regulation.

Bybliography

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, 9(3), 185–211.

Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. Bantam Books.

Rivoltella, P. C. (2014). *Fare didattica con gli EAS. Episodi di apprendimento situato*. La Scuola.

Bonaiuti, G., Calvani, A., & Picci, P. (2014). *Didattica digitale. Formazione e insegnamento con le tecnologie*. Carocci.

Bilotti, U., Bisogni, C., De Marsico, M., & Tramonte, S. (2024). Multimodal emotion recognition via convolutional neural networks: Comparison of different strategies on two multimodal datasets. *Engineering Applications of Artificial Intelligence*, 130, 107708. <https://doi.org/10.1016/j.engappai.2023.107708>

Savia, G. (2016). *Universal Design for Learning e inclusione scolastica* (verifica titolo e editore esatti nella tua bibliografia).

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3–4), 169–200.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, 3(4), 364–370.

Clavel, C., & Callejas, Z. (2016). Sentiment analysis: From opinion mining to human–agent interaction. *IEEE Transactions on Affective Computing*, 7(1), 74–93.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715–734.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>

Pan, B., Hirota, K., Jia, Z., & Dai, Y. (2023). A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods. *Neurocomputing*, 561, 126866.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>

Wu, T., Purwanto, E., Huang, Y., & Yang, S. (2025). Phy-FusionNet: A Memory-Augmented Transformer for Multimodal Emotion Recognition With Periodicity and Contextual Attention. *IEEE Transactions on Affective Computing*.

Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A. (2021). Application of deep learning on student engagement in e-learning environments. *Computers & Electrical Engineering*, 93, 107277.

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE*, 13(5), e0196391. <https://doi.org/10.1371/journal.pone.0196391>

S. Zhalehpour, O. Onder, Z. Akhtar and C. E. Erdem, "BAUM-1: A Spontaneous Audio-Visual Face Database of Affective and Mental States," in *IEEE Transactions on Affective Computing*, vol. 8, no. 3, pp. 300-313, 1 July-Sept. 2017, doi: 10.1109/TAFFC.2016.2553038.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, 2, 674-679.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4), 357-366.

Chowdhury, J. H., Ramanna, S., & Kotecha, K. (2025). Speech emotion recognition with light weight deep neural ensemble model using hand crafted features. *Scientific Reports*, 15(1), 11824.

Alfarano, A., Maiano, L., Papa, L., & Amerini, I. (2024). Estimating optical flow: A comprehensive review of the state of the art. *Computer vision and image understanding*, 249, 104160.

Wellman, H. M., Carey, S., Gleitman, L., Newport, E. L., & Spelke, E. S. (1990). *The child's theory of mind*. The MIT Press.

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teoria della mente e autismo-Insegnare a comprendere gli stati psichici dell'altro*. Edizioni Erickson.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37-46.

Flavell, J. H., Shipstead, S. G., & Croft, K. (1978). Young children's knowledge about visual perception: Hiding objects from others. *Child Development*, 1208-1211.

Perner, J., & Davies, G. (1991). Understanding the mind as an active information processor: Do young children have a “copy theory of mind”? *Cognition*, 39(1), 51-69.

Baron-Cohen, S. E., Tager-Flusberg, H. E., & Cohen, D. J. (1994). Understanding other minds: Perspectives from autism. In *Most of the chapters in this book were presented in draft form at a workshop in Seattle, Apr 1991.* Oxford University Press.

Aiello, P., & Zappalà, E. (2023). Future support teachers' opinions about their role on inclusion of students with ASD: an exploratory study to determine training impact. *Education Sciences & Society*: 1, 2023, 342-357.

Aiello, P., Di Gennaro, D. C., Palumbo, C., Zollo, I., & Sibilio, M. (2014). Inclusion and universal design for learning in Italian schools. *International Journal of Digital Literacy and Digital Competence (IJDLC)*, 5(2), 59-68.

Conclusion

This work, which analyzed fundamental principles, defined the emerging context of the Educational Metaverse, and applied inclusive strategies across three critical areas, offers a holistic perspective on how pedagogical research can not only respond to but also actively guide the evolution of immersive technologies and Artificial Intelligence toward genuinely inclusive didactic models. The Educational Metaverse, far from being merely a futuristic fantasy, has been conceptualized as the evolution of *onlife* existence, an ecosystem that merges immersion in Virtual Reality with the potential of Blockchain and the crucial role of Artificial Intelligence (AI). Historical analysis highlighted that its precursors and theoretical pillars (such as immersion, presence, and communicative dynamics) are already rooted in the history of educational technologies and learning theory. However, it was also outlined that the ethical and practical challenges, security, accessibility, and the digital divide, hinder the Metaverse from achieving its democratic and universal potential. The thesis provided concrete answers to these challenges through the empirical application of three distinct lines of research. Firstly, the intervention on Accessibility and Human-Machine Interaction addressed the core deficit of dyslexia, demonstrating that optimizing typographic design for wearable technologies, through the experimentation of a high-readability font, represents a low-cost yet high-impact intervention for improving Accessibility (HCI) in textual content. This establishes a crucial precedent for ensuring that future technologies do not inherit the barriers present in traditional media. Secondly, the focus on Virtual Reality and Neurodiversity led to the design of specific Learning Units for students with ADHD. The analysis demonstrated that VR, when intentionally designed through an inclusive lens and informed by models such as Hein's constructivism, becomes a powerful tool for manipulative didactics and behavior management in immersive environments, thanks to strategies like task decomposition and adaptive scaffolding to minimize extraneous

cognitive load. Finally, the third line of research explored Emotional Learning in students with ASD, integrating an *edugame* based on the Howlin program with facial recognition technologies (FER). The most significant conclusion in this area is that the effectiveness of such new-generation tools intrinsically depends on their capacity to be aligned with solid pedagogical models and, crucially, to rigorously adhere to current regulatory frameworks, ensuring that the use of AI is strictly didactic and does not violate data protection principles. The historical period we are currently living in is not a period of progress; rather, it is a period of uncertainty. While a few centuries ago we knew exactly that the engineer was behind the machine, now we do not know the machine, and what is even more serious, we do not know who is behind the machine. What this work sought to achieve, as anticipated by the title, is to know who is Beyond the Machine. In our case, beyond the machine are the interpreters of the learning process, the components of the educational-didactic framework. However, this is not a recent endeavor. The pedagogical community has always faced the responsibility of knowing who is beyond the machine. The word itself, which Di Tore asserted is a technology, or for a more current example, beyond the screen of an online Meet, there is someone who manifests a part of their identity.

Reflections on our extended existence, which began long before the pandemic period, must not accept an endpoint but only new starting points. With the adoption of this paradigm, any technological equipment, any content delivery modality, and any didactic practice will feel physiologically restrictive; in other words, to foster inclusion, it is inevitably necessary to become uncomfortable. The common thread connecting the works presented in this thesis is precisely a discomfort that penetrates all components of the pedagogical framework. It is uncomfortable for a student to use a new font, it is uncomfortable for a teacher to propose a virtual experience within a learning unit, and it is uncomfortable for the pedagogical community to talk about Artificial Intelligence. However, even when we are uncomfortable in the best armchair available in the room, we do not simply start moving, but rather experimenting with other configurations of "sitting down." In the same way, when a certain real or virtual environment feels uncomfortable, we begin to elaborate strategies to modify it, in other words, to adapt to the environment or to adapt the environment. This discomfort, which may almost always appear accessory at first glance, is therefore actually necessary,

and research and commitment to fostering the process of inclusion in educational contexts cannot transcend it.

Appendix

Below is the questionnaire used for the font experimentation.

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη: *



Η Δημιουργία του Αδάμ

Είναι νωπογραφία του 16ου αιώνα (περ. 1508-1512) που δημιουργήθηκε από τον Ιταλό καλλιτέχνη Μιχαήλ Άγγελο και αποτελεί μέρος της οροφής της Καπέλα Σιστίνα.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα



Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη: *



Η Δημιουργία του Αδάμ

Είναι νωπογραφία του 16ου αιώνα (περ. 1508-1512) που δημιουργήθηκε από τον Ιταλό καλλιτέχνη Μιχαήλ Άγγελο και αποτελεί μέρος της οροφής της Καπέλα Σιστίνα.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη: *



Η Δημιουργία του Αδάμ

Είναι νωπογραφία του 16ου αιώνα (περ. 1508-1512) που δημιουργήθηκε από τον Ιταλό καλλιτέχνη Μιχαήλ Άγγελο και αποτελεί μέρος της οροφής της Καπέλα Σιστίνα.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα



Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη: *



Η Δημιουργία του Αδάμ

Είναι νωπογραφία του 16ου αιώνα (περ. 1508-1512) που δημιουργήθηκε από τον Ιταλό καλλιτέχνη Μιχαήλ Άγγελο και αποτελεί μέρος της οροφής της Καπέλα Σιστίνα.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη:

*



Η Αφροδίτη των Σόλων

Η Αφροδίτη της Κύπρου είναι ένα μαρμάρινο άγαλμα που χρονολογείται γύρω στον 1ο αιώνα π.Χ., και αναπαριστά την θεά Αφροδίτη γυμνή. Το άγαλμα αυτό γνωστό και ως η Αφροδίτη των Σόλων, βρέθηκε στην ομώνυμη αρχαιολογική περιοχή των Σόλων της Κύπρου.



- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη:

*



Η Αφροδίτη των Σόλων

Η Αφροδίτη της Κύπρου είναι ένα μαρμάρινο άγαλμα που χρονολογείται γύρω στον 1ο αιώνα π.Χ., και αναπαριστά την θεά Αφροδίτη γυμνή. Το άγαλμα αυτό γνωστό και ως η Αφροδίτη των Σόλων, βρέθηκε στην ομώνυμη αρχαιολογική περιοχή των Σόλων της Κύπρου.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη:

*



Η Αφροδίτη των Σόλων

Η Αφροδίτη της Κύπρου είναι ένα μαρμάρινο άγαλμα που χρονολογείται γύρω στον 1ο αιώνα π.Χ., και αναπαριστά την θεά Αφροδίτη γυμνή. Το άγαλμα αυτό γνωστό και ως η Αφροδίτη των Σόλων, βρέθηκε στην ομώνυμη αρχαιολογική περιοχή των Σόλων της Κύπρου.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη: *

Η Αφροδίτη των Σόλων



Η Αφροδίτη της Κύπρου είναι ένα μαρμάρινο άγαλμα που χρονολογείται γύρω στον 1ο αιώνα π.Χ., και αναπαριστά την θεά Αφροδίτη γυμνή. Το άγαλμα αυτό γνωστό και ως η Αφροδίτη των Σόλων, βρέθηκε στην ομώνυμη αρχαιολογική περιοχή των Σόλων της Κύπρου.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα



Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη:

*



Τάλαντο Χαλκού

Το τάλαντο σε σχήμα δοράς βοδιού, η μορφή με την οποία εξήγαγαν τον κυπριακό χαλκό από το 1600-1050 π.Χ., είναι ένα από τα πιο χαρακτηριστικά αντικείμενα της αρχαίας Κύπρου.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη: *



Τάλαντο Χαλκού

Το τάλαντο σε σχήμα δοράς βοδιού, η μορφή με την οποία εξήγαγαν τον κυπριακό χαλκό από το 1600-1050 π.Χ., είναι ένα από τα πιο χαρακτηριστικά αντικείμενα της αρχαίας Κύπρου.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη:

*



Τάλαντο Χαλκού

Το τάλαντο σε σχήμα δοράς βοδιού, η μορφή με την οποία εξήγαγαν τον κυπριακό χαλκό από το 1600-1050 π.Χ., είναι ένα από τα πιο χαρακτηριστικά αντικείμενα της αρχαίας Κύπρου.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η ακόλουθη λεζάντα είναι ευανάγνωστη:

*



Τάχαντο Χαλκού

Το τάχαντο σε σχήμα δοράς βοδιού, η μορφή με την οποία εξήγαγαν τον κυπριακό χαλκό από το 1600-1050 π. Χ., είναι ένα από τα πιο χαρακτηριστικά αντικείμενα της αρχαίας Κύπρου.

- ☐ 1=Διαφωνώ απόλυτα
- ☐ 2=Διαφωνώ
- ☐ 3=Ούτε Συμφωνώ/Ούτε Διαφωνώ
- ☐ 4=Συμφωνώ
- ☐ 5=Συμφωνώ απόλυτα

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι το Πυθαγόρειο Θεώρημα είναι ευανάγνωστο με κάθε μία από τις παρακάτω γραμματοσειρές:

1. $\alpha^2 = \beta^2 + \gamma^2$

2. $\alpha^2 = \beta^2 + \gamma^2$

3. $\alpha^2 = \beta^2 + \gamma^2$

4. $\alpha^2 = \beta^2 + \gamma^2$

	1=Διαφωνώ Απόλυτα	2=Διαφωνώ	3=Ούτε Συμφωνώ/ Ούτε Διαφωνώ	4=Συμφωνώ	5=Συμφωνώ Απόλυτα
Γραμματοσειρά 1	☐	☐	☐	☐	☐
Γραμματοσειρά 2	☐	☐	☐	☐	☐
Γραμματοσειρά 3	☐	☐	☐	☐	☐
Γραμματοσειρά 4	☐	☐	☐	☐	☐

Σε κλίμακα από 1 μέχρι 5, όπου το 1= Διαφωνώ απόλυτα και το 5= Συμφωνώ απόλυτα, δηλώστε κατά πόσο συμφωνείτε ή διαφωνείτε ότι η μαθηματική φόρμουλα είναι ευανάγνωστη με κάθε μία από τις παρακάτω γραμματοσειρές: *

1. $\psi = \omega\chi + \varphi$

2. $\psi = \omega\chi + \phi$

3. $\psi = \omega\chi + \varphi$

4. $\psi = \omega\chi + \varphi$

	1=Διαφωνώ Απόλυτα	2=Διαφωνώ	3=Ούτε Συμφωνώ/ Ούτε Διαφωνώ	4=Συμφωνώ	5=Συμφωνώ Απόλυτα
Γραμματοσειρά 1	☐	☐	☐	☐	☐
Γραμματοσειρά 2	☐	☐	☐	☐	☐
Γραμματοσειρά 3	☐	☐	☐	☐	☐
Γραμματοσειρά 4	☐	☐	☐	☐	☐

Φύλο:

- ☐ Γυναίκα
- ☐ Άντρας
- ☐ Άλλο

Ηλικία:

- ☐ 18-22 ετών
- ☐ 23-25 ετών
- ☐ 26-29 ετών
- ☐ Άνω των 30 ετών

Έχεις επίσημη διάγνωση αναγνωστικών δυσκολιών (π.χ. Δυσλεξία)

- ☐ Ναι
- ☐ Όχι
- ☐ Δεν γνωρίζω

