



**Politecnico
di Torino**

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (38th cycle)

Digital Transformation Technologies for Virtual Experiences

Jacopo Fiorenza

* * * * *

Supervisors

Prof. Fabrizio Lamberti

Prof. Anna Osello

Doctoral Examination Committee:

Prof. Bill Kapralos, Ontario Tech University

Prof. Giacinto Barresi, University of the West of England Bristol

Prof. Renato Ferrero, Politecnico di Torino

Politecnico di Torino

This thesis is licensed under a Creative Commons License, Attribution - Noncommercial-NoDerivative Works 4.0 International: see www.creativecommons.org. The text may be reproduced for non-commercial purposes, provided that credit is given to the original author.

This doctoral thesis is part of the project PNRR-NGEU which has received funding from the MUR – DM 352/2022



I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

.....
Jacopo Fiorenza
Turin, May 11, 2026

Summary

Over the past two decades, digital transformation has gained increasing attention, driven by the convergence of multiple Digital Transformation Technologies (DTTs) across several sectors. Among them, Virtual Reality (VR) has emerged as a key technology for creating immersive and interactive Virtual eXperiences (VXs), often enhanced through integration with other DTTs, such as laser scanning, descriptive CAD tools, and large language models. Despite their potential, the creation of VXs still faces significant challenges. Current VX design practices often lack principles for specific target users in various domains, while technical complexity and interoperability issues frequently hinder the creation of virtual environments. Although the integration of other DTTs could possibly address these challenges, several of them remain underexplored in the context of VXs.

This doctoral thesis investigates current challenges in relation to VXs and their combination with other DTTs, aiming at analyzing their strengths, limitations, and applicability across various domains. The presented investigations, conducted over the years of the Ph.D. program, address a wide range of topics, including virtual design for user accessibility, automated reconstruction of virtual environments for VX creation, and the use of generative AI-based technologies to develop conversational agents. In this context, the research explores the convergence of VR with several DTTs, such as 3D photogrammetry, Building Information Modeling (BIM), and large language models.

The findings of the thesis contribute to a deeper understanding of key aspects of VXs, including accessibility, user experience, and usability, while also underlining the benefits and limitations introduced by the integration of complementary DTTs. Overall, this thesis highlights both the opportunities and challenges in the VR and other DTTs domains, outlining promising directions for future research in the design and development of advanced VXs.

Acknowledgements

I would like to thank all the people I have met and who have supported me during this long journey.

First and foremost, I would like to thank my supervisor, Prof. Fabrizio Lamberti, who had the patience to review, critique, and improve my work for three years. I am especially grateful to him for providing me with the motivation I needed to continue my research and for always being available and direct in his assessments. His approach to work and his experience will always stay with me.

I would also like to thank my colleagues in the laboratory, the members of g10, from the bottom of my heart. I will always remember these years with them, filled with hard work and sleepless nights, but also laughter and satisfaction. I would like to thank Prof. Anna Osello for giving me the opportunity to work on an amazing project, and all the members of Palazzo Carignano, with whom I spent months preparing research activities and who were always available to help me on every occasion. I would like to also thank Massimo Migliorini for supporting me during my research period at Links Foundation, as well as all the members of the XR laboratory. I hope I have left something inside each of them, just as they have done with me.

I would also like to take this opportunity to thank Prof. Ruck Thawonmas, who guided and supported me during my research in Japan. I am especially grateful to him for giving me the opportunity to spend six months in Osaka, a period of my life that profoundly changed me and will remain with me forever. In this regard, I would like to thank the members of the laboratory at Ritsumeikan University and all the other people I met in Japan. It would not have been such a unique experience without them.

I would like to thank my parents, who supported me throughout my journey, and my friends, who put up with my talk and complaints about my scientific research for three years. I would like to thank my brother, with whom I can always talk and who has always supported me in my decisions, and my nephews, who are growing up faster and faster every year. A deep thank you also goes to my uncle, who has always cheered me up with his cooking, and my aunt, wherever she may be now.

Finally, I thank myself for persevering through these three years. I could list a million reasons why I am different now, but I don't think even all the pages of this

thesis would be enough to explain it. So, you'll have to settle for this...

*“We are what we are because we have
been what we have been.”*

- Sigmund Freud

Contents

List of Tables	X
List of Figures	XI
1 Introduction	1
1.1 Context and Motivations	1
1.2 Thesis Structure	5
2 VX Design and Immersive VR	9
2.1 Accessible Virtual Museum Experiences for ASD Individuals	10
2.1.1 Overview	10
2.1.2 Background	13
2.1.3 Materials and Methods	20
2.1.4 Use Case	29
2.1.5 Results and Discussion	45
2.2 Closing Remarks	56
3 Virtual Environment Reconstruction	59
3.1 HBIM for Virtual Heritage Experiences	60
3.1.1 Overview	60
3.1.2 Background	62
3.1.3 Materials and Methods	72
3.1.4 Use Case	85
3.1.5 Results and Discussion	89
3.2 Closing Remarks	94
4 Generative AI for Conversational Agents	97
4.1 Unified UX Design Across Multiple Virtual Museum Experiences	98
4.1.1 Overview	98
4.1.2 Background	99
4.1.3 Materials and Methods	101
4.1.4 Results and Discussion	106
4.2 Vision-Language Models for Environment Aware Agents	107

4.2.1	Overview	107
4.2.2	Background	109
4.2.3	Materials and Methods	111
4.2.4	Use Case	115
4.2.5	Results and Discussion	120
4.3	Closing Remarks	125
5	Conclusions	129
5.1	Summary of Contributions	129
5.2	Limitations and Future Works	133
	Glossary	135
	Bibliography	140

List of Tables

2.1	Structural representation a POI procedure; beyond the interaction routines, interaction steps and tasks, it is possible to observe in the last column the default values, in seconds, for the timeouts.	38
3.1	Characteristics of existing VHE platforms analyzed in Section 3.1.2, in comparison with the proposed solution. ✓ = Available, - = Not available, ✓* = Planned as future work, G = Generated, P = Pre-defined	70
3.2	IFC entities and parameters defined within the HBIM model.	75
4.1	Detailed list of current agents employed in the state-of-the-art. The not yet defined acronyms are reported as follows: EK, Environmental Knowledge; GSE, Game System Extracted; VLMG, VLM-Generated; MM, Memory Management; AP, Action Planning; PSLTY, Personality; BEX, Body Expressions; FEX, Facial Expressions; UG, User Guidance; TM, Task Management;	112
4.2	Example of template system prompt for the VLM module	116

List of Figures

1.1	Some of the current most representative DTTs according to [130, 57]. VR, which is part of the XR technologies, is often combined with other DTTs across several applications and domains, such as BIM for architectural projects, robotics for manufacturing, and IoT for urban planning.	3
2.1	Key points of the analyzed works, including the inclusive immersion principles addressed, the technologies used, the methodologies adopted, and the ID types considered.	21
2.2	Example of the proposed approach combining VE modularity, through progressive enrichment, and application modularity in a VR application. At the beginning of the Module Procedure (MP), both the cultural element (a) and the cultural description (b) are respectively invisible and inaccessible. After completing a predefined number of tasks (c), the cultural element becomes visible (d), and the cultural description (e) can be accessed and explored by the user.	23
2.3	Example of a module procedure (MP) integrating the predictability principle. The procedure includes two interaction routines (IRs) within a single Interaction Pattern (IP), composed of multiple interaction steps (IS), a cultural element (a), and a cultural description (b). Differently from the configuration adopting only modularity (Figure 2.2), the tasks are embedded within the interactions steps. Once the final task of the first routine is completed (c), the cultural element (d) and its description (e) become accessible.	25
2.4	Conceptual representation of an interaction routine (IR), composed of multiple interaction steps (IS1-2-3), each associated with a specific task and a timeout constraint (a, b, and c).	26

2.5	Simplified representation of a head-gaze-based interaction routine using a visual facilitator (VF). The routine includes three interaction steps (IS1-2-3), each associated with a timeout, and three spatial positions within the VE (x_i, y_i, z_i). In accordance with the modularity principle, the cultural element is initially hidden (a). After completion of the final interaction step, the cultural element becomes visible (c), while the visual facilitator disappears (d).	27
2.6	Overview of the three proposed inclusive immersion principles and their integration within the design of a VR-based VME.	28
2.7	Scheme representing the key elements of the proposed methodology, namely the inclusive immersion principles, the material preparation, and the co-design refinement.	30
2.8	Picture of the “Hall of Seasons” of Palazzo Carignano, Turin, Italy, representing the environment to reconstruct for the use case (©MiC – DRM Piemonte).	31
2.9	Image of the frescoed vault of the “Hall of Seasons” (©MiC – DRM Piemonte).	33
2.10	Golden decorations in the “Hall of Seasons” depicting the allegories in correspondence to the four seasons (©MiC – DRM Piemonte). . .	34
2.11	Tutorial scene. As observable on the floor, the unlocked (blue) and locked (orange) TPs enable the user’s movement within the VE. . .	34
2.12	The image depicts the digital VE, i.e., the “Hall of Seasons”, as visualized in the VME scene of the VR application. On the floor, the pathway and the TPs defining the POIs of the VME are visible. . .	35
2.13	The images represent the four interaction steps (1/2/3/4) in the first interaction routine of a POI. Specifically, the example corresponds to the cultural element linked to the winter season decoration. . . .	39
2.14	Demographic representation of the participants (part 1): age group (a), level of education (b), gender (c), frequency of use of VR technology (d), and occupation (e).	44
2.15	Percentage breakdown of participant demographics (part 2): specific syndrome (a), sensory processing difficulties (b), reactions to stress or overstimulation (c), and ASD trait level (AQ-Short [97]) (d). . .	45
2.16	Overall comparison for the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars.	48
2.17	Overall comparison regarding for the sections of the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars. . .	49
2.18	Overall analysis in terms of age for the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars.	50

2.19	Overall analysis in terms of ASD trait level (AQ-Short [97] score) for the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars.	52
3.1	Schematic representation of the three automatic stages executed by the VR platform after the selection of the environment to be generated.	79
3.2	Schematic representation of the handling of POIs and LAOs. In this example, the i -th POI manages all the LAOs associated with it. As illustrated, both POIs and LAOs are capable of issuing TTS requests through the TTS Manager.	83
3.3	Examples of UI elements integrated within the application (main menu UI shown as reference): a) selection of the environment to be reconstructed and visualization of related information; b) option to make the 3D miniature of the environment transparent in order to display POIs or to initiate the virtual visit.	84
3.4	Example of a UI panel used to present a LAO within the VHE. . .	85
3.5	Picture of the real-world environment, i.e., the “Hall of Seasons” of Palazzo Carignano, in comparison to its HBIM model.	86
3.6	Picture of the three LODs for one of the meshes representing the “Fall” season. In this case, the number of vertices for the LODs is ~ 200 (LOD0), $\sim 80k$ (LOD1), and $\sim 15k$ (LOD2).	88
3.7	Frescoed vault texture UV mapping (a) and visual representation (b) in Blender.	89
3.8	Miniature representation displayed at the end of the VE reconstruction displayed a) using a flat gray material, and b) in wireframe to enable the visualization of POIs.	90
3.9	Representation of the state of the VE during the virtual visit: a) overview of the hall and of its features, both in terms of graphic and rendering (e.g. mirror reflections), as well as examples of automatic handling of textures and UVs for b) the frescoed vault and c) other meshes (e.g., wall paintings).	91
3.10	a) The first LAO associated with the initial POI of the virtual tour, emphasized by the activation of a spotlight. In this instance, the LAO corresponds to the decorative element symbolizing the Summer season; b) UI panel shown for the first LAO linked to the initial POI of the virtual tour, presenting cultural information directly extracted from the HBIM model.	92
4.1	Visual representation of the two VEs: the meta-museum experience (a) and the virtual workshop (b).	103
4.2	Schematic representation of the ECA architecture;	104
4.3	Preliminary data for the users’ perception of the ECA	107

4.4	Devised VLM-based environment perception module integrated in a traditional architecture of an LLM-based ECA. Over both the VLM module and the Conversational Module, their respective system prompts can be observed.	114
4.5	One of the rooms part of the VE within the video game. The four images represent the frames extracted before passing them to the VLM module to generate the SG. An NPC is standing within the room, surrounded by different objects.	117
4.6	An example of the devised interaction modality with one of the NPCs within the VE. In the first step (a,b) the player can freely have a conversation by using a chat interface, while within the second step (c,d) it can select questions from a UI showing dialogue options. In the third step (e,f,g), the player can explore the environment in search of clues, freely interacting again with the NPCs.	119
4.7	The graphs depict the results obtained from the experimental phase. Specifically, an overall comparison (a) and a more detailed analysis on the individual dimensions: PK (b), SPK (c), PI (d), FCTY (e), and WFI.	121
4.8	The graphs show the results obtained from the experimental phase. Specifically, an overall comparison (a) and a more detailed analysis on the individual dimensions of the PMQ: CCD (b), HLC (c), and CF(d).	123

Chapter 1

Introduction

1.1 Context and Motivations

Since the definition of Information Technology (IT) in 1958 [125], the concept of Digital Transformation (DT) has substantially evolved over time. Along with the digitization of traditional processes started in the 1990's, this concept has been initially associated with the conversion of analog data into digital information [118]. However, the advancement in technology reached in the last twenty years has enabled a broader interpretation of what actually DT represents. Essentially, if digitization has brought a transition towards digital information, DT has taken a step further by representing a process in which IT is progressively integrated into existing organizations (e.g., companies, institutions) [183]. This has eventually implied an important shift in how these organizations operate and engage with users and customers [164].

In the last decade, the DT process has become even more relevant, representing a fundamental driver to improve entities across several fields. This improvement has been enabled over time by the exponential growth of information and communication technologies [199], as well as by their employment and combination [214]. Given their importance, these technologies – in this context referred to as Digital Transformation Technologies (DTTs) – and their applications have been extensively explored by both industry and academia [130]. The evolution of modern DTTs, such as Internet of Things (IoT), Artificial Intelligence (AI), and cloud computing, had a substantial impact on several sectors, such as healthcare [147], education [66] and tourism [89]. Additionally, a concurrent usage of these technologies to build innovative tools and systems has been extensively investigated in modern literature [194, 47]. On the one hand, this combination of DTTs, commonly referred to as “technological convergence” [23], increased the value of existing technologies and fostered innovation [141]. On the other hand, it introduced new challenges to allow this convergence to happen efficiently. This consideration becomes substantially relevant when the involved DTTs are not commonly employed in the same

scenarios, or they present peculiar hardware requirements.

Currently, the DT process is still ongoing and involves a wide set of rapidly evolving and converging DTTs (Figure 1.1). Some of the most promising in terms of innovation fall under the umbrella term of eXtended Reality (XR), namely Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). Particularly, VR has become in recent years a trendy argument in the IT domain. From a general perspective, VR is a computer simulation aiming at immersing users in a Virtual Environment (VE) [221]. Fostered by the rising production of affordable Head-Mounted Displays (HMDs), such as Meta Quest¹ and HTC Vive², this concept has become popular both in research studies and industrial organizations. From a more detailed perspective, VR comprises two key elements: *presence* and *immersion* [220]. The former can be associated with the feeling of users being located in a certain environment [188]. The latter can be defined as the degree to which a certain VR system can produce a comprehensive, engaging, and realistic sense of reality [196, 117]. Thus, immersion not only depends on the employed technologies, i.e., immersive (e.g., HMD) or non-immersive (e.g., LCD screen), but also by the experience within the VE as perceived by users [220].

With the previously highlighted considerations, VR is among the most relevant DTTs to create meaningful and effective Virtual eXperiences (VXs). At a global level, the scientific community has widely explored the topic for numerous applications. Eventually, the technological convergence trend consisting in combining multiple DTTs has become a common practice also for the creation of VXs.

From the perspective of Human-Computer Interaction (HCI) and User eXperience (UX), VXs typically involve one or multiple users interacting with each other and/or with their surrounding environment. The presence of users into VXs can be considered as a core aspect, as it assumes the need of key design principles specifically tailored to the considered individuals and application field. The nature of these principles ultimately defines most of the core features of a VX, including the VE, the interaction and locomotion modalities (e.g., hand-gestures and teleportation), as well as the DTTs employed. This kind of user-centered design has been widely explored by the research community, with the common objective of enhancing UX and HCI in various contexts. For instance, applications employing immersive VR have been regularly developed for educational and training purposes [10, 144]. Moreover, the benefits provided by VR technologies of building safe and secure VEs, as well as simulating real-world tasks safely [155], have enabled the creation of highly user-specific VR applications, such as for rehabilitation or social inclusion of individuals with cognitive or physical impairments (i.e., users with Autism Spectrum Disorder (ASD)) [174, 187].

¹Meta Quest: <https://www.meta.com/en/quest/>

²HTC Vive: <https://www.vive.com/us/>

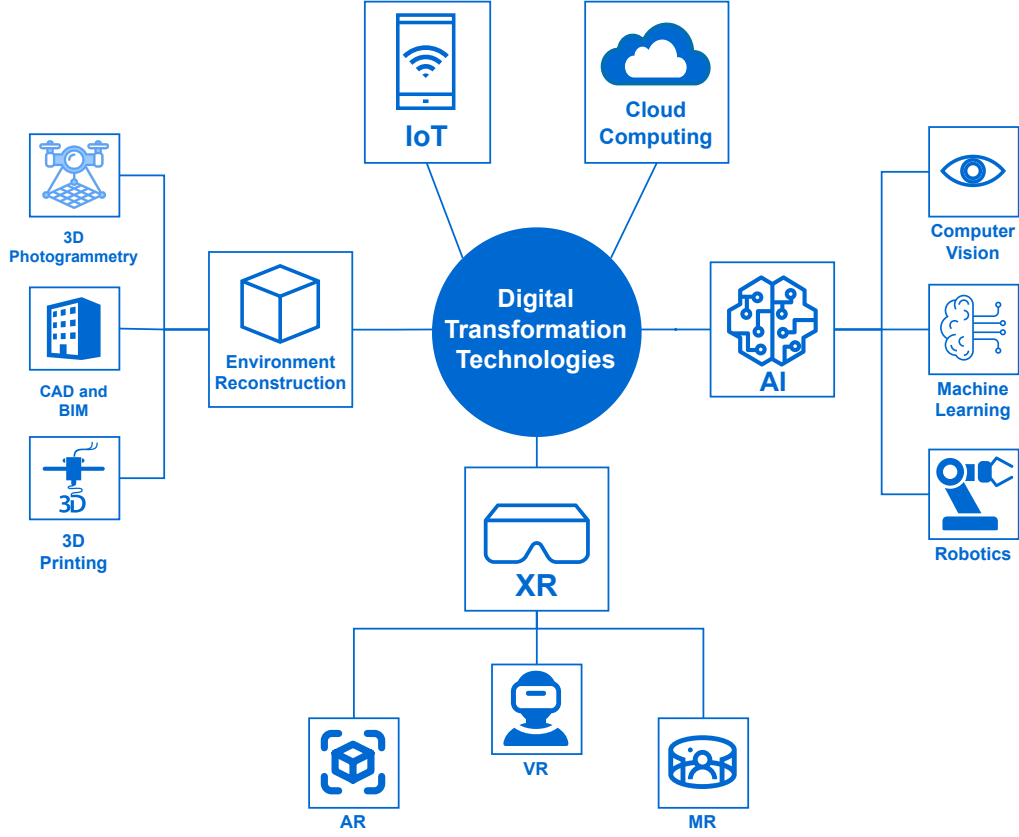


Figure 1.1: Some of the current most representative DTTs according to [130, 57]. VR, which is part of the XR technologies, is often combined with other DTTs across several applications and domains, such as BIM for architectural projects, robotics for manufacturing, and IoT for urban planning.

Despite the advantages, for some categories of users and application domains, the employment of effective design principles for VXs is still underexplored. On the one hand, this lack of adoption is due to the fact that some criteria to define these principles have not been widely accepted yet. On the other hand, their practical implementation is often hindered by the long technical process of developing the VE, which is another core aspect to consider when creating VXs.

The environment of a VX may vary across applications and be characterized by several static and dynamic elements, such as user-interfaces, 3D meshes, textures, as well as lighting and reflection systems. Depending on the users involved and the purpose of the VX, the VE may reach considerably high levels of complexity. Consequently, this complexity substantially increases the efforts required during the development phase of a VR application. To partially tackle this challenge,

workflows adopting DTTs for Environment Reconstruction (ER) have proven to be really useful [67]. This kind of DTTs (e.g., photogrammetry, descriptive models) has been widely explored by the research community for mirroring real-world environments in VXs, e.g., to create educational platforms for the maintenance of buildings [4], and to design VR applications for accessibility within cultural heritage sites (e.g., museums) [34]. However, some challenges still remain. Firstly, utilizing a specific DTT requires a certain knowledge for building functional and efficient systems, highlighting the need of competences in both the VR and the ER domains. Secondly, issues such as the affordability and costs of closed-source tools for reconstructing environments are not yet fully addressed challenges in this context.

Further considerations can be made regarding other elements typically presents in VXs. In both research and commercial VR applications, one method of promoting HCI and improving the UX is the inclusion of embodied agents, such as the traditional Non-Player Characters (NPCs) found in videogames. When equipped with conversational capabilities, these entities are named Embodied Conversational Agents (ECAs), and their design encompasses aspects ranging from mental (e.g., personality, perception, reasoning) and embodiment (i.e., external look) characteristics to those related to the performed actions [225]. These agents may improve VXs for different reasons: firstly, they enhance the realism of VEs by imitating human behaviour; secondly, they can provide guidance to users for navigating the environment, as well as completing certain tasks; finally, they often contribute positively to user engagement and UX [9]. With advances in generative AI-based DTTs, ECAs have been greatly enhanced in their conversational abilities and explored within a wide range of VR applications [218]. Nevertheless, this kind of DTTs is relatively new in VXs and not yet fully exploited. Some technologies, such as Large Language Models (LLMs), are widely employed to create interactive and engaging ECAs, while others, such as Vision Language Models (VLMs) have received little attention. A deeper exploration of these technologies could be beneficial to improve not only agents' conversational abilities, but also their reasoning and perception skills.

To summarize, several challenges still persist within the field of VXs, both when VR is employed in isolation, and when it is combined with other DTTs. These challenges, which constitute the focus of this doctoral thesis, can be summarized as follows:

- The presence of users is a core aspect to consider when creating a VX, assuming the need of user-centered design principles specifically tailored to the target individuals [174, 187]. For some categories of users in certain domains, this element is often under-explored. In this thesis, a specific population of overlooked users is taken into account, namely the one with ASD.
- The creation of VEs is another challenge in terms of time and effort, which

has been partially addressed by combining ER technologies with VR [67]. However, supporting this convergence of DTTs requires expertise in both VR and ER; additionally, despite the presence of open-source alternatives, expensive tools are often employed to efficiently reconstruct the environments. These considerations indicate the need of applications that can easily mediate between technologies, adopting workflows possibly based on open-source systems. Within this thesis, this topic is tackled by taking advantage of Building Information Modeling (BIM) for the creation of VXs.

- Finally, the integration of generative AI-based DTTs within VXs for creating ECAs has not yet reached its full potential, suggesting the need of a deeper exploration on the subject [225]. In this thesis, this challenge is addressed by exploring the use of LLMs and VLMs for creating and/or enhancing ECAs in VXs, and their impact on UX and users' perception.

Building on the previous considerations, this thesis seeks to investigate the above challenges, aiming at filling the current gaps and enhancing VXs across different scenarios. To achieve these objectives, the following Macro Research Questions (MRQs) are formulated:

- **MRQ1:** How can user-centered design principles tailored to specific target individuals can be applied in VXs to create accessible experiences across educational, entertainment, aesthetic, and usability dimensions?
- **MRQ2:** To what extent can the convergence of open-source ER-based DTTs and VR optimize the workflow for reconstructing VEs and developing VXs?
- **MRQ3:** How can generative AI-based DTTs be leveraged to design and enhance ECAs in VXs, and how do these approaches impact UX and users' perception?

The following chapters present a series of investigations that further elaborate on the above MRQs by declining them into specific research questions and addressing the identified challenges. Each investigation focuses on specific aspects of the devised research questions, providing possible solutions and empirical insights that collectively contribute to the overall objectives of this thesis. The author strongly believes that the outcomes of the carried out studies during the years of the Ph.D. could possibly highlight new paths within this domain, as well as potential developments in the future.

1.2 Thesis Structure

This thesis is divided into five chapters. Chapter 1 serves as an introduction to the context in which the thesis is placed, providing an overview of DTTs and their

combination within VXs. The three following chapters regard the activities carried out during the years of author’s Ph.D. research. Each chapter is divided in sections representing one or multiple investigations.

In particular, Chapter 2 includes an investigation set in the context of VX design employing immersive VR. Specifically, it considers a category of users often overlooked within certain domains, namely those with ASD. Since these types of users usually experience discomfort in relation to social presence (e.g., crowds) and sensory overload (e.g., bright lights, sudden noises), the research explores the creation of inclusive VXs that leverage immersive VR technologies. By defining a methodology based on user-centered design principles, a specific emphasis is put on VXs within the Cultural Heritage (CH) domain. Since most research on ASD in this context remains theoretical, the methodology was applied to a real-world use case, developing an immersive virtual visit in a museum specifically tailored to users with ASD. The workflow for creating the VE of the application involved manual creation and the adoption of ER technology, specifically 3D photogrammetry, to digitally reproduce a hall of a real-world museum. Experiments with ASD users were also conducted, obtaining results that confirmed the efficiency of the developed methodology. Additionally, although the adoption of ER technology was not directly the main focus of the investigation, it highlighted the high effort and time-consuming process required for reconstructing the environment and creating the experience.

The above considerations have led research activities to focus on workflow automation for reconstructing VEs in VXs. Chapter 3 addresses this challenge by describing an investigation involving alternative ER technologies for creating VEs. More specifically, it investigates the usage of BIM, usually adopted for the planning and maintenance of buildings, in the scenario of VXs for CH sites. Since most of the current solutions rely on proprietary tools, the investigation proposed a semi-automated BIM-to-VR workflow that adopts an open-source file format to reconstruct the VE and create the experience. The innovative potential of the workflow is showcased in a real-world use case, leveraging a VR platform that automatically reconstruct the VE of a museum and build a virtual tour starting from an enriched BIM file.

Subsequently, to face the third aspect of VXs mentioned in Section 1.1, Chapter 4 presents two investigations exploring the use of DTTs based on generative AI for the creation of conversational agents. The first is placed in the context of UX within the CH domain. In order to achieve an optimal design for a unified UX across multiple VXs, the research takes into account two elements: VE and ECA. For the latter element in particular, the investigation proposed an ECA architecture that leverages LLMs and prompt engineering guidelines for building efficient agents. From this preliminary investigation, the research shifted towards the enhancements of specific capabilities of ECAs, namely their perception of the environment. Thus, the second investigation in this chapter regards the creation of

environment-aware ECAs. Since this capability is often a core feature for improving interaction and conversation quality between ECAs and users, a possible solution relying on VLMs is proposed. Therefore, the investigation introduces an approach to integrate a VLM-based module into traditional ECAs to enable environmental awareness. The approach was validated within an existing VR application derived from the state-of-the-art, i.e., a videogame. Additionally, an exploratory testing phase with users was conducted to evaluate the efficiency of the devised approach and the potential of VLMs in this context.

Finally, Chapter 5 presents a comprehensive summary of the conducted investigations and the overall conclusions of the thesis. The chapter also discusses the main limitations of the proposed approaches and outlines possible future implementations and research directions.

Chapter 2

VX Design and Immersive VR

The work described in this chapter has been formerly published in [71].

The design of an immersive VX can be influenced by several factors. Starting from the general purpose of VR applications, VX design should aim at establishing a relationship between the user and the VE [157]. To achieve this goal, a user-centered approach should be adopted for understanding the user's needs in the specific context of use [226]. The identification of these needs should be functional to the definition of key design principles to be later employed in a virtual application. This aspect gains extreme importance especially in studies related to accessibility for users with disabilities, whose VR experience may be hindered by physical or perceptual barriers [95]. Once these principles have been defined, it is then possible to select the modalities of interaction, locomotion, and devices from the wide range of VR technologies, as well as the concurrent usage of other DTTs. At the same time, iterative design and active user participation should be taken into account during the entire development process of a VX. Over the years, such an approach has been adopted to develop engaging and effective VXs in several domains. However, some categories of users in certain contexts are still not sufficiently considered. One of these categories is represented by users with ASD.

This chapter presents an investigation employing immersive VR to explore the topic of ASD accessibility within CH sites, mainly focusing on museums. Such environments are often challenging for users with ASD, who may experience several discomforts caused by sensory and social elements (e.g., bright lights, crowded spaces) [85]. The use of immersive VR has often proven to be effective in addressing these issues [52]. However, a solid methodology for designing VR-based museum experiences for ASD users seems to be lacking. Thus, the investigation proposes a structured methodology involving the definition of core VX design principles specifically tailored to users with ASD. Additionally, it applies these principles in a use case, creating a VX based on the reconstructed environment of a real-world museum hall. The reconstruction of the VE was made possible by manually creating the base structure of the museum hall, and by employing an additional DTT related

to ER, namely 3D photogrammetry. The outcome of the investigation, finalized with a series of experiments with ASD users, helped the author understand the strengths and limitations of the devised methodology, and reflect on the substantial effort required to build the VE and create the VX.

2.1 Accessible Virtual Museum Experiences for ASD Individuals

This section describes the said investigation, carried out in the context of CH accessibility for ASD users employing immersive VR as the main DTT. The following subsections include an overview of the topic, the related research within the context, the devised methodology and its application in a use case, and a user study with ASD individuals.

2.1.1 Overview

Over the last twenty years, the notion of accessibility has progressively emerged as a key topic within the CH field, attracting growing attention from both cultural institutions and academic research. In literature, accessibility is commonly defined as the capability of a service or product to be effectively used by all individuals, independently of personal attributes such as disability, gender, and age [54]. This increasing awareness has contributed to a profound shift in the objectives of museums and heritage sites, which have gradually moved from collection-centered approaches toward approaches that emphasize visitor experience and inclusivity [69]. Within this perspective, accessibility has come to be recognized as a fundamental factor for promoting social inclusion, supporting participation in terms of individuality, community, and society [187].

Despite this progressive shift toward inclusivity, significant challenges remain in achieving effective accessibility within CH contexts. Such difficulties are particularly evident for individuals with specific needs, including people with a physical and/or Intellectual Disability (ID). In this regard, individuals with IDs often encounter several obstacles that may hinder learning processes, personal well-being, and opportunities for social participation [86]. Within this category, individuals on the autism spectrum, i.e., people with ASD, represent a population with highly heterogeneous characteristics. By definition, ASD indicates a wide diagnostic range, encompassing several challenges that differ considerably across individuals. These may involve impairments in social communication, the presence of restricted or repetitive behaviors in terms of physical movements and interests, as well as a high prevalence of concurrent neurodevelopmental conditions, such as Attention-Deficit/Hyperactivity Disorder (ADHD) [7].

In museum environments, the main challenging elements for ASD visitors are the environmental adaptation and the cultural learning [46, 80]. Specifically, the former concerns the ability of individuals with ASD to become accustomed to the physical and sensory characteristics of museum spaces. This process is frequently hindered by the presence of intense or unpredictable stimuli, such as sudden noises and bright lighting conditions, as well as by common sensory overload factors (e.g., crowded areas, elevated sound levels) [74, 98, 101, 209]. The presence of such elements may result in discomfort for visitors with ASD, along with the increase of the levels of anxiety, stress, or fatigue [46]. The latter element, cultural learning, refers to the difficulties that visitors with ASD may experience when engaging with museum content [80, 145]. Such difficulties are often present when inclusive strategies to present the cultural information are not employed [99]. The combination of these elements can significantly hinder UX and limit cultural engagement, both of which are central components of contemporary museum practice [85]. As a result, improving accessibility for visitors with ASD represents an important challenge for worldwide cultural institutions.

Over the years, numerous studies have proposed strategies to mitigate the challenges faced by museum visitors with ASD. The most commonly adopted measures include the introduction of sensory-friendly activities and specific quiet hours [204], the development of collaborative initiatives with autism support organizations [140], and the use of co-design strategies that directly involve individuals on the autism spectrum in the design process [96]. As an alternative solution, the use of VR has emerged as a particularly promising approach to support accessibility in CH contexts. The application of VR technologies for individuals with ASD has been extensively investigated across different domains, including social skills training [3, 52] and educational settings [40]. However, their adoption in the context of museum experiences has so far remained more limited [37, 53]. The usage of VR encompasses several advantages, such as the creation of immersive, isolated, and safe VEs, which can help reduce the sensory overload experienced by visitors with ASD [52]. Additionally, by digitally reconstructing real-world spaces, VR can facilitate the transfer of acquired skills to physical settings, allowing users either to virtually explore a museum [77] or effectively prepare for an in-person visit [60]. Lastly, VR supports the design of controlled and highly customizable simulations centered on user needs, an aspect that is particularly beneficial for individuals with ASD, who often prefer predictable environments [156].

Despite the benefits of employing VR, other crucial factors must be taken into account. Among them, there is the integration of inclusive design and immersion practices specifically tailored to ASD users [61]. Beyond the CH domain, such practices have been extensively investigated, aimed at creating ASD-oriented design guidelines within VXs. For instance, some studies highlight the importance of enabling task repetition, minimizing unnecessary informational complexity, and structuring activities around clear and consistent learning routines [26]. Moreover,

the creation of highly realistic VEs has been identified as a relevant design strategy, as realism may facilitate faster adaptation to the virtual space and enhance user motivation to engage with learning activities [45, 135]. Other studies have emphasized the role of gaze-based interaction and attention-guidance mechanisms [121], since individuals with ASD tend to process visual information more effectively than other sensory modalities [219].

The previously mentioned design guidelines should also be taken into account within the CH domain, particularly when developing inclusive Virtual Museum Experiences (VMEs). In function of the ongoing trend toward the digitization of CH [228], such VMEs, including virtual tours and virtual exhibitions, have seen widespread adoption by museums worldwide [136, 177, 197]. However, significant challenges remain within this context. With respect to CH accessibility, most of the criteria proposed to support ASD users have not yet been universally accepted [158]. This may also be the reason why the topic is currently underexplored. Moreover, this consideration remains valid also from the perspective of virtual design, where a generally accepted methodology for designing inclusive VMEs seems to be lacking.

The investigation described in this section aims to overcome the challenges in the use of VR for individuals with ASD within museum contexts. To this end, a methodology for designing tailored, accessible, and engaging VMEs is proposed. Specifically, the methodology is structured into three main phases: the identification of several inclusive principles specifically developed for ASD visitors, which shape the design of a VME; the definition of guidelines for preparing the materials required to develop the application, such as the selection of cultural contents and the digital reconstruction of the VE; and finally, the implementation of co-design sessions to evaluate and refine the VME design, ensuring alignment with the needs of ASD users and the expectations of all the involved stakeholders.

The developed methodology was employed for a specific scenario, namely the “Hall of Seasons” in Palazzo Carignano, Turin, Italy. Since the selected scenario involved an environment characterized by multiple elements potentially hindering for ASD users, such as decorations and bright lighting, it was deemed suitable for this investigation. The objective was to develop an inclusive VME, following the developed design principles, inspired by a real-world visit. Subsequently, co-design sessions were conducted for the validation of the VME with ASD users and stakeholders, including museum curators, ASD organizations, a psychologist, and a psychiatrist. These sessions were essential for refining the VME through iterative feedback, allowing for targeted adjustments and improvements. Lastly, a testing phase involving ASD visitors was conducted in order to analytically compare the developed inclusive VME with its real-world counterpart.

2.1.2 Background

A substantial body of research has investigated accessibility and inclusive virtual design practices for supporting individuals on the autism spectrum. This section presents an overview of the current state of the art across CH, VR, and VMEs specifically related to ASD. Specifically, it highlights key findings, recent advancements, and unresolved challenges, providing the conceptual context for the proposed approach.

ASD and Cultural Accessibility

In recent years, increasing attention has been devoted to accessibility within both scientific research and CH institutions. Historically, museums were conceived as spaces accessible only to selected groups of visitors; however, contemporary practice increasingly recognizes the necessity of ensuring that cultural resources are available and meaningful for all audiences [81]. In order to comprehend this necessity, it is essential to clearly define the concept of accessibility and to examine the existing guidelines that support it within museums and cultural institutions.

A comprehensive conceptualization of accessibility in the CH domain has been introduced by Deffner et al. [54], who distinguished between two primary phases: physical and perceptual accessibility. Physical accessibility regards the visitor’s ability to interact with cultural assets through sensory channels, such as vision, hearing, and touch. Barriers at this level may arise from physical impairments as well as from external factors (e.g., admissions costs). Perceptual accessibility, by contrast, relates to the processes of understanding and interpreting cultural content, which depends on the cognitive abilities and the educational background of every individual. As discussed in the study, this dimension may present significant challenges for visitors with cognitive or learning disabilities. Moreover, the existence of a third phase is highlighted, namely “appropriational accessibility”, which integrates physical and perceptual dimensions. This phase enables visitors to establish a meaningful emotional relationship with a CH site.

Based on these definitions, it is essential for CH institutions to adopt practices that address all the previous dimensions of accessibility. In particular, physical accessibility in relation to visitors with ASD has been the focus of several research investigations.

For example, Hutson and Hutson [99] examined existing best practices and guidelines aimed at improving accessibility for ASD individuals within CH sites (e.g., museums). Among these guidelines, the authors highlighted the use of sensory maps designed to inform visitors about potentially challenging elements, such as crowded areas, intense lighting, or specific smells. In contrast, initiatives such as sensory-friendly days or special opening hours were found to be negative elements in terms of social inclusion. The study further proposed additional strategies, including the provision of sensory-support tools (e.g., headphones or earplugs), alternative

instructional materials, such as Picture Exchange Communication System (PECS), and audio guides. Lastly, the authors emphasized the value of collaboration with organizations and support groups specialized in ASD to enhance the overall visitor experience.

Similarly, Lussenhop et al. [140] investigated cultural participation among families with ASD children. Through qualitative interviews, the study explored museum visit experiences, highlighting both positive aspects and recurring challenges encountered during real-world visits. A key outcome of the analysis was the identification of specific environmental features of museum spaces that can significantly influence the cultural experience of families with ASD children. In line with the findings of Hutson and Hutson [99], the authors emphasized the importance of establishing connections with ASD support programs and associations to improve accessibility. However, the study also reported that environmental barriers, such as excessive noise levels and overcrowded spaces, could negatively affect museum experiences, even when mitigation strategies are employed.

Another relevant study was carried out by MacLennan et al. [143], who examined the sensory experiences of autistic adults in public environments. By utilizing a focus group methodology, the study analyzed participants' responses in accordance to a set of evaluative principles. Although museums were perceived as more accessible than other public spaces, such as restaurants or public transportation, challenges related to environmental predictability were still reported. In particular, unpredictability was associated with uncertainty in relation to potential sensory stressors (e.g., crowd density and noise levels), as well as with inconsistencies, unfamiliar settings, and insufficient information to support effective pre-visit planning.

Building on the studies discussed above, the presence of disruptive environmental factors, such as excessive noise, intense lighting, and overcrowded spaces, appears as a critical aspect regarding accessibility for visitors with ASD. These elements not only affect visitors' well-being, but may also interfere with perceptual accessibility. As a result, several approaches have been proposed to enhance perceptual accessibility for visitors with learning and neurodevelopmental impairments, including individuals on the autism spectrum.

Within this context, Rappolt-Schlichtmann and Daley [182] investigated the application of Universal Design for Learning (UDL) principles in museum environments. The set of UDL principles, represented by seven items, have been extensively explored in literature across various fields, aiming at designing products and environments usable by individuals of all ages and abilities [202]. In the study of Rappolt-Schlichtmann and Daley, the authors observed that, while traditional physical accommodations, such as ramps or subtitled videos, have contributed to improving access to museum spaces, they often were not efficient in supporting deeper learning and engagement. To address this limitation, the study encouraged the adoption of UDL strategies in museums, including the use of multimedia resources, interactive elements, and personalized challenges tailored to diverse

learning needs.

Similarly, Guedes et al. [86] emphasized the importance of actively involving individuals with IDs in the design of accessible museum technologies. Their work focused on the development of the *ACCESS+* application through an iterative co-design process, incorporating features such as Augmentative and Alternative Communication (AAC) pictograms, simplified user interfaces, and text-to-speech functionalities. The results indicated that participatory design approaches could possibly enhance both usability and emotional engagement in digital applications for museums.

ASD and VR

The adoption of VR to support individuals with ASD and other cognitive impairments has been widely investigated, largely due to the advantages offered by this type of DTT. VR enables users to practice skills within safe and controlled environments, allows precise regulation of sensory stimuli, supports adjustable levels of task complexity, and facilitates the design of highly structured and predictable scenarios tailored to individual needs [26]. Thus, VR has been applied across a broad range of contexts, including rehabilitation [50], the development of real-world social [90, 186], emotional [17], and safety-related skills [88], as well as special education settings [24]. Research in this area frequently relies on immersive VR solutions, often delivered through HMDs, to enhance engagement and learning outcomes. Furthermore, the adoption of inclusive immersion strategies has been shown to improve accessibility and usability of VR applications for users with diverse sensory and cognitive profiles [61].

A relevant contribution to the clarification of these concepts was provided by Dudley et al. [61], who distinguished between inclusive design and inclusive immersion. Inclusive design was described as a set of practices aimed at maximizing accessibility across heterogeneous user groups, whereas inclusive immersion specifically addresses accessibility challenges within immersive VR experiences. Their analysis of existing inclusive immersion approaches revealed several open challenges, including the difficulty of applying universal design principles in practice given the diversity of user needs, as well as the limitations faced by researchers when relying on too general guidelines.

Building on the concept of inclusive immersion and the challenges associated with its practical application, several studies have proposed design principles tailored to specific users. In this direction, Guitton et al. [87] introduced a set of guidelines for immersive serious games intended for users with cognitive impairments. These principles focused on the design of intuitive interfaces with simple navigation, careful management of environmental level of detail, the use of clear visual and auditory cues, and reduced control complexity. The authors also highlighted the relevance of providing real-time feedback and adopting inclusive design

processes that actively involve individuals with cognitive disabilities throughout the development phase.

A further example is provided by Cai et al. [35], who described the design and implementation of an immersive and interactive Virtual Dolphinarium. The main objective was to support children with ASD with respect to learning nonverbal communication skills. Unlike conventional dolphin-assisted therapy, the proposed system offered a safe and fully controlled virtual setting in which children assumed the role of dolphin trainers, interacting with virtual animals through hand gestures. The experience was supported by a 320-degree curved display, stereoscopic 3D visualization, and gesture-based interaction. While a pilot study involving participants with ASD was conducted, the results also highlighted limitations, as more than half of the users experienced sensory overload during the VR session.

In addition to generic VR applications, a substantial portion of research on inclusive immersion for users with ASD has concentrated on training and educational scenarios. Within this domain, Lorenzo et al. [135] proposed an immersive VR system designed to support the development of emotional skills in students on the autism spectrum. The approach relied on realistic social situations that allowed participants to practice emotion recognition and appropriate responses. The results reported significant improvements in emotional competencies, underscoring the importance of realism in VEs to facilitate adaptation and to promote the transfer and retention of skills acquired during the VX.

From an educational perspective, Cecil et al. [40] investigated the design and evaluation of Virtual Learning Environments (VLEs) aimed at supporting Science, Technology, Engineering, and Mathematics (STEM) education for students with ASD. The study leveraged immersive VR to develop interactive and modular learning units, each addressing a specific topic such as robotics or manufacturing. By conducting pre- and post-intervention assessments, the authors demonstrated that VR-based VLEs can improve both engagement and comprehension in ASD learners, highlighting the potential of modular immersive tools to enhance accessibility and effectiveness in STEM education.

Another relevant contribution adopting a modular perspective is the work by Meins et al. [149], which addressed the treatment of individuals with social impairments through VR-based interventions. The authors highlighted how modular treatment structures can support personalized approaches by targeting specific difficulties related to social functioning, such as social anxiety and paranoia. In this context, they introduced VR-SOAP, a VR-based treatment composed of multiple independent modules designed to encourage social engagement and participation. Each module focused on a particular challenge and relied on realistic social simulations implemented within several VEs. Although the study was presented as preliminary, its emphasis on modularity suggested promising potential for scalable and customizable interventions.

Beyond modularity, several studies in the field of VR for ASD have focused on

visual attention as a key design element. Compared to neurotypical individuals, users with ASD often show differences in both selective attention, such as difficulties in rapidly shifting focus between visual elements, and sustained attention, characterized by maintaining focus over time on specific tasks [219, 18, 76]. These characteristics indicate that the integration of explicit attention-guidance mechanisms may play a crucial role in the design of effective VXs for individuals with ASD.

With the above considerations, multiple strategies for guiding user attention within VEs have been proposed. Lee et al. [126], for example, introduced a diegetic attention guidance system aimed at improving turn-taking in social VR group meetings. The proposed approach addressed several challenges, such as reduced peripheral awareness caused by limited fields of view and the difficulty of perceiving non-verbal cues, including gestures. The results demonstrated a reduction in response times, suggesting that such multi-modal guidance techniques may be beneficial for users with reduced social abilities, including individuals with ASD.

Another relevant study is provided by Amat et al. [6], who developed a system specifically targeting joint attention skills in children with ASD. In this scenario, joint attention was defined as the ability to align one’s focus with another individual toward a specific object or event. With the objective of improving this ability for autistic children, the proposed desktop-based VR application integrated gaze tracking to monitor attention toward predefined Region of Interest (ROI), as well as adaptive gameplay mechanisms. While the findings indicated improvements in gaze-related behaviors, the authors acknowledged methodological limitations, including a small sample size and the absence of a control group.

VR and Cultural Accessibility for ASD

Although still relatively under-explored, some research studies have proposed approaches specifically leveraging VMEs to enhance cultural accessibility. One example is the work by Messina et al. [151], which introduced “Social Stories” as a medium to support museum accessibility for visitors with cognitive impairments. The authors employed low-cost wearable VR devices to develop an immersive VME in the form of a virtual tour based on 360° videos. The experience was further enriched with visual elements (e.g., geometric shapes and images), auditory cues, and interactive components to increase engagement. An evaluation of the VME was conducted through workshops involving domain experts, including educators and therapists, who recognized the potential of the approach and suggested improvements. In particular, it was emphasized the need for explicit visual “facilitators”, such as highlighted areas guiding users’ attention toward specific regions of the VE.

Meanwhile, other studies have addressed cultural accessibility through VMEs with a more specific emphasis on ASD. Giaconi et al. [78], for instance, conducted a pilot investigation aimed at improving CH accessibility for users on the autism

spectrum through the adoption of VMEs. Their methodology relied on a co-design process involving both experts (e.g., psychologists and museum curators) and individuals with ASD, and was structured into three main phases: identification of accessibility challenges, on-site data collection (including 360° images and cultural data), and the prototyping of an inclusive VME. The resulting experience, implemented using the Marzipano web-based tool, incorporated several accessibility criteria, such as anticipatory exposure to museum spaces, realistic environmental representation, improved access to cultural content, and clarification of symbols used within the museum context.

Another contribution is represented by Drageset et al. [60], who presented an exploratory study on a pre-visit VME designed to support individuals with ASD and their families. The proposed solution consisted of a non-immersive virtual tour developed using Matterport and 360° images, with the goal of reducing sensory overload and increasing familiarity with the museum environment prior to a physical visit. The VME allowed users to explore exhibits in detail through features such as zooming on cultural objects and navigation via interactive hotspot or Point of Interest (POI). The study reported positive outcomes related to the absence of crowds and audio stimuli, while also identifying limitations concerning POI design and textual clarity.

Practical Use Cases and Stakeholder Perspectives

Beyond academic research, many of the accessibility practices discussed above are already adopted by museums worldwide. The British Museum in London, for example, implements a range of accessibility measures tailored to different visitor groups, including individuals with ASD. These measures involve offering preparatory resources, such as sensory maps, before the visit, as well as providing sensory backpacks and other supportive tools during the museum experience. A similar strategy is employed by the V&A Museum of Childhood (now Young V&A), which uses sensory kits and PECS-based communication symbols to facilitate cultural engagement for children with ASD, and also provides quiet rooms and designated low-stimulation visiting hours. However, as highlighted by Hutson and Hutson [99], such approaches may not necessarily foster social inclusion.

In recent years, some museums have begun experimenting with VMEs as an alternative way of improving access to CH. The Solomon R. Guggenheim Museum in New York, for instance, has incorporated virtual activities into its “Guggenheim for All” program [208]. As discussed by Di Lello [55], this initiative, launched in 2014 and grounded in UDL principles, aims to support cultural education for students with ASD. The program is structured around several principles, such as offering multiple forms of representation (e.g., text, images), reducing exposure to potentially overwhelming sensory stimuli (e.g., loud sounds), and promoting on-site learning within the museum environment as an alternative pedagogical approach.

A more recent exploratory study at the National Archaeological Museum of Naples (MAN) investigated the use of an immersive video tour specifically created for ASD visitors [169]. In this case, the designers intentionally adopted a stylized, non-realistic visual aesthetic to mitigate sensory overload caused by lighting, shadows, or reflections—factors frequently reported as challenging for individuals with ASD. However, as noted in other studies, such stylization may limit the user’s ability to “anticipate” the real-world CH environment.

Despite these efforts, VR-based accessibility initiatives remain relatively rare in museum contexts, particularly for individuals with perceptual or cognitive disabilities. Some curators express concern about VR, perceiving it either as incompatible with traditional curatorial principles or just as a trend-driven technological addition [192]. Moreover, the development of inclusive VR experiences often requires coordinated work among clinicians, caregivers, technical developers, and museum professionals, making these projects demanding in terms of resources and organization.

In any case, existing research indicates that VMEs can provide distinctive benefits compared to in-person visits, especially regarding narrative structuring, interactive engagement, and the ability to modulate sensory input [107].

In conclusion, although interest in digital and VR-based accessibility tools is increasing, their practical implementation within museums is still limited. This scenario underscores the importance of balancing institutional priorities, user requirements, and interdisciplinary collaboration to enable accessibility within the CH sector.

Considerations

Figure 2.1 provides an overview of the main findings emerging from the analyzed literature. Overall, the reviewed works clearly indicate that the adoption of targeted practices can significantly contribute to improving both physical and perceptual accessibility for ASD users. Several of these practices occur several times across different studies and application contexts, including the provision of pre-visit resources (e.g., sensory maps), in-visit support tools (e.g., sensory aids and quiet rooms), simplified representations of cultural content (e.g., PECS and AAC systems), and the involvement of individuals with ASD and related support organizations through co-design and participatory approaches. Despite the adoption of such strategies, some studies still reported issues related to users’ well-being, particularly sensory overload. These difficulties appear to be closely linked to the unpredictability of museum environments, which remains a critical barrier for visitors with ASD.

In response to the challenges posed by real-world unpredictability, numerous studies have explored alternative solutions based on controlled VR environments.

Within this research area, a number of design principles and guidelines aimed at inclusive and effective immersive VXs have emerged. In particular, studies conducted in training and educational contexts suggest that VR applications benefiting users with ASD often rely on structured routines, modular designs, and realistic scenarios. Moreover, research regarding attention has highlighted the role of visual (e.g., lighting cues and ROI) and auditory signals in effectively directing users' focus within VEs.

With specific regard to VMEs designed for individuals with ASD, the use of 360° imagery represents a widely adopted solution for achieving realistic representations of museum spaces. This approach enables users to preview environments and gradually familiarize with them, positively influencing their overall experience. However, the static nature of 360° images tends to restrict user interaction, thereby revealing the potential advantages offered by more interactive and immersive 3D VEs.

Across the literature, predictability is frequently defined in terms of “environmental anticipation,” particularly in relation to VMEs that allow users to explore a museum before a physical visit. While anticipation supports familiarization with visual and spatial aspects, such as layouts and exhibits, this interpretation only partially reflects the broader concept of predictability. Particularly, this concept also involves the capacity to consistently interpret and navigate an environment in real time, thus minimizing uncertainty and cognitive stress. This distinction is crucial for individuals with ASD, who typically benefit from environments characterized by stable and clearly structured cues. Consequently, approaches aimed at enhancing predictability should extend beyond pre-visit experiences and provide continuous support throughout the visit itself.

Finally, although an increasing number of studies address accessibility for individuals with IDs, relatively few contributions specifically focus on ASD within the context of VMEs. Furthermore, even when inclusive immersion principles are considered, a lack in terms of standardization and user-specific design seems to be present. These gaps highlight the need for a more systematic and target-oriented approach.

2.1.3 Materials and Methods

This section describes the proposed methodology, which is structured around three main phases: a series of inclusive immersion principles intended to guide the design of a VR application featuring a VME specifically conceived for visitors with ASD; a collection of guidelines supporting the preparation of materials required for the application development, including criteria for selecting cultural contents and defining the workflow for constructing the VE; and a set of co-design sessions aimed at evaluating and iteratively refining the proposed solution, ensuring its suitability for ASD users and stakeholders.

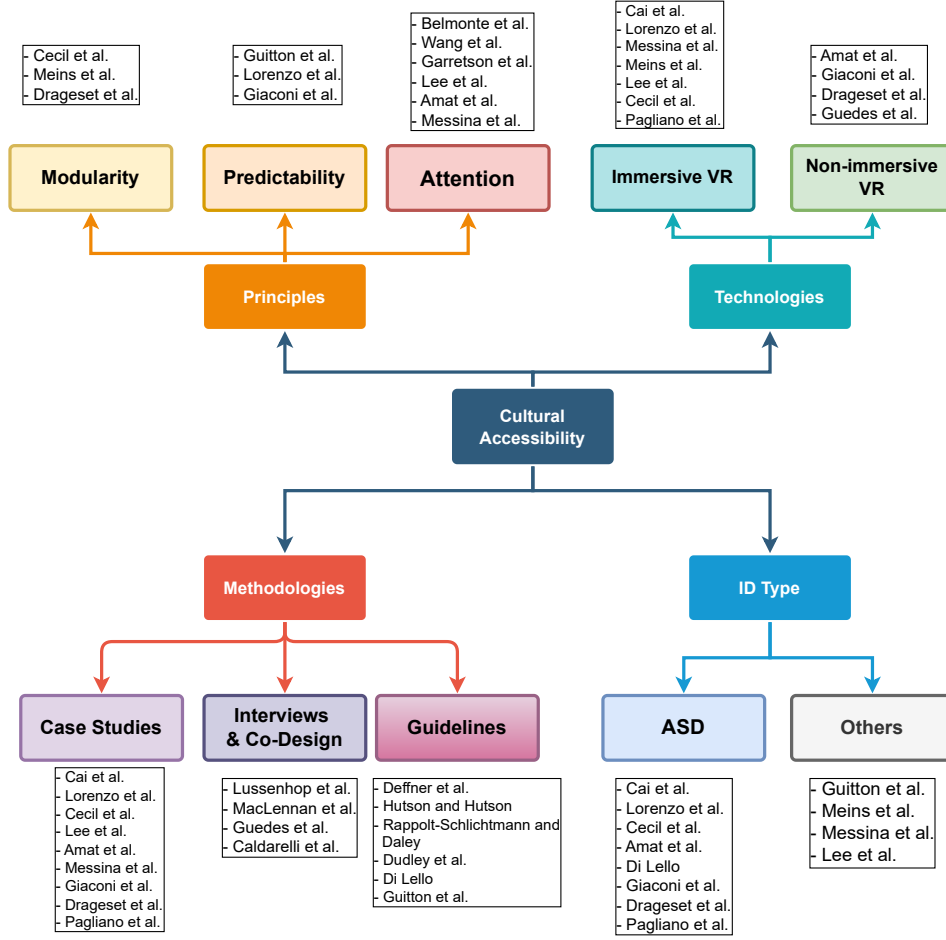


Figure 2.1: Key points of the analyzed works, including the inclusive immersion principles addressed, the technologies used, the methodologies adopted, and the ID types considered.

Inclusive Design Principles

Based on the analysis of the state of the art, in this investigation a set of inclusive immersion principles specifically tailored to ASD users are introduced, which can be integrated into the design process of a generic VR application featuring a VME. The proposed principles, namely modularity, predictability, and attention, are intended to improve the VR experience of users with ASD while interacting with virtual elements, like User Interface (UI) components, navigating the VE, and accessing cultural content.

Modularity Principle

The first principle, modularity, aims to address potential challenges experienced by users with ASD, particularly those related to sensory overload and learning. As highlighted by previous studies, the adoption of structured modules, simplified navigation mechanisms, intuitive controls, and realistic environments can address these issues [40, 26]. However, these measures alone are often insufficient to fully prevent sensory overload. This limitation is especially evident in museum contexts, which typically contain a high density of visual stimuli, as well as in VMEs, where multiple virtual objects and interactive interfaces may coexist. When presented simultaneously or within a short temporal window, these elements may negatively affect the cultural experience of the target users. To address this problem, the proposed methodology defines modularity according to two complementary dimensions: *VE modularity* and *application modularity*.

VE modularity refers to the conceptualization of the VE as a collection of sub-elements, including both 3D components and UI elements (e.g., buttons and panels). This structure enables the sequential presentation of these sub-elements, rather than exposing users to all elements at once. Accordingly, the VE should initially not include some of its defining elements, which are then progressively introduced and enriched as the user advances within the VX. This process is referred to as “progressive enrichment” of the VE. To support this enrichment, visual techniques such as fade-in animations can be employed to smoothly reveal new elements.

Application modularity, instead, concerns the capability of a VR application to offer discrete and structured modules aimed at supporting the presentation of cultural content. Each module is defined through a task-based process, referred to as a module procedure, which includes a sequence of user actions (e.g., pressing a virtual button or moving to a specific location) and associated cultural information (e.g., historical descriptions). Access to the cultural content should be granted only upon completion of all the required tasks. In addition, the VR application should provide clear instructions for task execution, as well as support access to cultural information (e.g., through textual UI panels or auditory cues).

Given the importance of providing a clear and well-defined routine for users with ASD, the modules should be arranged sequentially, such that each module becomes available only after the completion of the previous one. Consequently, in a typical VR experience, all modules except the first should initially be unavailable, either hidden or non-interactive, and should be unlocked progressively as the user proceeds.

Within the proposed methodology, VE modularity and application modularity are implemented through a linked approach. Each module is defined by a module procedure, a textual cultural description, and a reference to a specific cultural element (e.g., an artwork or architectural detail), which represents a sub-element of the VE. The tasks composing the module procedure are designed to be completed

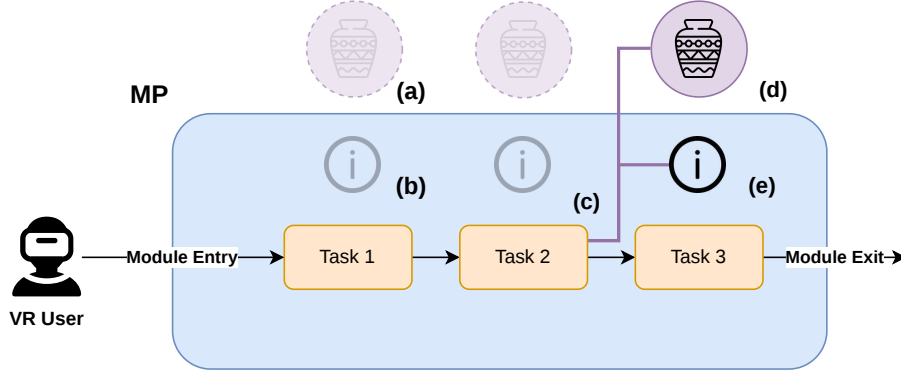


Figure 2.2: Example of the proposed approach combining VE modularity, through progressive enrichment, and application modularity in a VR application. At the beginning of the Module Procedure (MP), both the cultural element (a) and the cultural description (b) are respectively invisible and inaccessible. After completing a predefined number of tasks (c), the cultural element becomes visible (d), and the cultural description (e) can be accessed and explored by the user.

individually through simple interactions with virtual elements (e.g., UI buttons). Once a predefined number of tasks has been completed, the associated cultural description becomes accessible, and the linked cultural element is revealed within the VE. An example of this approach is illustrated in Figure 2.2, showing how VE progressive enrichment and application modularity can be implemented within a VR application.

Finally, task execution within the modules may benefit from the inclusion of a comprehensive tutorial, particularly when combined with the principle of predictability, which is discussed in the following subsection.

Predictability Principle

The second principle, predictability, is intended to mitigate the challenges arising from the dynamic and often unpredictable nature of CH sites. As discussed in Section 2.1.2, unexpected sensory stimuli, such as sudden noises, bright lighting conditions, and overcrowded spaces, have been identified as major factors negatively affecting both well-being and learning experiences of individuals with ASD [140, 143]. Several studies have suggested that VR can partially address these challenges by providing controlled and predictable VEs. However, predictability within a VME cannot just rely on environmental control and should also be supported by inclusive immersion practices.

From a design perspective, fostering predictability requires the adoption of intuitive interaction mechanisms, as widely acknowledged in interface design research

[20]. At the same time, it should tailor interaction procedures to the specific needs of users with ASD. In this regard, best practices identified in the literature emphasize the relevance of structured learning routines and the possibility of repeating tasks. These considerations suggest that integrating repetitive and consistent interaction schemes throughout the VME may significantly improve the cultural experience for the target users.

Based on these premises, the proposed methodology conceptualizes predictability as the capability of a VR system to provide two fundamental elements: intuitive UIs and iterative *interaction routines*.

Intuitive UIs can be achieved by adopting simple and direct interaction paradigms with virtual elements (e.g., touching or selecting a UI button), as well as by employing symbols that are easily interpretable by users with ASD. Examples include pictograms and icons derived from AAC or PECS-based systems [22, 152]. Interaction routines, on the other hand, ensure that users are exposed to consistent *Interaction Patterns (IPs)* across the entire VR experience. Each IP consists of a series of recurring interaction routines, which are in turn composed of a variable number of *interaction steps*. To further support learning and reduce frustration, the VR system should allow task repetition, enabling users to perform the same interaction routine multiple times if additional attempts are required to master a given IP [26].

These interaction patterns are designed to be reused across multiple module procedures, thereby enabling a tight integration between the principles of modularity and predictability (Figure 2.3). To limit cognitive load and maintain a uniform interaction experience, a particularly critical aspect to consider when users with ASD are involved [140, 143], the number of distinct IPs should be kept as low as possible. Each interaction routine is associated with a specific cultural element, while each interaction step corresponds to a single task within the module procedure and is regulated by a timeout. The timeout defines the maximum duration allowed for completing a task. If the timeout expires before task completion, the interaction routine resets to its initial interaction step, reinforcing repetition and consistency. This mechanism is intended to establish a stable and repetitive interaction flow, facilitating task learning for users with ASD. A conceptual representation of such an interaction routine is illustrated in Figure 2.4.

Attention Principle

The third and final principle is attention, which plays a fundamental role in directing users' visual focus toward specific spatial locations or elements within a VE. As discussed in Section 2.1.2, individuals with ASD may experience difficulties both in shifting attention between different objects or events of interest and in maintaining focus on externally imposed activities. Such difficulties can negatively affect cultural engagement, particularly with respect to perceptual accessibility and

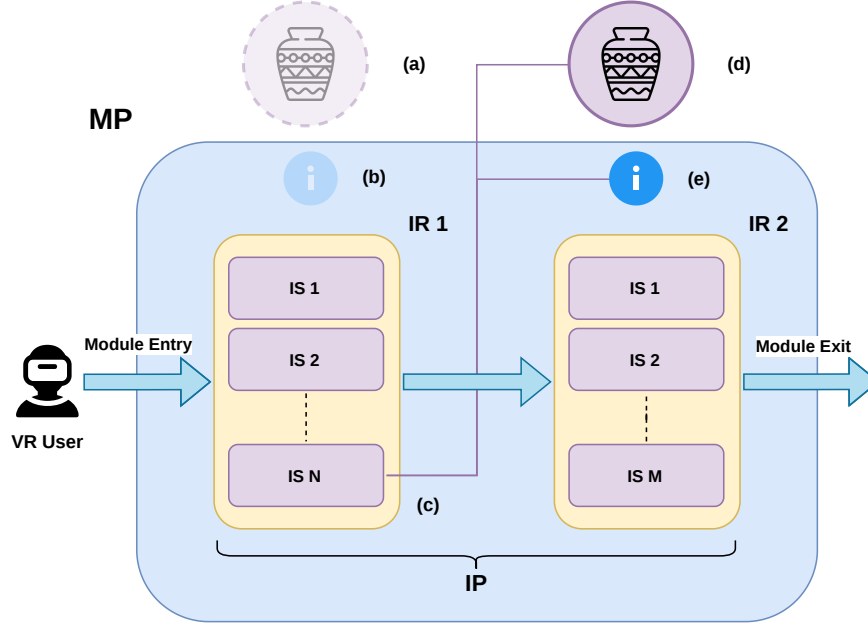


Figure 2.3: Example of a module procedure (MP) integrating the predictability principle. The procedure includes two interaction routines (IRs) within a single Interaction Pattern (IP), composed of multiple interaction steps (IS), a cultural element (a), and a cultural description (b). Differently from the configuration adopting only modularity (Figure 2.2), the tasks are embedded within the interactions steps. Once the final task of the first routine is completed (c), the cultural element (d) and its description (e) become accessible.

learning processes. Prior research in VR indicates that the use of visual cues (e.g., UI elements, highlights), ROIs, and gaze-based interactions can effectively support attention guidance within VEs [6]. Moreover, in the context of VMEs, additional guidelines tailored to CH scenarios have been proposed, including the adoption of *visual facilitators* [77] and zoom-based mechanisms to emphasize specific cultural elements. While these approaches have been mostly applied to VMEs based on 360° images or videos with limited interactivity, adapting them to fully 3D-VEs appears particularly promising.

Based on these considerations, the principle of attention is defined here as the ability of a VR system to effectively guide users' focus toward specific regions or elements of interest within a VE. To further support attention, the involved virtual elements should dynamically adapt to user interactions and be presented in visually accessible ways. Within the proposed methodology, this principle can be implemented through two complementary modalities, which may be employed either independently or simultaneously:

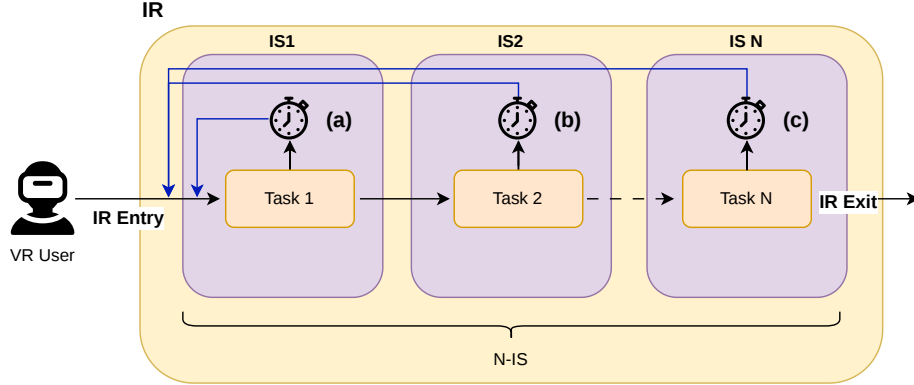


Figure 2.4: Conceptual representation of an interaction routine (IR), composed of multiple interaction steps (IS1-2-3), each associated with a specific task and a timeout constraint (a, b, and c).

- The adoption of a head-gaze-based interaction paradigm, in which the user’s head orientation defines the gaze direction and enables interaction, combined with visual facilitators to direct attention toward a cultural element;
- The spatial repositioning of cultural elements toward visually accessible areas of the VE.

In coherence with the principles of modularity and predictability, both modalities can be formalized as interaction routines embedded within a module procedure. An illustrative example is reported below.

1. **Head-gaze-based interaction routine:** this interaction routine relies on the use of a visual facilitator, such as a UI panel, which dynamically moves within the VE according to the user’s head-gaze direction, detected through the HMD sensors. The interaction steps require the user to visually follow the facilitator as it shifts across predefined spatial positions. To sustain engagement and progressively orient attention, multiple interaction steps are employed, gradually leading the user toward the target cultural element. After a predefined number of interaction steps, the visual facilitator stops at a specific position, then disappears, revealing the cultural element located at that position. In line with the predictability principle, each interaction step is associated with a timeout; if the timeout expires, the interaction routine resets, allowing the user to repeat the sequence if necessary. A conceptual example of this routine is shown in Figure 2.5.
2. **Visually accessible positioning:** once the cultural element becomes visible, a secondary interaction routine composed of a single interaction step

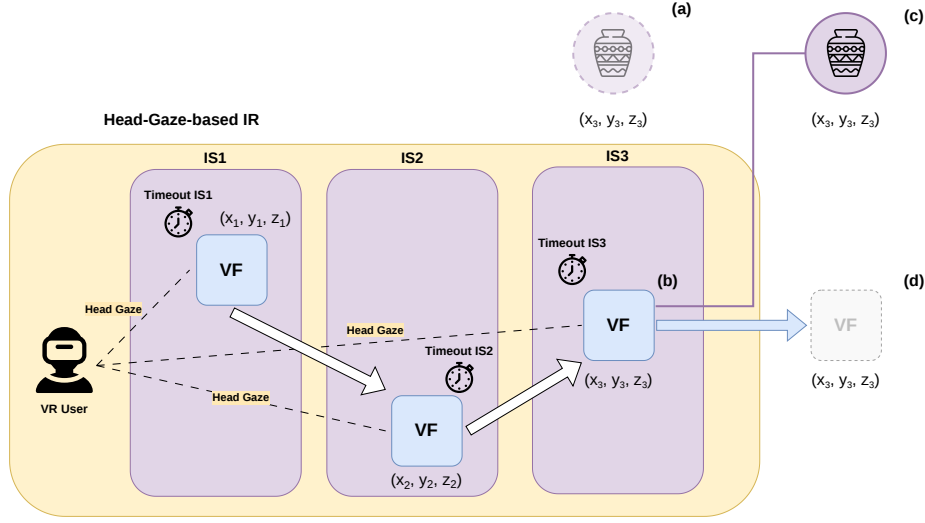


Figure 2.5: Simplified representation of a head-gaze-based interaction routine using a visual facilitator (VF). The routine includes three interaction steps (IS1-2-3), each associated with a timeout, and three spatial positions within the VE (x_i, y_i, z_i) . In accordance with the modularity principle, the cultural element is initially hidden (a). After completion of the final interaction step, the cultural element becomes visible (c), while the visual facilitator disappears (d).

can be used to reposition the element within the VE, ensuring optimal visual accessibility. Although this routine may not be required for all modules, it becomes particularly relevant when cultural elements are initially placed in positions that hinder effective observation. The interaction task associated with this step can be implemented through a simple and intuitive action, such as pressing a UI button, allowing the cultural element to automatically move to a predefined accessible location.

A summary diagram illustrating the proposed inclusive immersion principles, modularity, predictability, and attention, and their application to the design of VR-based VMEs is provided in Figure 2.6.

Material Preparation

Defining which design principles should be adopted and how they can be utilized through interaction mechanics represents a core component of the proposed methodology. Additionally, equally important is the preparation of the materials required to build a VME tailored to users with ASD. This phase focuses on the digital reconstruction of a suitable museum environment and on the definition of a structured virtual visit. These aspects are considered essential to support both

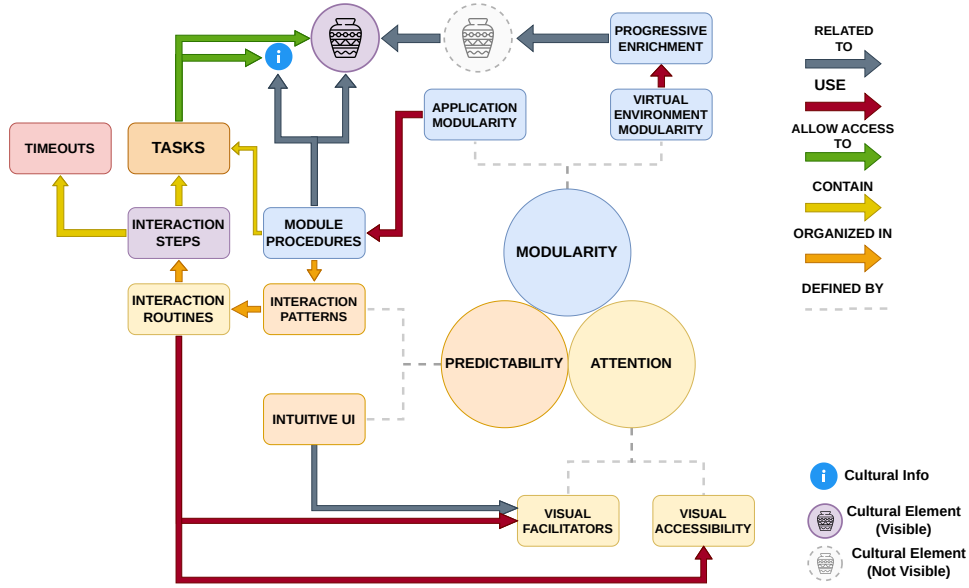


Figure 2.6: Overview of the three proposed inclusive immersion principles and their integration within the design of a VR-based VME.

physical and perceptual accessibility, with the aim of mitigating common issues experienced by individuals with ASD during museum visits, such as sensory overload.

Since the methodology targets VMEs inspired by real-world museums, it is crucial to ensure coherence between the virtual and physical experiences in terms of cultural contents, including artworks and historical information. To this end, within a generic museum environment (e.g., a museum hall), a set of POIs should be identified in collaboration with museum curators. In accordance with the modularity principle, these POIs represent the modules of the VME. Each POI corresponds to a specific spatial location where visitors can access cultural information and observe selected elements. Furthermore, particularly in the case of guided visits, the definition of a clear and well-structured pathway through the museum environment is strongly recommended. Pathway design plays a critical role, as unclear or ambiguous navigation may severely disrupt the experience of users with ASD, limiting their engagement with cultural contents.

Once the cultural elements and the pathway have been established, the workflow for creating the VE can be defined. This workflow aims to digitally reconstruct the museum environment in 3D, resulting in an architectural model of the space and a collection of 3D meshes representing the cultural elements associated with the POIs. This process can leverage widely adopted 3D reconstruction DTTs, such as laser scanning and photogrammetry, as well as mesh modeling and processing tools, including Maya, Blender, and Meshlab. For the 3D meshes representing

cultural elements, the management of the Level of Detail (LOD) is highly recommended. This practice is particularly important for optimizing performance in VR applications, ensuring smooth interaction and efficient rendering during the VME experience [190].

Notably, the material preparation phase does not strictly depend on the application design phase and can be carried out either before or after it. Two alternative approaches can be followed. In the first approach, material preparation is performed in advance, collecting all data related to the museum environment and the virtual visit. These data shape the subsequent design and development of the VME, ensuring the implementation of modularity, predictability, and attention principles utilizing the prepared materials.

In the second approach, the application design precedes the material preparation phase. In this case, the initial focus is on defining how the three principles will be implemented through specific interaction mechanics and design choices. Once the design framework is established, the material preparation phase follows, gathering the required data to develop the VME in alignment with the predefined design objectives.

Co-Design Refinement

As emerged from the analysis of the state of the art, co-design involving all relevant stakeholders is a key component of inclusive design processes. Accordingly, after the development of an initial application prototype, a series of co-design sessions should be conducted to evaluate and refine the system. These sessions may benefit from the involvement of museum curators, individuals with ASD, and representatives of ASD associations, as well as healthcare professionals such as psychiatrists, psychologists, and ASD specialists, in order to comprehensively address user needs and well-being.

During these sessions, each design iteration should be systematically evaluated by engaging ASD users for testing the VR application. Standardized assessment tools, such as the System Usability Scale (SUS) [28], can be employed to quantitatively measure usability and guide further refinements. After a sufficient number of iterative cycles, the resulting application can be considered mature enough for deployment within a real museum context.

2.1.4 Use Case

To validate the effectiveness of the previously explained methodology, whose overall structure is illustrated in Figure 2.7, this investigation presents a use case centered on the design and implementation of a VME. Specifically, the use case concerns a virtual visit tailored for individuals with ASD within the “Hall of Seasons” of Palazzo Carignano, a historical building and museum located in the city

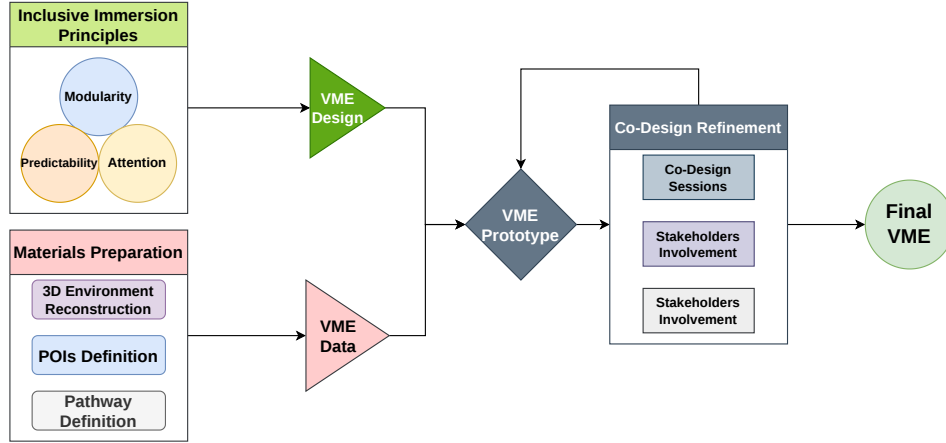


Figure 2.7: Scheme representing the key elements of the proposed methodology, namely the inclusive immersion principles, the material preparation, and the co-design refinement.

center of Turin, Italy (Figure 2.8). As previously mentioned, the selected scenario featured an environment characterized by multiple elements that may be potentially hindering for ASD users (e.g., golden decorations, bright lights); therefore, it was considered suitable for this investigation.

The resulting VR application, named *Palazzo Carignano VR*, was developed in accordance with the inclusive immersion principles introduced in this investigation and was conceived as an integrated component of the museum experience rather than as a standalone application. To assess the outcomes of the proposed approach, a user study involving participants with ASD was conducted, comparing the VR-based experience with a traditional in-person visit to the same exhibition space.

Palazzo Carignano VR

In line with the selected use case, the development process followed the stages defined by the proposed methodology. First, a material preparation phase was carried out through iterative interactions with the museum curators. Afterwards, the three inclusive immersion principles were applied to design and develop the VR application. Finally, the system was refined through a co-design phase conducted in collaboration with both museum professionals and ASD support associations (i.e., caregivers and psychiatrists), ensuring alignment with curatorial objectives and the needs of the target users.



Figure 2.8: Picture of the “Hall of Seasons” of Palazzo Carignano, Turin, Italy, representing the environment to reconstruct for the use case (©MiC – DRM Piemonte).

Material Preparation Phase

The material preparation phase was carried out in close collaboration with the museum staff and focused on the selection and digital reconstruction of an exhibition space suitable for evaluating the proposed methodology. The “Hall of Seasons” was chosen as the target environment, as it presents characteristics that may pose challenges for visitors with ASD, particularly due to the numerous and complex visual decorative elements.

To support the identification of culturally relevant contents, the curators provided documentation routinely used during on-site guided visits, including descriptive brochures and interpretative materials. Based on this information, a set of cultural elements was selected to structure the virtual visit. Specifically, seven elements were identified as particularly significant: the frescoed vault (Figure 2.9), two paintings portraying historical figures, and four golden sculptural decorations symbolizing the allegories of the seasons, namely summer, autumn, winter, and spring (Figure 2.10). Given their spatial proximity and their unified interpretative role within traditional narratives, the two paintings were treated as a single cultural entity. Consequently, six distinct POIs were defined along a predetermined pathway

inside the hall, each associated with one cultural element. For each POI, concise and accessible descriptions were defined with the museum staff to ensure consistency with the curatorial perspective. The geometric data required for the digital reconstruction were then acquired through on-site surveys. These surveys aimed at acquiring high-quality textures for several architectural elements (e.g., floor, ceiling) and at performing detailed 3D scans of the decorative sculptures representing the four seasons. Data collection was performed using an F6 Smart Handheld 3D scanner, producing point clouds that were imported in Echo software to generate meshes with millimetric accuracy. Mesh processing and export operations were performed using MeshLab version 2022.02, an open-source software widely adopted for point cloud and mesh generation. Each point cloud was processed three times to generate three distinct LODs by progressively reducing the number of vertices. Mesh reconstruction was performed using MeshLab’s “Screened Poisson Surface Reconstruction” algorithm, based on the method proposed by Kazhdan and Hoppe [110]. The resulting meshes were further refined in Blender to improve visual quality and usability within a real-time VR application, and exported as FBX files. In parallel, a simplified architectural model of the hall was created in Revit and integrated with the reconstructed meshes to provide a coherent spatial structure.

The resulting VE was designed to closely resemble the physical exhibition space while incorporating two minor simplifications. In particular, the chandelier and the central table present in the real hall were replaced with less detailed representations. This decision, made in agreement with the curators, was motivated by accessibility considerations, as these elements were not essential to the cultural narrative of the visit. Additionally, they were identified as potential sources of unnecessary and possibly hindering sensory stimulation for users with ASD.

Design and Development: Adopting the Principles

Once the material preparation phase was completed, the methodological focus moved to the design of interaction mechanics and to the technical development of the VR application. This phase aimed at applying the three inclusive immersion principles, modularity, predictability, and attention, within a coherent system tailored to users with ASD. The application was developed using Unity¹ 2021.3, integrating the SteamVR framework to provide immersive VR support. This choice ensured compatibility with a broad range of HMDs and enabled access to a consolidated set of interaction and tracking functionalities.

Regarding locomotion, teleportation was selected as the primary navigation technique, as it represents a widely adopted compromise between usability, comfort, and accessibility [25]. Although controller-based locomotion can introduce

¹Unity Engine: <https://unity.com/>



Figure 2.9: Image of the frescoed vault of the “Hall of Seasons” (©MiC – DRM Piemonte).

additional cognitive and motor demands for users with ASD, alternative solutions present notable issues: smooth locomotion is often associated with cybersickness and still requires controller use, while real walking necessitates large physical tracking spaces that are rarely available in standard museum setups [38]. To guarantee stable performance and high visual fidelity, the system was designed for PC-based VR platforms rather than standalone headsets.

To enhance perceptual accessibility, the user interface of the application was designed using symbols from AAC systems [41], which are commonly employed in applications targeting users with ASD. This design choice was intended to improve the clarity and immediate comprehensibility of UI elements such as buttons and panels, thereby supporting both usability and learning. In addition, a pre-recorded audio guide was integrated to provide verbal instructions and guidance throughout the experience, further supporting user orientation and task understanding.

From a structural perspective, the application was organized into three distinct scenes: a main menu, a tutorial, and a VME. Since the system was conceived as a supervised experience rather than a standalone application, the virtual environment is simultaneously rendered on the HMD worn by the user and on a desktop monitor used by a supporting operator. The desktop interface includes dedicated UI components, such as clickable buttons, not visible within the immersive view. These elements allow the operator to manage the session, monitor user progression, and intervene when necessary. Importantly, a persistent desktop UI button

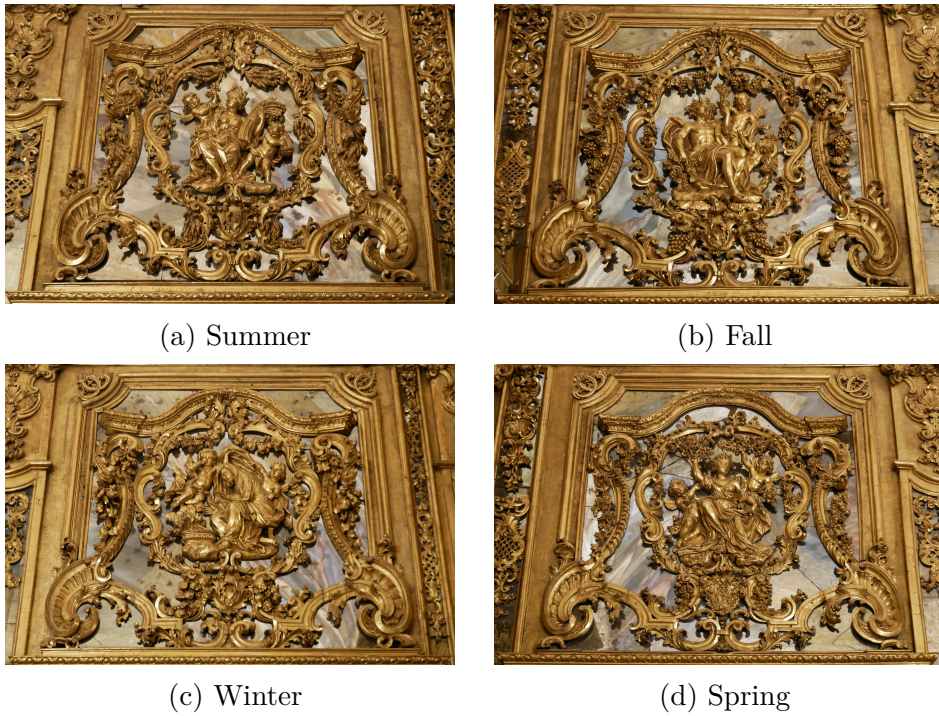


Figure 2.10: Golden decorations in the “Hall of Seasons” depicting the allegories in correspondence to the four seasons (©MiC – DRM Piemonte).

enables the immediate termination of the experience in case the user shows signs of discomfort or distress.



Figure 2.11: Tutorial scene. As observable on the floor, the unlocked (blue) and locked (orange) TPs enable the user’s movement within the VE.



Figure 2.12: The image depicts the digital VE, i.e., the “Hall of Seasons”, as visualized in the VME scene of the VR application. On the floor, the pathway and the TPs defining the POIs of the VME are visible.

The main menu, accessible exclusively from the desktop interface, allows the operator to configure several parameters prior to starting the experience. Among these, the user’s height can be either manually specified or automatically retrieved from HMD tracking data via a dedicated control. This information is used to dynamically adapt the positioning of virtual UI elements to accommodate users with different statures. In accordance with the modularity principle, each POI corresponds to an independent module; therefore, additional settings are provided to manage several parameters, such as the timeout associated with each interaction step and the activation or deactivation of particle effects used to reinforce visual feedback. Once the configuration is completed, the operator can initiate the VR session.

At the start of the session, the user is placed within a tutorial scene (Figure 2.11), designed to facilitate a gradual familiarization with the VE. The tutorial comprises multiple POIs, each equipped with a Teleport Point (TP) to support movement within the scene. At each POI, the user is introduced to the fundamental interaction mechanisms of the application, including teleportation and interaction with virtual UI elements, with guidance provided through the audio narration. The tutorial also presents contextual information about the museum hall featured in the experience, such as its location within the building and a brief historical overview. Upon completion of all tutorial POIs, the user is smoothly transitioned to the VME scene to begin the main virtual visit.

The VME scene represents the inclusive virtual visit itself (Figure 2.12) and is organized around the six POIs previously identified during the material preparation

phase. These POIs define a structured visit path that becomes incrementally visible through directional UI arrows displayed on the floor after each completed POI. Furthermore, a module procedure consistent with the proposed inclusive immersion principles is embedded within each POI. In the current implementation, each interaction step of a procedure is uniquely associated with a single POI, thereby reinforcing modularity at both the spatial and interaction levels. These procedures are hereafter referred to as *POI Procedures*, and their structure is detailed in the following subsection.

POI Procedure

At the beginning of the virtual visit, the user is positioned at the first POI, where an introductory narration about the hall is delivered through the audio guide. Following this initial contextualization, the POI Procedure associated with the selected module is activated. This procedure governs the sequence of interaction routines and interaction steps that the user must complete in order to progress within the experience. As discussed in Section 2.1.3, each interaction step is formally characterized by a specific task and an associated timeout, while task instructions are conveyed verbally via the audio guide.

To clarify how a POI Procedure is concretely implemented within the VR application, a representative example is outlined below.

- **Interaction routine 1 – Cultural element appearance**
 - **Interaction step 1 – POI activation:** a UI button featuring an AAC-based icon is presented to the user and functions as an initial visual facilitator. At this stage, the button is positioned in front of the user and dynamically adjusts its placement based on head orientation and user position, maintaining a consistent distance and orientation. The audio guide instructs the user to activate the button by reaching out and pressing it using the controller.
 - **Interaction steps 1/2/3 – Head-gaze guidance:** once the activation task is completed, the behavior of the UI button transitions to a head-gaze-based interaction paradigm, consistent with the attention principle described in Section 2.1.3. The button is augmented with a particle system to reinforce its role as a visual facilitator. The user is required to visually track the movement of the button within the VE across a predefined number of interaction steps (three in the present implementation). Each successful gaze detection triggers visual feedback, i.e., a temporary color change to green, and causes the button to relocate to a new position in the environment. Upon completion of the final interaction step, the button disappears and the cultural element associated with

the POI becomes visible. In the case of the four seasonal decorations, an additional particle effect, thematically customized for each season, is employed to further draw attention to the revealed element.

- **Interaction routine 2 – Cultural element learning**
 - **Interaction step 1 – Observation trigger:** the audio guide introduces the cultural element and explains the subsequent task. A second UI button then appears, allowing the user to initiate an automatic repositioning of the cultural element to a predefined location within the VE that offers improved visual accessibility. Once the button is pressed, it disappears, and the cultural element moves accordingly. This step corresponds to the visually accessible positioning strategy discussed in Section 2.1.3.
 - **Interaction step 2 – Return control:** while the audio narration provides detailed information about the cultural element, the user can observe it from a close and visually optimal perspective. A final UI button is then displayed, enabling the user to return the cultural element to its original position within the VE.
- **Interaction routine 3 – Transition to the next POI**
 - **Interaction step 1 – Teleportation:** after the learning phase is completed, the TP associated with the subsequent POI is unlocked. The audio guide encourages the user to use teleportation to reach the next location, also providing spatial cues (i.e., position of the TP on the floor) to facilitate navigation. Once the teleportation is performed, the current POI Procedure is considered concluded.

Table 2.1 summarizes the general structure of a POI Procedure. Within the developed VR application, this procedural structure is fully adopted for the POIs associated with the four seasonal decorations, which are positioned at elevated locations and therefore present greater observational challenges. Consequently, all three interaction routines are employed for these POIs. Unlike the latter ones, the remaining two POIs rely solely on the first and third interaction routines. Timeout values associated with interaction steps can be configured through the main menu scene, allowing the operator to adapt the experience to the specific learning pace and needs of individual users.

Figure 2.13 illustrates an example of the first interaction routine of a POI, shown from the perspective of a user immersed in the VR environment.

Interaction Routine	Interaction step	Task	Timeout (s)
1 - Cultural element appearance	POI activation	Press UI button	30
	Head-gaze guidance 1	Follow UI button	15
	Head-gaze guidance 2	Follow UI button	15
	Head-gaze guidance 3	Follow UI button	15
2 - Cultural element learning	Observation trigger	Press UI button	45
	Return control	Press UI button	60
3 - Transition to the next POI	Teleportation	Press, point and release	30

Table 2.1: Structural representation a POI procedure; beyond the interaction routines, interaction steps and tasks, it is possible to observe in the last column the default values, in seconds, for the timeouts.

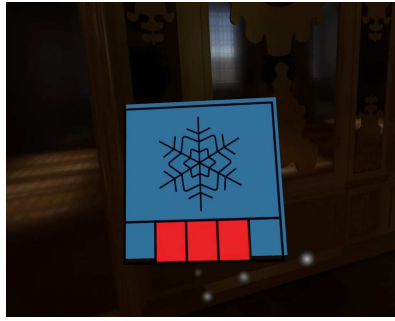
Co-Design Refinement with Museum Curators and ASD Associations

In accordance with the proposed methodology, once a functional prototype of the VR application was developed, a structured co-design process was carried out to assess and iteratively refine the system. The refinement process was organized into two distinct rounds of sessions, each involving different categories of stakeholders and targeting complementary evaluation objectives.

The first round involved domain experts, namely a psychologist, a psychiatrist specialized in ASD, and three museum curators from Palazzo Carignano. Their contribution focused on evaluating the prototype with respect to accessibility requirements, cognitive load, and the accuracy and appropriateness of the cultural contents presented within the VE. Specifically, the psychologist and the psychiatrist assessed the usability of the prototype based on their expertise and experience, ensuring that no discomfort would arise for ASD users engaging with the VR experience. In contrast, the museum curators were responsible for verifying the correctness of the content, including the audio guide narration and the visual representation of the elements within the museum hall.

The second round of co-design sessions targeted end users and included 17 adults with ASD (14 males and 3 females), aged between 18 and 60 years ($M = 24.17$, $SD = 10.85$), characterized by heterogeneous backgrounds and different levels of familiarity with VR technologies. Participants were accompanied, when required, by caregivers, tutors, or parents.

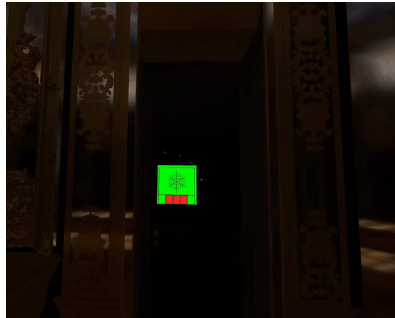
All sessions were conducted in a controlled laboratory environment, consisting of a medium-sized room arranged for immersive VR experimentation. During each session, participants experienced both the tutorial scene and the VME scene of the application. The immersive setup relied on an HTC Vive Pro 2 HMD connected to a desktop PC running the application through SteamVR. Each experimental session lasted approximately 35-40 minutes and followed a structured protocol comprising: an initial welcome and procedural briefing by the research team (approximately 10 minutes); a hands-on interaction with the VR prototype (approximately 15 minutes); and completion of the SUS questionnaire, complemented by open-ended



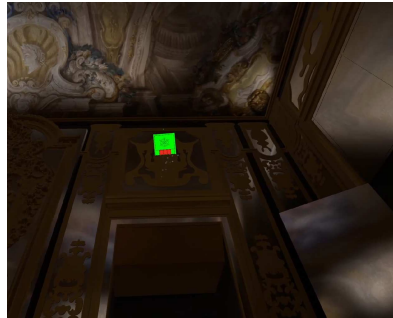
(a) IR1-IS1



(b) IR1-IS2



(c) IR1-IS3



(d) IR1-IS4



(e) Fall season appearance

Figure 2.13: The images represent the four interaction steps (1/2/3/4) in the first interaction routine of a POI. Specifically, the example corresponds to the cultural element linked to the winter season decoration.

questions aimed at collecting qualitative feedback.

Participant safety and comfort were treated as primary concerns throughout the experimental process. Users were explicitly informed that they could interrupt

the experience at any time and were encouraged to report any discomfort. A human operator continuously supervised the session from the desktop interface and retained the ability to immediately terminate the VR experience if necessary.

Feedback gathered during the first round of co-design sessions led to several refinements of the application. In particular, unnecessary textual UI elements were removed (i.e., descriptions of the cultural objects), as the psychiatrist identified them as potentially distracting and unsuitable for certain ASD users with reading-related perceptual difficulties. Additionally, suggestions provided by the museum curators resulted in revisions of some cultural descriptions narrated by the audio guide to ensure consistency with the real-world museum narratives. Specifically, the descriptions regarding the ceiling and one of the paintings in the hall, which were initially based on publicly available online data, were updated based on the feedback of the museum curators.

Results from the second round of sessions, involving ASD users, showed that participants were able to complete the entire virtual visit and reach all POIs. However, the usability evaluation resulted in a mean SUS score of 66.62 ($SD = 22.72$), slightly below the acceptability threshold of 70 defined by the adjective ratings proposed by Bangor et al. [14]. Observations highlighted specific usability issues: participants with concomitant upper-limb motor impairments encountered difficulties in interacting with UI buttons using controllers, while other users reported discomfort caused by the sudden appearance of UI elements at close visual distances.

To address these issues, the UI appearance mechanism was redesigned. In the updated version, UI buttons initially appear at a greater distance from the user and then smoothly transition to an ergonomically reachable position after a short interval. Furthermore, additional configuration options were introduced in the main menu to allow operators to adjust UI distances based on individual motor capabilities. Moreover, one participant presenting short-term memory impairments experienced difficulties in retaining the cultural information conveyed during the experience. To support this aspect of accessibility, a recap UI panel was added at the end of the virtual visit, displaying a summary video of the cultural contents explored.

As a result of the iterative co-design process and the subsequent refinements, the *Palazzo Carignano VR* application was validated by museum curators and progressively adapted based on feedback from ASD users and their caregivers, and was therefore deemed suitable for the experimental phase. In particular, a final co-design session involving 10 more ASD users (7 males and 3 females) resulted in a SUS score corresponding to a “Good” level of usability ($M = 76.47$, $SD = 17.22$), according to the classification proposed by Bangor et al. [14]. Despite these improvements, a limited number of usability issues still emerged during the main experiment, as discussed later in relation to the results. These findings further underline the importance of carefully refining attention-guidance mechanisms and

spatial UI layouts when designing VR-based VMEs for users with ASD.

Experimental Procedure

To assess the effectiveness of the VR application developed according to the proposed methodology, a between-subjects user study was designed and conducted. The experimental protocol involved a total of 34 participants diagnosed with ASD (28 males and 6 females), with ages ranging from 19 to 67 years ($M = 27.24$, $SD = 9.00$). Participants presented heterogeneous levels of familiarity with immersive VR technologies, ranging from no prior experience to regular use. All the participants were selected to ensure no overlap with the co-design groups.

The study was conducted in compliance with the “Carignano VR” experimental protocol (ID: 101072/2022), which received approval from the Ethics Committee of the Politecnico di Torino, Italy. All experimental activities were carried out under the supervision of the research team and in accordance with ethical guidelines for research involving human participants.

Participants were welcomed in small groups of five to six individuals by the research staff. Each group was accompanied by caregivers and/or representatives from ASD associations. During this introductory phase, the objectives of the study and the overall experimental procedure were explained in accessible terms, after which written informed consent was obtained. Following this initial briefing, participants completed the experimental experience individually.

Prior to the VR experience, each participant was asked to complete a pre-experience questionnaire aimed at collecting demographic and background information. The questionnaire included items related to age, gender, education level, occupation, self-identification as being on the autism spectrum, self-reported or clinically diagnosed ASD level (when available), and frequency of VR usage. Additionally, the abridged version of the Autism-Spectrum Quotient (AQ-Short) [97] was administered to provide a standardized estimate of autistic traits.

Age information was subsequently used to categorize participants into four age groups: Young Adults (18–25), Adults (26–40), Middle-aged Adults (41–60), and Seniors (61+). Given that self-assessed or clinically diagnosed ASD levels were unavailable for more than 50% of the sample, AQ-Short scores were adopted as a proxy measure to characterize participants. Based on the thresholds validated through ROC analysis by Hoekstra et al. [97], scores below 65 were classified as low, scores between 65 and 69 as moderate, and scores equal to or above 70 as high. These categories are hereafter referred to as *ASD trait levels*.

An overview of the demographic characteristics of the participant sample, including age distribution and ASD trait levels, is reported through percentage-based charts in Figure 2.14 and Figure 2.15.

The user study was guided by two main research questions:

- **RQ1:** Is a VME developed according to the methodology proposed in this

investigation, and explicitly tailored to the needs of individuals with ASD, capable of delivering a more educational, entertaining, visually appealing and comfortable museum experience than a traditional in-person visit featuring the same cultural contents?

- **RQ2:** Do participants' age and level of autistic traits significantly affect the perceived quality of the VME and its potential benefits in terms of education, entertainment, esthetics and museum experience when compared to a traditional museum visit?

Given that the primary goal of the study was to compare two different visit modalities, namely, a traditional museum visit and a virtual one, participants were equally divided into two groups, each composed of 17 individuals. The first group, hereafter referred to as the Control Group (CG), took part in a conventional guided visit inside the "Hall of Seasons" of Palazzo Carignano, led by a professional museum guide from the institution's staff. The second group, referred to as the VR Group (VG), experienced the same visit through the proposed VR application led by the virtual audio guide.

The experimental sessions for the VG were conducted using an HTC Vive Pro 2 HMD, tethered to a desktop PC running the VR application. Throughout each session, an experimenter supervised the experience via the PC monitor, which allowed both the configuration of application parameters (e.g., registering the participant's height) and the immediate interruption of the experience in case of discomfort. Considering the involvement of individuals with ASD, caregivers were allowed to be present during the sessions for participants in both groups.

The traditional visit was designed in close collaboration with the museum curators to ensure full correspondence with the contents presented in the VME. This alignment in terms of cultural elements, visit structure, and overall experience was considered essential to enable a fair comparison between the two conditions. The CG sessions were conducted with small groups of approximately two to three ASD visitors, accompanied by a museum guide. After an initial introduction to the museum, participants followed the same predefined pathway adopted in the VR experience, reaching the same locations identified as POIs in the VME. At each POI, visitors were presented with the same cultural contents established during the material preparation phase and implemented in the VR application. Explanations were provided verbally by the museum guide, who accompanied participants throughout the visit.

To further ensure consistency between the two modalities, the AAC icons used in the VR application to represent cultural elements were printed on A4 sheets and employed as supporting materials by the guide during the traditional visit. Each experimental session lasted approximately 45 minutes and was structured as follows:

1. Welcome and introduction to the experience (approximately 10 minutes);
2. Administration of a demographic questionnaire, signing of the informed consent form, and delivery of an information sheet describing the study (approximately 10 minutes);
3. Participation in the museum visit, either traditional (CG) or virtual (VG), with comparable durations (approximately 15 minutes);
4. Administration of a post-experience questionnaire.

The post-experience questionnaire was designed to assess the overall quality of the museum experience for participants in both groups. Due to the absence of standardized and validated instruments specifically targeting UX in immersive museum contexts, which is a limitation also highlighted by recent literature reviews [15], an existing questionnaire previously adopted in similar research scenarios was selected to ensure methodological coherence and comparability. In particular, the instrument proposed by Lee et al. [127] was adapted to fit the scope of the present study.

This questionnaire builds upon established evaluation tools and is grounded in the 4E framework (Education, Entertainment, Esthetics, Escapism) derived from the Experience Economy model by Pine and Gilmore [176]. Differently from other applications of this framework, the instrument by Lee et al. was explicitly designed to evaluate VMEs conceived as digital counterparts of existing physical exhibitions [127]. Owing to its structure, it has also been shown to be suitable for adaptation to real-world museum contexts [198], such as traditional guided visits.

Within the current study, the sections related to *Education* (defined as the absorption of unfolding events at a CH site through active and interactive participation [163]), *Entertainment*, *Esthetics*, *Overall museum experience*, and *Offline museum visit intention* were retained. Conversely, the *Escapism* dimension, since it was specific to immersive VR experiences, was excluded. Moreover, the original *Offline museum visit intention* construct, which assessed the extent to which a VX encourage a real-world visit, was adapted to measure participants' intention to explore the entire museum (Palazzo Carignano) associated with the exhibition area of the "Hall of Seasons." Accordingly, this dimension was renamed *Entire museum visit intention*. Hereafter, the adapted instrument is referred to as the *Museum Experience Evaluation Questionnaire (MEEQ)*.

Finally, an open-ended feedback section was included at the end of the questionnaire to allow participants to freely report additional comments or observations regarding their experience.

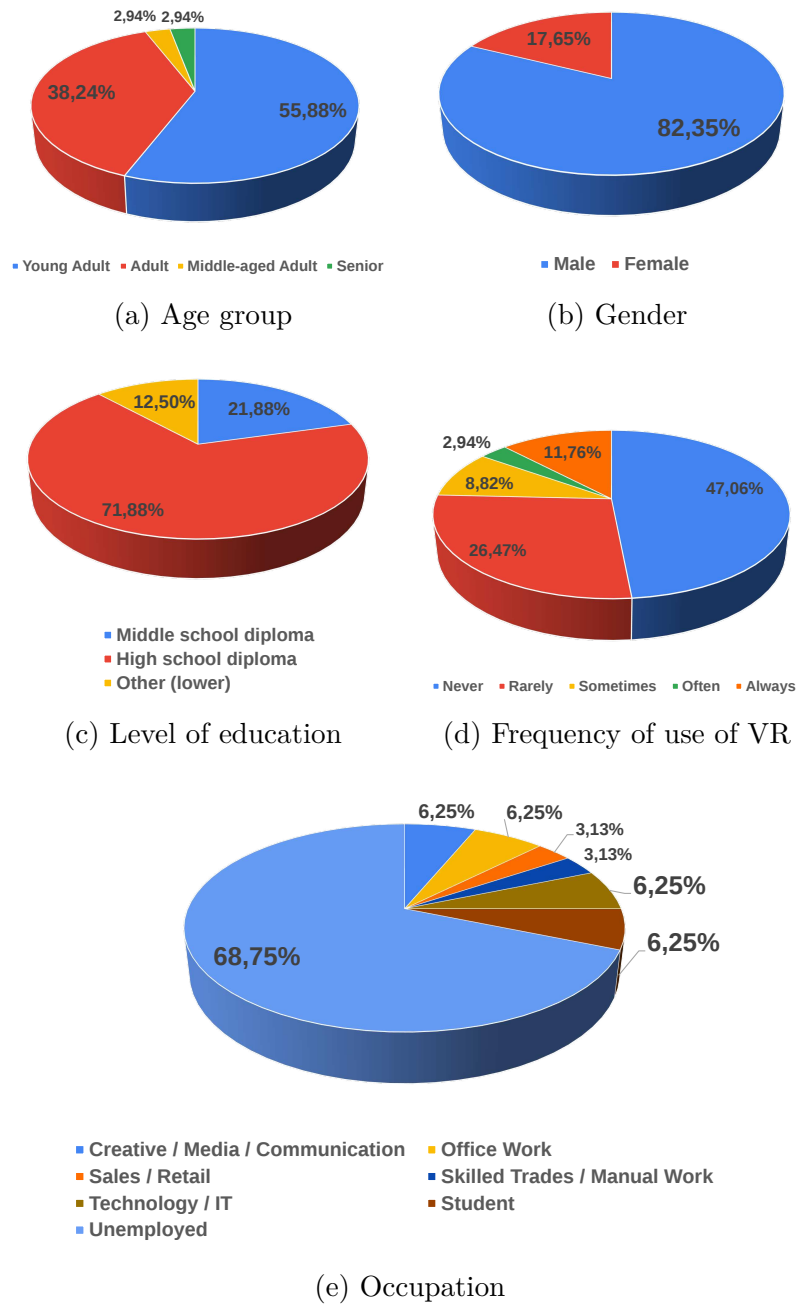


Figure 2.14: Demographic representation of the participants (part 1): age group (a), level of education (b), gender (c), frequency of use of VR technology (d), and occupation (e).

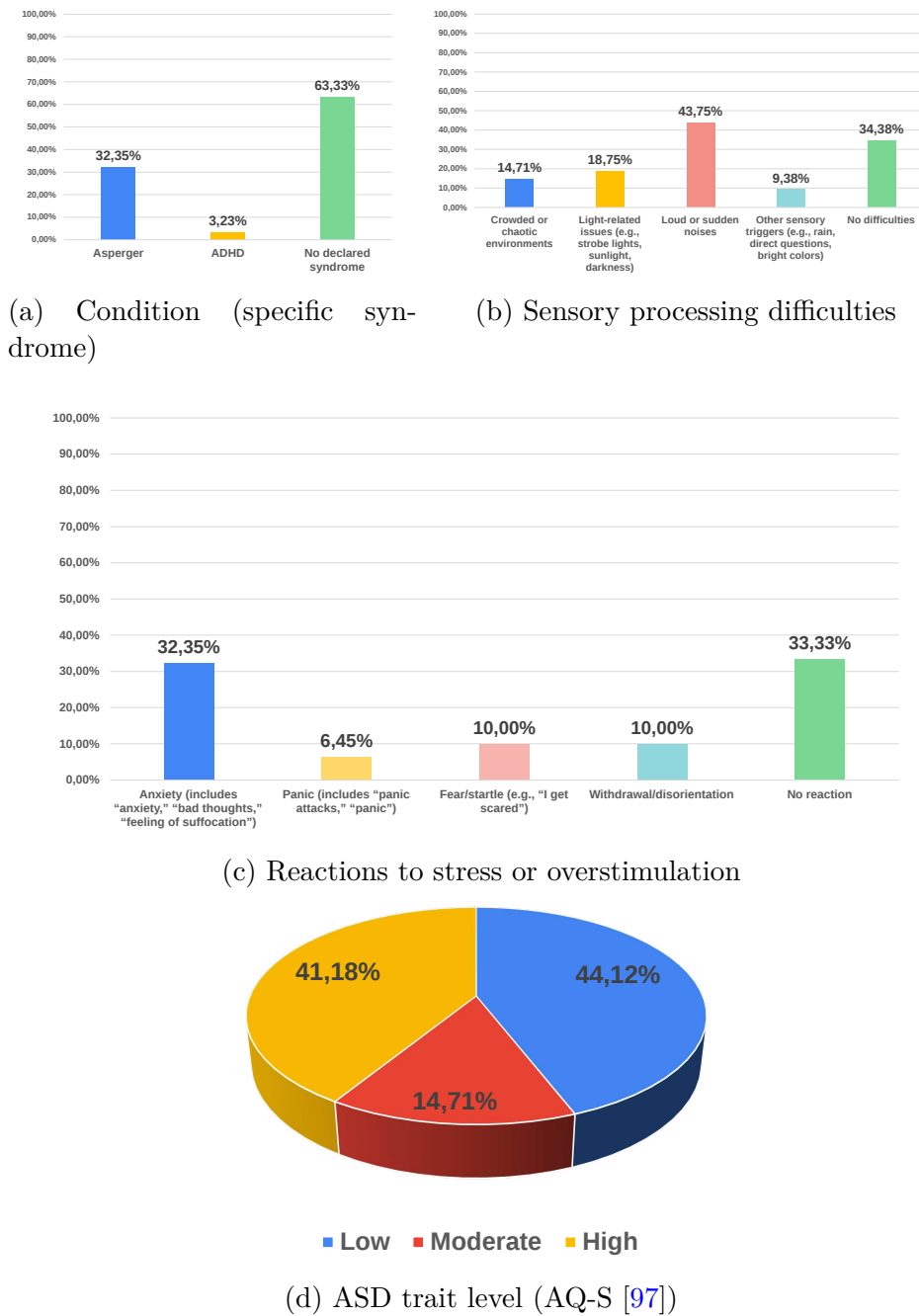


Figure 2.15: Percentage breakdown of participant demographics (part 2): specific syndrome (a), sensory processing difficulties (b), reactions to stress or overstimulation (c), and ASD trait level (AQ-Short [97]) (d).

2.1.5 Results and Discussion

The data analysis followed a three-step procedure. First, an overall comparison between the VG and the CG was performed by considering the entire participant

sample. Subsequently, to investigate the presence of potential moderating elements, participants were stratified according to age and ASD trait levels, and the comparison between the VR and traditional conditions was repeated within each resulting subgroup. These two variables were selected for stratified analysis not only due to their relevance from a demographic and clinical perspective, but also because they were the only factors that allowed the formation of sufficiently balanced and comparable groups, as shown in Figure 2.14 and Figure 2.15. Other variables, such as gender or self-reported ASD severity classifications, did not provide an adequate distribution of participants across categories, thus preventing statistically meaningful comparisons.

To ensure group comparability, it was necessary to merge certain categories. This approach was adopted both to remove categories characterized by very small sample sizes and to avoid three-group comparisons, which would have required the use of ANOVA procedures and subsequent post hoc adjustments. In all cases, category merging was also supported by conceptual considerations. With respect to age, for instance, the “middle-aged adult” and “senior” categories (each accounting for only 2.94% of the sample) were combined with the “adult” category (38.24%). This resulted in a binary division between participants aged 18–25 and those older than 25 years, with balanced group proportions (44.12% vs 55.88%). A similar strategy was applied to ASD trait levels, where the moderate (14.71%) and high (41.17%) categories were merged into a single “high” trait level group. This provided a well-balanced split in terms of both sample size (53.88% vs 44.12%) and score ranges between the “low” (AQ-Short scores 28–64) and “high” (AQ-Short scores 65–112) ASD trait levels [97].

Data normality was assessed using the Shapiro–Wilk test. As the distributions deviated from normality, non-parametric statistical analyses were employed. Specifically, the Mann-Whitney U test was used for all comparisons, adopting a significance level of 5% ($p < .050$).

VR vs Control: Overall Comparison

The first set of analyses focused on the overall comparison between the VG and CG across the full participant sample. The results obtained using the MEEQ [127], summarized in Figure 2.16 and Figure 2.17, indicate that the museum experience delivered through the VR application significantly outperformed the traditional visit in several dimensions, namely Education, Entertainment, Esthetics, and Overall museum experience.

Examining the individual items within each questionnaire section, starting from Education, defined as the active and interactive absorption of unfolding events during the visit (Figure 2.17a), it emerges that participants in the VG reported higher levels of perceived learning compared to those in the CG. In particular, the VR

visit was more effective in fostering the acquisition of new knowledge and in stimulating curiosity toward further learning. These findings suggest that the application of the proposed design principles, especially the attention-related mechanisms, contributed to a more effective transmission of cultural content than traditional approaches. Nevertheless, the absence of statistically significant differences in the items assessing whether participants felt more knowledgeable overall or considered the visit a good learning experience indicates that both visit modalities were reasonably aligned in achieving these outcomes.

Turning to the Entertainment dimension (Figure 2.17b), the VR-based visit was perceived as more enjoyable and fun by participants in the VG compared to those in the CG. Although no statistically significant differences were observed for captivation and overall entertainment, the associated p -values approached the significance threshold ($p = .067$ and $p = .057$), suggesting that these effects might reach significance with a larger sample size. These results further support the suitability of VR for delivering experiences in contexts where the real-world counterpart may lack an inherent entertainment component. This advantage can be attributed to the possibility of integrating gamified elements and enabling interactions that are not feasible in physical environments. Importantly, the application of the predictability principle, resulting in more repetitive interaction mechanics compared to conventional VR applications, did not appear to negatively affect entertainment. Likewise, the attention principle, which often requires longer interaction times than traditional paradigms (e.g., maintaining gaze on a UI element until it reaches a target location), did not reduce perceived enjoyment. This aspect is particularly relevant, as such mechanisms could have been perceived as tedious by participants used to the variability typical of commercial VR games and interactive applications.

With respect to Esthetics (Figure 2.17c), participants in the VG rated the visit as more visually appealing and pleasant than those in the CG, confirming the effectiveness of the VR experience from a visual design standpoint. However, no significant differences were observed in terms of perceived attention to detail, indicating that both the virtual and traditional visits were considered adequate in this regard.

Considering the Overall museum experience dimension (Figure 2.17d), the results show that the virtual visit contributed more positively to participants' overall experience at Palazzo Carignano compared to the contribution of the traditional visit for the CG. Additionally, the VR condition resulted in a more enjoyable perception of the museum as a whole. It is important to note that both the virtual and traditional visits took place within the museum setting. This suggests that the VR application, developed according to the proposed methodology, was able to enhance the museum experience in a way that better accommodated the specific needs of the target audience, namely individuals with ASD traits. At the same time, no significant differences were found between the two conditions regarding the perceived meaningfulness of the museum experience, indicating that the traditional visit was

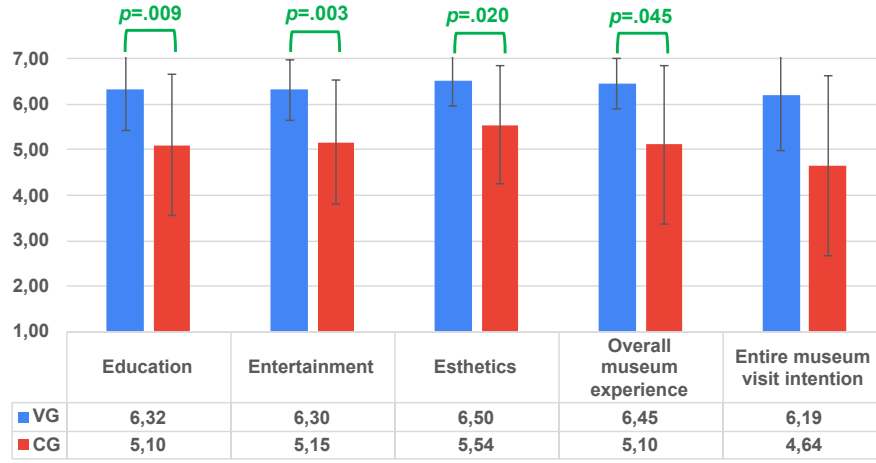


Figure 2.16: Overall comparison for the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars.

also regarded as highly meaningful.

Finally, for the Entire museum visit intention section, as well as for the aggregated score, no statistically significant differences emerged across any of the individual items. This finding indicates that replacing a traditional visit with a VME did not negatively affect participants' intention to visit other areas of the museum in the future. From the perspective of museum curators, this result is particularly relevant, as it suggests that the adoption of VR-based solutions is unlikely to discourage this audience from engaging with the physical museum. During questionnaire administration, it was explicitly clarified that the expression “the rest of the museum” referred exclusively to areas accessible through traditional visits, and not to additional VR experiences.

Overall, these findings provide strong support for **RQ1**, which investigated whether a VR-based museum experience developed following the proposed methodology could improve the perceived quality of the visit for individuals with ASD traits when compared to a traditional guided visit. The significant advantages observed across multiple dimensions, namely Education, Entertainment, Esthetics, and Overall museum experience, demonstrate the effectiveness of the VR application in engaging users holistically. Moreover, the absence of detrimental effects on participants' intention to explore the rest of the museum further highlights the potential of VR technologies to complement, rather than substitute, conventional museum offerings.

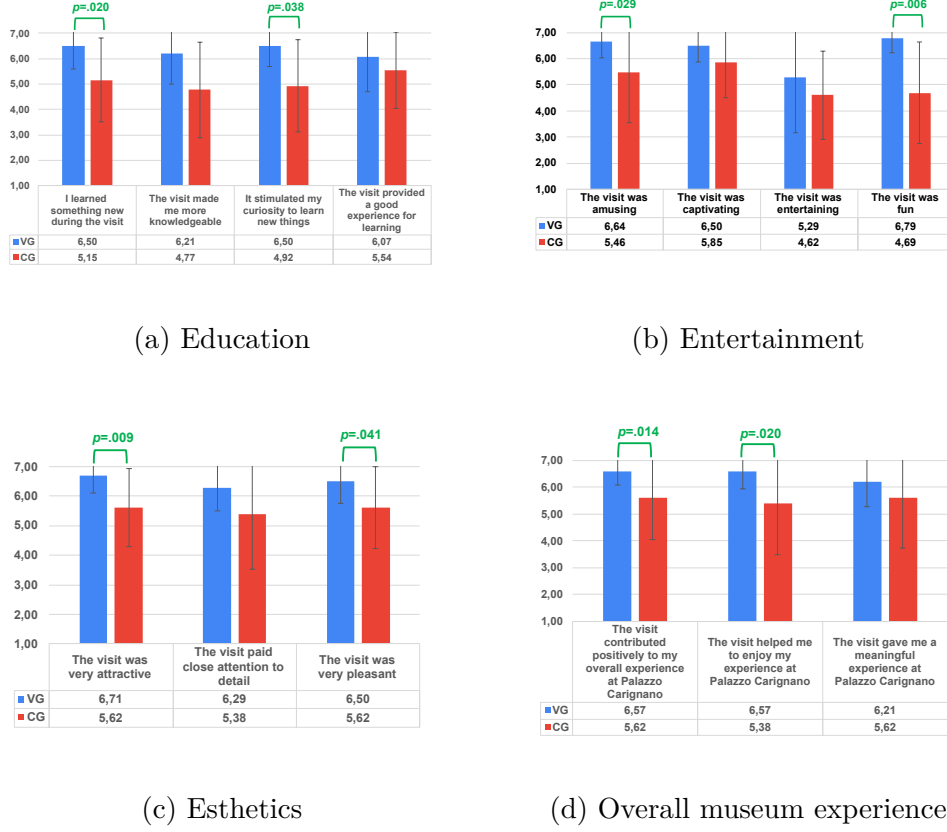


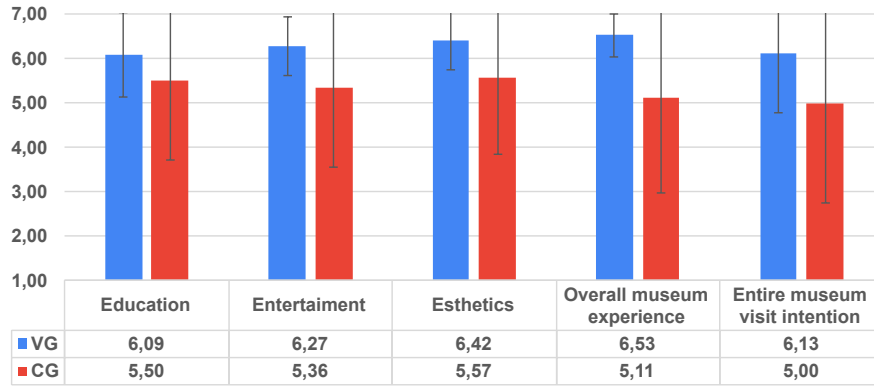
Figure 2.17: Overall comparison regarding for the sections of the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars.

VR vs Control: Age-Based Analysis

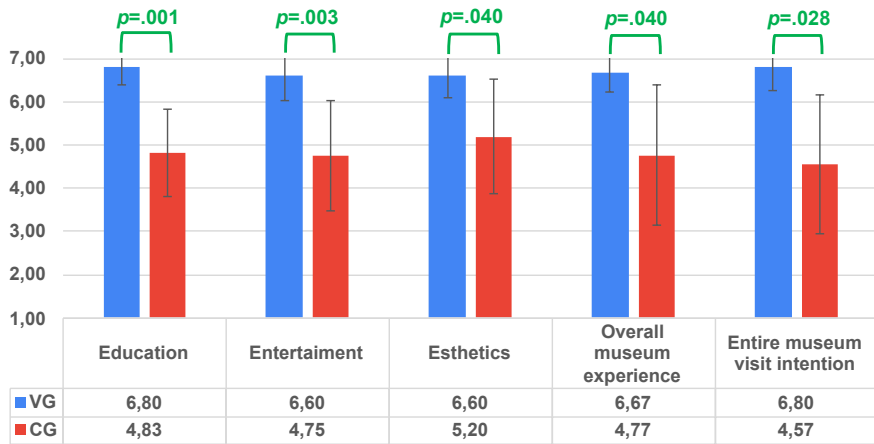
A second set of analyses was conducted by stratifying participants according to age, with the aim of investigating whether the impact of the VR museum experience differed across age groups. As described earlier, due to the limited number of participants in some categories, age groups were merged, resulting in two balanced cohorts: participants aged 18–25 and participants older than 25 years. The comparison between the VG and the CG was then performed separately within each age group.

The MEEQ results for the age-based analysis, reported in Figure 2.18, reveal a marked divergence in outcomes between the two age groups.

Contrary to expectations, within the younger group, where greater openness toward immersive technologies, lower technological barriers, and a potentially weaker appreciation for traditional museum visits might reasonably be assumed, the significant differences observed in the overall analysis were no longer present. In



(a) Young-Adults (18-25)



(b) Adults (>=26)

Figure 2.18: Overall analysis in terms of age for the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars.

contrast, among adult participants, who could be expected to exhibit higher levels of skepticism toward novel technologies and a stronger preference for conventional cultural experiences, not only were the overall significant differences preserved, but they were in some cases more pronounced. Notably, this included a statistically significant difference in the *Entire museum visit intention* indicator, as illustrated in Figure 2.18.

A more detailed examination of the adult subgroup highlights that participants who experienced the VR visit reported significantly higher levels of perceived knowledge acquisition compared to those who took part in the traditional visit (6.80 vs 5.10, $p = .013$), and judged the experience as more effective from a learning perspective (6.80 vs 4.70, $p = .028$). Additionally, the VR experience was rated as more captivating (7.00 vs 5.50, $p = .013$), and participants expressed a stronger intention to visit the museum in its entirety when given the opportunity (6.60 vs 4.40, $p = .040$). At the same time, several differences identified in the overall comparison were not replicated within this age group. These include items related to curiosity stimulation, perceived fun, visual attractiveness, and overall meaningfulness of the experience; however, it is worth noting that the corresponding p -values for these measures were only marginally above the conventional significance threshold of .050.

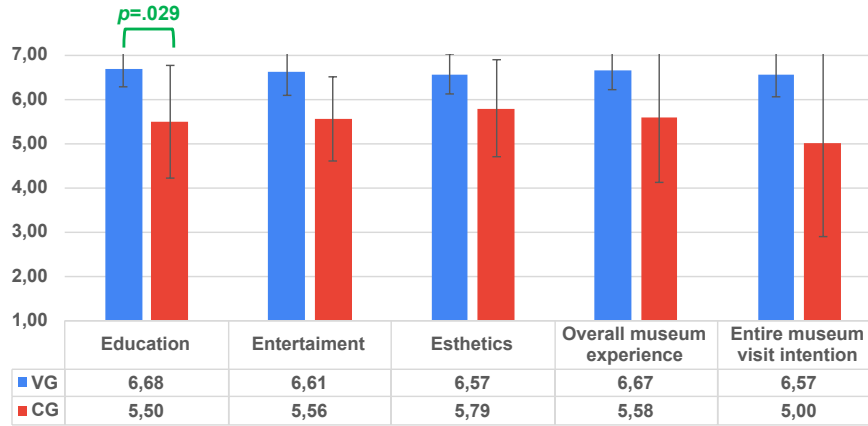
These findings can be further interpreted in light of both the demographic information collected and the qualitative feedback provided by participants. In particular, a relationship emerged between younger age and greater familiarity with VR technologies, with several younger participants reporting frequent prior use. This observation suggests two complementary explanations. First, adult participants, being less accustomed to immersive VR, may have been more susceptible to a novelty effect associated with the technology. Second, younger participants, already exposed to more sophisticated and varied interactive experiences that are not specifically tailored to the needs of individuals with ASD, may have perceived the VME, designed according to the proposed methodology, as overly constrained or simplified. This perception may have reduced engagement and, consequently, attenuated the difference between the virtual and traditional visit conditions within this age group.

VR vs Control: ASD Trait Level Analysis

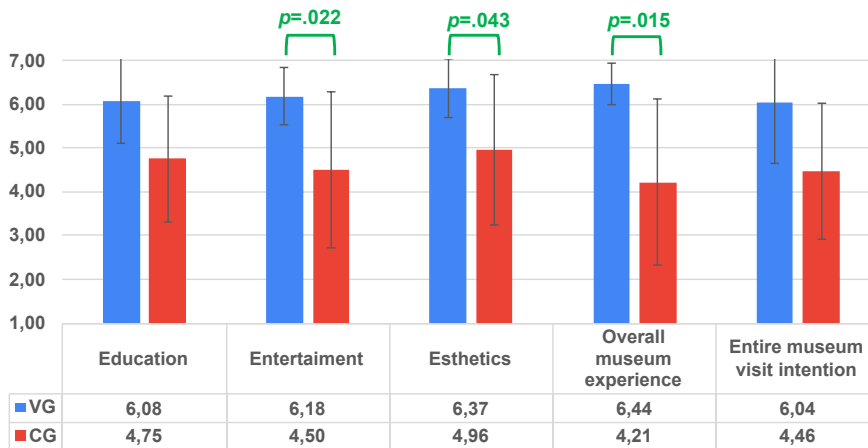
The final set of analyses examined whether the level of ASD traits moderated the effects of the VR-based museum experience. Participants were classified into two groups, namely *low* and *high* ASD trait levels, according to their AQ-Short scores, as previously described. Within each trait level category, comparisons were then carried out between the VG and the CG.

The MEEQ results for this analysis, illustrated in Figure 2.19, reveal a pattern of outcome differentiation similar to that observed in the age-based analysis (although with a lower overall magnitude).

In the overall comparison, conducted without stratification, statistically significant differences between the VG and CG were observed across all major indicators, with the exception of the *Entire museum visit intention*. However, once the sample was divided according to ASD trait level, more pronounced trends emerged. In the low-trait subgroup, nearly all previously significant differences vanished, with the sole exception of the *Education* indicator, which remained statistically significant.



(a) Low (28-64)



(b) High (65-112)

Figure 2.19: Overall analysis in terms of ASD trait level (AQ-Short [97] score) for the MEEQ [127] (the higher the better). The baffles were inserted to represent significant differences ($p < .050$). Standard deviation is reported with error bars.

Conversely, in the high-trait subgroup, most of the significant effects identified in the overall analysis were preserved. An exception was again represented by the *Education* section, for which neither the individual items nor the aggregated score reached statistical significance. This outcome may be interpreted in relation to the broader learning difficulties often associated with higher ASD trait severity.

In addition, differences related to the perceived amusement and visual attractiveness of the visit were no longer significant within this subgroup, suggesting that these dimensions may have been partially driven by a technology-related bias in the aggregated analysis.

Notably, participants with higher ASD trait levels who experienced the VR visit reported a significantly stronger intention, given the opportunity, to visit Palazzo Carignano in its entirety compared to their counterparts in the CG (6.00 vs 4.13, $p = .036$), even though the aggregated indicator for this section remained marginally above the conventional significance threshold ($p = .059$).

Overall, these findings provide partial support for **RQ2**. While the VR-based experience appeared to yield a stronger positive effect for participants with higher ASD trait levels, its impact was considerably attenuated among those with lower trait levels, for whom most of the significant differences observed in the overall analysis were no longer present. This pattern suggests that the proposed methodology may be particularly effective in supporting users who encounter greater cognitive or perceptual barriers during traditional museum visits, while offering more limited benefits for individuals with lower ASD traits. Rather than producing uniform advantages across the entire autism spectrum, the results indicate that ASD trait severity acts as a moderating factor in shaping the effectiveness of the VME, underscoring the importance of further refinement and personalization to better accommodate the diversity of ASD users.

Limitations and Future Developments

Despite the encouraging outcomes, several limitations of the present study must be acknowledged. During the experimental phase, a subset of ASD participants with particularly severe levels of disability had to be excluded for different reasons. Some participants reported discomfort when wearing the HMD and were therefore unable to experience the VR application, while others, despite completing the VME successfully, were not able to provide reliable responses when completing the questionnaires. When comparing these issues with previous works in the literature, similar challenges related to immersive technologies emerge, including discomfort associated with HMD usage [169] and excessive levels of immersion [35]. These observations suggest that future VR studies targeting ASD users should more carefully account for differences in disability severity. One possible direction would be to adopt multiple levels of immersion, for instance by combining immersive VR with desktop-based VR solutions, as well as tailoring evaluation methods to the specific needs and capabilities associated with different disability levels.

Additional limitations emerged in relation to interaction. Some participants experienced difficulties when interacting with virtual UI buttons, identifying their spatial position within the VE, or using the controllers to teleport between POIs. Notably, these issues only became evident during the experimental phase, further

emphasizing the importance of iterative testing when designing experiences for this user group. A potential strategy to mitigate these difficulties would be the introduction of varying degrees of automation in the management of POI procedures. In this way, users with greater difficulties in performing specific actions could benefit from more automated procedures while still accessing the experience, whereas more autonomous users could retain a higher degree of interaction freedom. Furthermore, alternative interaction modalities, such as eye tracking or gesture-based interaction systems, may represent promising avenues for addressing these challenges.

From a methodological perspective, another limitation concerns the relatively small sample size, comprising 34 participants in the user study and 17 in the co-design phase. While recruiting larger and more homogeneous samples of ASD participants remains challenging due to ethical and practical constraints, it must be recognized that the limited sample size and the wide age range (19–67 years) reduce the generalizability of the findings. Although stratified analyses based on age and ASD trait level were performed to partially address this issue, future research should aim to involve larger and more controlled groups.

A further limitation relates to the scope of the evaluation, which was confined to a single CH setting, namely the Hall of Seasons at Palazzo Carignano. Although this context provided a realistic and well-controlled use case, the specificity of the architectural layout and exhibition content restricts the extent to which the results can be generalized to other museums or CH environments. Consequently, future work should seek to replicate the proposed methodology across different museums and cultural contexts in order to assess its broader applicability. In addition, the present methodology and prototype primarily focused on visual attention guidance, while other sensory channels, such as auditory and haptic feedback, were not systematically investigated. Given that multisensory overload represents a major challenge for many ASD individuals in real museum environments, future studies should incorporate auditory and tactile dimensions to develop a more comprehensive accessibility framework.

Further considerations concern the application of the inclusive immersion principles. The progressive enrichment process related to VE modularity was only partially implemented in the present use case, as the modular sub-elements corresponded exclusively to the cultural elements associated with the POIs, rather than to other components of the VE. On the one hand, this design choice assumes that modularity is sufficiently handled through independent POIs, without requiring additional mechanisms to control the appearance of cultural elements. On the other hand, it implies that other elements of the VE are not included in the progressive enrichment process, which may be suboptimal in scenarios where ASD users are exposed to environments with a high density of visual stimuli. A more comprehensive application of VE modularity would require additional strategies to regulate the appearance and interaction of all environmental sub-elements, as well as experimentation with different degrees of modularity. To date, this aspect has not

been extensively explored in ASD-related VR research, which has predominantly focused on application-level modularity [40, 149]. However, in the field of theory of mind, the existence of varying degrees of modularity within the human brain has been hypothesized [2]. Adopting this perspective could inform the design of more effective VXs for ASD users by also considering cognitive and sensory modules (e.g., vision, hearing, touch).

With respect to the attention principle, several participants experienced difficulties in noticing certain cultural elements. As described in the material preparation phase, two paintings within the VE were grouped into a single POI because they belonged to the same wall. Participants of the VG, however, unlike the ones of the CG, often focused on only one of the two paintings, perceiving the wall as a single cultural element associated with the POI. This observation further supports the need to explore different levels of VE modularity, such as grouping or separating cultural elements, and highlights the close relationship between modularity and attention. An alternative approach could involve the use of multimodal attention-guidance strategies, as suggested by existing literature [126], or the development of independent mechanisms to support selective and sustained attention [219]. These solutions should be systematically evaluated through additional co-design sessions with ASD users. As observed during the sessions conducted for this study, certain elements, such as the sudden appearance of UI buttons, may disrupt the experience and cause discomfort. This again underlines the importance of carefully calibrating attention-guidance mechanisms in future iterations to minimize potential sources of sensory overload.

Regarding the predictability principle, the current implementation primarily focused on intuitive UI design and repetitive interaction routines. However, previous studies have also interpreted predictability as the ability of VR applications to prepare or “anticipate” cultural environments for ASD users. Exploring this dimension of predictability could be particularly valuable, for example by developing VR tools aimed at preparing users for real-world museum visits. Such tools may help mitigate common challenges faced by ASD individuals, including anxiety or panic triggered by environmental factors such as sudden noises or intense lighting. Preliminary work in this direction already exists [60], and could be further extended through the application of the inclusive immersion principles proposed in this research.

Finally, it should be noted that the present study focused on immediate usability and museum experience, without assessing whether repeated exposure to the VR application could improve long-term accessibility in physical museums or enhance the retention of cultural knowledge. These aspects remain open questions and represent important directions for future investigation, particularly in light of the potential role of VR as a preparatory tool for real-world cultural experiences.

Future research may also explore alternative interaction techniques and additional accessibility features to be integrated into VMEs. For example, reliance on

handheld controllers may be perceived as restrictive or unnatural by some users, compared to controller-free solutions such as hand tracking or VR gloves. Similarly, teleportation-based locomotion may be less intuitive than room-scale navigation, which allows users to physically explore the environment and may improve their perception of spatial relationships within the VE. Moreover, the integration of 3D avatars and conversational agents within VMEs represents a promising research direction. Although these solutions are widely adopted in social VR applications [135, 6], their use in CH contexts for ASD users remains relatively underexplored. Investigating how avatars and agents could support communication, provide guidance, or enhance engagement within VMEs could contribute to the development of more inclusive and effective VR experiences tailored to the needs of individuals on the autism spectrum.

2.2 Closing Remarks

This chapter explored the adoption of VX design principles and immersive VR to enhance the creation of user-centered VR experiences.

The presented investigation introduced a structured methodology for the design of VR-based VMEs specifically addressing the needs of individuals with ASD. The methodology was grounded on three core design principles, i.e., modularity, predictability, and attention, identified through a systematic analysis of the existing literature. These principles were designed to support the creation of accessible and engaging VMEs, preserving the educational and experiential value of real-world museum visits while reducing potential sensory and cognitive overload.

The proposed methodology was applied in a real-world context through the development of a VR experience of the Hall of Seasons at Palazzo Carignano (Turin, Italy), designed in close collaboration with museum curators and ASD associations. To create the VE, manual creation was supported by 3D photogrammetry, employed to transpose some of the elements within the real-world environment into their digital correspondents. The obtained environment was enhanced with the proposed VX design principles tailored to ASD individuals, thus providing to the users an isolated and secure VE. The resulting application, *Palazzo Carignano VR*, was evaluated through a between-subjects user study comparing the VR-based visit with a traditional, in-person museum experience. The study aimed to investigate whether the proposed VME could foster a more positive and comfortable museum experience for users with ASD (**RQ1**), and to explore the influence of age and autism severity on the observed outcomes (**RQ2**). While a within-subjects design might have allowed for tighter control over individual differences, a between-subjects approach was intentionally adopted to limit participant fatigue and stress, which are particularly relevant considerations when working with ASD populations. This methodological trade-off is acknowledged as a limitation of the study.

The obtained findings provided empirical support for the effectiveness of the proposed VR-based VME methodology in enhancing the museum experience of individuals with ASD, thus addressing **RQ1**. The VR application was positively received and was perceived as more engaging, accessible, and enjoyable than the traditional visit across several dimensions, including interactive absorption, entertainment, esthetics, and overall museum experience. Participants reported increased learning, heightened curiosity, and appreciation for the visual quality of the environment. Notably, the interaction mechanics introduced to promote predictability and attentional guidance did not negatively affect the perceived enjoyment of the experience. This result suggests that accessibility-oriented design choices, when carefully implemented, can coexist with engaging interaction paradigms, even for users familiar with the more dynamic conventions of commercial VR applications. In relation to **RQ2**, results indicated an influence of ASD trait levels on the methodology effectiveness. Participants with a higher ASD severity were observed to gain more benefits from the experience than those with lower ASD traits. These results should still be interpreted with caution, as the assessment tool used for ASD traits evaluation (AQ Short [97]) cannot replace a full clinical diagnosis. Further studies should be direct towards the involvement of fully diagnosed ASD users for strengthening the value of the results.

Overall, the investigation described in this chapter contributed to the **MRQ1** of this thesis by proposing several user-centered design principles tailored to ASD users for enhancing VXs in the context of CH. The outcomes showed that the VME developed for the user study following the devised principles was successful in providing a more engaging and accessible experience across several dimensions, in comparison to its real-world counterpart. Future developments should be directed toward the application of the proposed user-centered design principles in contexts beyond CH, such as social training (e.g., public transportation) or formal education (e.g., in school) scenarios. The exploration of such contexts might enhance the scalability of the principles and their value for creating accessible VXs for ASD users.

Beyond the discussed limitations in the previous section, additional considerations for further research were made:

- **Effort and Interoperability:** the workflow to put the proposed methodology in practice, comprising the reconstruction of the VE, the creation of the VX, and the iterative refinement phases, represented a significant challenge in terms of human effort and invested time. Moreover, it was noted the considerable technical knowledge required for both employing techniques related to the ER domain to create the VE and manually designing the VR application for the VX. These issues led the research of the this thesis towards considering other methods for possibly automating the workflows to create VXs, while simultaneously fostering interoperability between VR and ER technologies.

- **User's Guidance:** another relevant consideration emerged in relation to user's guidance during the VX, which in the presented investigation was mainly achieved by utilizing a narrative audio guide without any embodied entity within the environment. This lack was also noticed by some participants during the experimental phase, who initially experienced some discomfort with being alone in the VE. Thus, the research of this thesis was further extended by considering the integration of embodied entities within VXs, possibly capable of providing users' guidance and in parallel simulating human behavior.

Building on the above considerations, further investigations were initiated by considering the integration of other DTTs in VXs to address the above concerns. The next chapter is focused on the use of one of these DTTs from the ER category, namely BIM, with the specific objective of fostering interoperability and the automation of workflows to create VXs.

Chapter 3

Virtual Environment Reconstruction

The work described in this chapter has been formerly published in [72].

The employment of ER technologies within VXs is not a new practice. Typically, this kind of DTTs enable the creation of VEs that mirror real-world environments. As described in the previous chapter, the advantages of the latter possibility are extremely valuable in scenarios involving CH sites (e.g., museums, historical buildings), where these DTTs are usually employed to create VXs for heritage preservation or accessibility purposes. By utilizing 3D reconstruction techniques, these real-world locations can be digitally transposed in order to create the VE of a Virtual Heritage Experience (VHE) (e.g., a virtual visit in a museum). Furthermore, the digitization of built heritage has been enhanced through the use of management technologies, such as BIM, also fostering the understanding of CH sites. However, challenges within this context are still present. As seen in Chapter 2, the VX creation process is often time consuming and requires extended knowledge of both VR and environment reconstruction DTTs. Although BIM-based DTTs could possibly address this issue, they are mainly used for the maintenance of CH sites and have very little consideration regarding the creation of VHEs. Moreover, current BIM-to-VR tools (e.g., software applications, plugins, file formats) to ease the transposition of VEs in VR platforms are often unavailable given their proprietary nature and/or high costs.

This chapter presents an investigation aiming to address the previous challenges by employing ER technologies for the creation of VXs. Specifically, it investigates the usage of BIM in the context of CH, which in literature is called Heritage Building Information Modeling (HBIM). Since the process of transferring HBIM models into VR applications, named HBIM-to-VR, is usually mediated by restrictive and proprietary tools, the investigation proposes a semi-automated workflow based on an open-source BIM format file and a VR platform. Moreover, the devised VR platform generates a VHE (i.e., a virtual visit) leveraging the

reconstructed VE.

3.1 HBIM for Virtual Heritage Experiences

This section describes an investigation conducted in the context of CH for creating VHEs employing HBIM-based DTTs in combination with VR. The following subsections provide an overview of the topic, the related research within the context, the devised methodology and its application in a real-world use case.

3.1.1 Overview

The rapid evolution of digital technologies has increasingly highlighted the role of digitization as a strategic means to enhance the value of built heritage. In this context, BIM has emerged as a central DTT, commonly defined as “a digital representation of physical and functional characteristics of a facility” [161]. BIM functions as a shared and structured repository of information related to a built asset, supporting informed decision-making throughout its entire life cycle, from early conception to operation and maintenance [161]. Although BIM was initially introduced for design-related purposes by Nederveen and Tolman in the early 1990s [162], its widespread adoption only gained momentum from the 2010s onward [139]. BIM-based workflows integrate heterogeneous data acquired during construction and documentation processes, enabling the creation of information-rich digital models for the management of built artifacts [5].

While BIM tools are primarily employed within the Architectural, Engineering, and Construction (AEC) sector, their application to heritage contexts has led to the development of HBIM [159]. This concept extends the advantages of BIM by supporting interoperability, structured data storage and updating, and the representation of complex and irregular geometries typical of historical buildings [139]. Over time, HBIM has become a widely adopted approach for the documentation, preservation, and valorization of CH sites.

In parallel with the advancement of HBIM, immersive technologies have increasingly been explored as a means to activate and experience digital heritage models. Among these technologies, VR has attracted particular interest due to its ability to support interactive and immersive experiences [5]. Within the CH domain, VR has been extensively employed to create VHEs, a broad category encompassing applications such as virtual tours and exhibitions aimed at supporting education [171], improving accessibility and inclusion [189], and assisting heritage conservation efforts [195].

VR-based VHEs share several characteristics with commercial VR applications, including the use of advanced information visualization techniques [148]. These

typically involve UIs based on text, images, or videos, attention-guidance mechanisms [138], and detailed 3D representations of heritage assets within the VE [44]. Such experiences may be designed for stationary usage (seated or standing), allowing users to freely look around [108], or may exploit room-scale VR configurations combined with artificial locomotion techniques such as continuous movement [177] or teleportation [197, 207]. Interaction with the VE is commonly supported through hand-held controllers, enabling manipulation of UI elements (e.g., buttons [16]) and virtual cultural objects [201]. Audio narration is also frequently integrated, either through pre-recorded guides [136] or real-time Text-To-Speech (TTS) solutions [16], to enhance accessibility and adaptability.

In some instances, VHEs are developed by directly exploiting BIM platforms and their associated tools [222]. More frequently, however, general-purpose real-time graphics engines such as Unity or Unreal Engine are adopted, leveraging technologies originally conceived for video games and interactive applications [109].

Within this latter approach, a major challenge lies in the data exchange between BIM platforms and VR engines. BIM-to-VR workflows often involve multiple conversion steps and require a combination of modeling and programming expertise. Furthermore, these pipelines frequently depend on proprietary plugins or add-ons, which can limit transparency and interoperability. Similar challenges characterize HBIM-to-VR workflows, where several preparatory operations and tools are typically required even to achieve basic visualization of heritage sites in VR [21].

To address these issues, recent research efforts have explored the automation of BIM-to-VR workflows by leveraging the Industry Foundation Classes (IFC) standard [160]. IFC is an open and internationally recognized format (ISO 16739-1:2024 [104]) that defines a comprehensive schema for representing not only geometric information, but also semantic data such as object identities, attributes (e.g., materials, colors), and relationships among building components [113].

Within BIM-to-VR applications, IFC data are commonly exploited to support tasks related to construction, restoration, and facility management. In the context of HBIM-to-VR, however, the structured retrieval of semantic information associated with heritage elements opens additional opportunities. For example, such data can be leveraged to enhance storytelling and educational narratives, thereby increasing the cultural value of CH sites [167]. These considerations suggest that enriched IFC models could serve not only as a basis for navigable VEs, but also as a foundation for fully interactive VHEs, including virtual visits and exhibitions.

Despite this potential, the use of IFC-based HBIM models for the automatic generation of VHEs remains largely underexplored. Existing VR platforms specifically designed to generate VHEs, such as those proposed by Hayashi et al. [93] and Kiourt et al. [116], typically rely on external sources for geometric and semantic data rather than directly exploiting HBIM models.

This limited adoption may be attributed, on the one hand, to the requirement for specialized expertise in both HBIM modeling and VR development, which can

represent a significant barrier for non-technical users [173]. On the other hand, technical limitations persist in IFC-based HBIM-to-VR workflows, including potential information loss during data import into game engines [68] and the absence of standardized practices for encoding CH-specific data within HBIM models [203]. Moreover, to the best of the author’s knowledge, no established methodology currently exists for embedding all the information required for generating a VHE directly into an IFC file.

Building on these considerations, the investigation presented in this section aims to exploit the potential of HBIM models and the IFC standard to support the generation of VHEs for cultural heritage sites. To this end, a novel semi-automated HBIM-to-VR workflow is introduced, structured into two main phases:

- A first phase focused on the manual enrichment of an IFC file associated with an HBIM model, with the aim of supporting the generation of a virtual visit. The required features were identified through an analysis of the state of the art in VHE creation workflows and platforms, and include the definition of points of interest with descriptive content, the planning of visitor routes, and the control of locomotion within the VE;
- A second phase based on a Unity-powered VR platform specifically designed to automate key steps of the HBIM-to-VR pipeline. This platform processes the enriched IFC file and related external resources, reconstructs the VE and its object hierarchy, and dynamically generates the interactive experience, enabling the creation of a complete virtual visit entirely at runtime.

The proposed workflow seeks to advance the state of the art in VHE development by simplifying the transition from HBIM to VR, while improving accessibility and scalability. Based on the definition of the devised workflow, the investigation seeks to answer the following research question:

- **RQ:** To what extent is the proposed semi-automated HBIM-to-VR workflow capable of reconstructing the VE of a CH site and generating a VHE starting from an enriched IFC file?

To answer this research question, also building on the considerations at the end of Chapter 2, the applicability of the workflow was demonstrated by leveraging the same environment of the previous investigation, namely the “Hall of Seasons” at Palazzo Carignano in Turin, Italy.

3.1.2 Background

This section reviews the current state of the art in the field, outlining the main research directions and methodological approaches related to BIM, HBIM, and openBIM, HBIM-to-VR workflows, VR-based VHEs integrated with museum contexts, and both academic and commercial VHE authoring platforms.

BIM, HBIM, and openBIM

As discussed in Section 3.1.1, the BIM methodology has proven to be an effective approach for the digitization and systematic archiving of the physical and functional information of buildings. As highlighted by Babalola et al. [12], BIM has led to substantial improvements in process management within the AEC industry over the past years and is now widely adopted at a global scale.

Although BIM was originally conceived and predominantly employed for new construction projects, its application has progressively extended to historic buildings and, more broadly, to the CH domain. In this context, the term HBIM is commonly adopted. HBIM differs from conventional BIM in that it must address specific requirements related to historic assets, aiming to produce 3D digital models that faithfully represent not only the geometry of heritage structures, but also their material characteristics and historical attributes. These models are typically generated by integrating data acquired through photogrammetric surveys and laser scanning, ensuring a high level of geometric and semantic accuracy [137].

As reported in [1], BIM authoring tools primarily focus on information management and are often limited in their ability to handle the complex and irregular geometries that characterize historic buildings. As a consequence, the creation of HBIM models frequently relies on the combined use of BIM software and dedicated 3D modeling tools, such as Blender or Rhinoceros, which are better suited for processing point clouds obtained from survey campaigns.

The enrichment of BIM models to support different phases of a building's life cycle is regulated by standards and norms, such as EN ISO 7817-1:2024 [64] in European Union countries. These standards define the required geometric and information levels to enable the use of BIM models across design, construction, and operational phases. However, as discussed by Bianchini et al. [19], the peculiar challenges posed by HBIM modeling are often not fully addressed by existing standards, thus necessitating the adoption of customized solutions.

Given the frequent need for data exchange across heterogeneous software environments, both during modeling activities and application development, interoperability remains a central topic in BIM and HBIM research. A widely adopted strategy to mitigate interoperability issues is the use of the openBIM methodology, which promotes a collaborative and vendor-neutral workflow aimed at ensuring interoperability throughout the entire life cycle of buildings and assets [31]. In practice, openBIM relies on standardized, non-proprietary formats, such as the BIM Collaboration Format for issue management during modeling phases, CityGML for city-scale representations, and the IFC format for the exchange of BIM model data [30]. The IFC standard, developed and maintained by buildingSMART¹, enables the structured representation of both geometric and semantic information and

¹buildingSMART: <https://www.buildingsmart.org/>

is formally defined in the ISO 16739-1:2024 international standard.

The advantages of adopting open-BIM approaches for data exchange and interoperability have been extensively documented in the literature, including the work by Oostwegel et al. [165].

HBIM-to-VR

In the last years, the integration of HBIM models with VR technologies for the development of VHEs has attracted increasing attention in the literature.

A widespread strategy adopted by HBIM-to-VR workflows relies on the use of commercial game engines, such as Unity or Unreal Engine, as target platforms for immersive visualization. For example, Graham et al. [84] investigated the transition from BIM to VR by introducing a system based on multiple LODs, referred to as LODIA. In their study, the authors designed a narrative-driven VR experience in which areas characterized by higher geometric and visual fidelity were used to convey increased heritage relevance. A key challenge identified in this work concerned the difficulty of transforming highly detailed BIM models into VR-ready assets while preserving visual quality, a process that demands careful data handling and advanced technical expertise. Comparable conclusions were drawn by Banfi et al. [13], who analyzed the evolution of interactivity, immersion, and interoperability in HBIM-based workflows. Their findings emphasized that the technical complexity of these pipelines, together with the requirement for substantial resources and specialized skills, often represents a barrier to the widespread adoption of VR solutions in the heritage domain.

A comprehensive HBIM-to-VR pipeline was proposed by Pybus et al. [179], who focused on the digital reconstruction of the *Parliament Hill* site in Canada. Their workflow comprised multiple stages, including data acquisition, geometric refinement of the HBIM model through external modeling software (e.g., Rhinoceros), and the optimization of materials and textures for real-time rendering. The primary objective was to deliver a high-fidelity VR experience suitable for public dissemination, while maintaining acceptable performance levels. Although the approach proved effective from a visual standpoint, it remained highly resource-demanding and relied heavily on manual operations, thus underscoring the challenges associated with managing large HBIM datasets and ensuring interoperability across heterogeneous software environments.

Interoperability issues have also been addressed by studies adopting IFC-based strategies for importing HBIM models into game engines. For instance, Ferretti et al. [68] developed a semantically enriched HBIM model of *Palazzo Ducale* in Urbino, Italy, featuring intelligent object parametrization. The authors employed Unity to implement a VR application aimed at digital curation and storytelling of cultural heritage contents. The HBIM model included detailed semantic information related

to interior elements, such as artworks’ catalogue identifiers, authorship, and spatial localization. Nevertheless, the study highlighted several technical limitations, notably the difficulty of importing IFC-based HBIM models into Unity without incurring information loss, as well as the considerable computational overhead required to visualize highly detailed heritage models in immersive VR environments.

In addition to geometric complexity, materials and textures defined within BIM models are generally not suitable for real-time VR applications. Aiming to address this issue, Mahecha et al. [63] explored the graphical capabilities of IFC by proposing a user interface for reassigning textures in IFC-based VR workflows. Their solution was built upon the IfcOpenShell library², an open-source toolkit for IFC processing and geometry generation, and enabled users to associate custom image files with materials defined in the IFC schema. The resulting output consisted of a DAE model and a corresponding texture repository. However, the proposed interface supported only basic texture transformations, such as scaling and rotation, with textures being repetitively mapped onto surfaces without explicit consideration of UV coordinates. Moreover, the import process led to the loss of most BIM-related properties within the VR environment, driving the research toward the introduction of an additional XML export to preserve semantic information.

Overall, existing HBIM-to-VR workflows provide effective solutions for transferring HBIM models into VR development environments. Nevertheless, as discussed in the following section, VR-based VHEs typically require a broader set of interactive and experiential features that go beyond mere geometric visualization. Such functionalities cannot be fully addressed either at the BIM modeling stage or through the standard IFC format alone.

VHEs Integrated with Museum Environments

In recent years, museums have increasingly explored the integration of VR-based experiences within their exhibition spaces, with the objective of extending their educational, cultural, and experiential roles through immersive technologies [201]. Despite this growing interest, the prevailing practice still consists in commissioning bespoke VR applications developed ad hoc for specific exhibitions or installations. As a result, museums often face the challenge of achieving outcomes comparable, in terms of quality and functionality, to those produced by professional VR developers. An analysis of representative use cases can therefore help identify a baseline set of features that characterize effective museum-oriented VHEs.

An early example of VR adoption in a real museum context is reported by Jung et al. [108], who selected the Geevor Tin Mine Museum in Eastern Cornwall as a case study due to its emphasis on technological innovation for visitor engagement.

²IfcOpenShell: <https://ifcopenshell.org/>

In their experiment, visitors experienced an immersive VR application delivered through a Samsung Gear headset hosting a smartphone, which simulated the descent into a mining shaft and narrated the miners' first encounter with their working environment. Experiences of this kind typically provide a limited degree of interactivity, as user input is restricted to head rotations exploiting the three Degrees of Freedom (DOF) system supported by mobile-based HMDs. Furthermore, the computational constraints of mobile devices limit real-time rendering quality, often necessitating the use of pre-rendered and non-interactive 360° imagery or video content.

A more articulated approach is described by Puig et al. [178], who reported on the development of two immersive experiences conceived as extensions of a museum exhibition: a 360° video delivered via Samsung Gear and a fully interactive, gamified VR experience implemented using the HTC Vive system, including head-mounted display and hand controllers. Both experiences were designed in the context of an exhibition dedicated to the Neolithic settlement of La Draga at the Museu d'Arqueologia de Catalunya in Barcelona. To assess the contribution of VR to the overall museum visit, the authors recruited volunteers among museum visitors, who accessed the VR experiences after attending an introductory workshop conducted by an archaeologist specialized in the Neolithic period. The findings indicated that UX was strongly influenced by participants' prior familiarity with VR technologies, underlining the importance of designing museum-oriented VHEs that account for potential usability barriers. The use of an advanced, tethered VR system such as the HTC Vive addressed several limitations associated with mobile VR, enabling room-scale, six DOF tracking and richer graphical rendering. This configuration allowed users to navigate the VE either through physical movement within the tracking area or by means of artificial locomotion techniques implemented via hand controllers, including teleportation and continuous movement. Additionally, the controllers supported interaction with virtual objects and UIs through both direct manipulation and ray-casting mechanisms.

A prominent example of a large-scale museum VHE is represented by the project "*Mona Lisa: Beyond the Glass*", developed by the Louvre Museum [136]. This experience was conceived to allow visitors to metaphorically enter the world of the famous painting, offering an alternative mode of access and engagement. The VHE is characterized by high-resolution digital reconstructions and interactive components that promote a multisensory form of exploration. An audio guide based on pre-recorded narration is also included to accompany users throughout the experience. While the use of pre-recorded audio contributes to narrative consistency and user immersion, it introduces limitations in terms of flexibility, as any modification to the content requires re-recording the associated audio clips. This contrasts with alternative approaches based on TTS technologies, which allow for dynamic generation of spoken content.

More recently, Shahab et al. [191] investigated the impact of VR installations

in museum settings with respect to visitor enjoyment and learning outcomes. Their study, conducted at the VRlab of the Deutsches Museum in Munich, Germany, involved four VR experiences centered on historically significant machines, including the Apollo lunar roving vehicle, the Sulzer brothers' steam engine, Otto von Lilienthal's glider, and Carl Benz's Patentmotorwagen. These artifacts were digitally reconstructed using laser scanning and satellite imagery to exploit the photorealistic rendering capabilities of VR. The experiences also incorporated interactive elements, such as a lunar golf simulation. The results suggested that carefully designed VHEs, which balance entertainment and educational content while fostering a strong sense of presence, can substantially enhance visitor engagement and learning. Nonetheless, the authors noted that museums must continuously adapt their VR installations to keep pace with rapid technological evolution and shifting user expectations.

Another recent deployment of VR in a museum context is described by Restivo et al. [184], who presented a VHE exhibited at the Kings' Gallery of the Museo Egizio in Turin, Italy. In this experience, visitors are invited to interact directly with virtual representations of museum artifacts within a controlled VE. The setup resembles a cultural center environment, where users are seated opposite a digital curator. Virtual remains are placed on a table, and interaction is enabled through hand controllers, allowing detailed inspection of the objects. The experience deliberately avoids artificial locomotion, focusing instead on seated interaction, and features a virtual agent equipped with TTS capabilities to provide contextualized explanations. Feedback from both visitors and museum staff was largely positive, highlighting the effectiveness of VR as a complementary tool that enhances, rather than replaces, traditional museum visits.

Overall, the reviewed works illustrate that VR applications integrated into museum environments differ substantially in terms of features, interaction paradigms, and technological sophistication. This heterogeneity complicates the standardization of development pipelines and makes the creation and maintenance of VHEs dependent on a broad set of specialized competencies. Within this context, platforms specifically designed for authoring VHEs may play a crucial role in simplifying and partially standardizing the production of VR-based CH experiences.

VHE Authoring Platforms

Beyond the scope of BIM and HBIM-based workflows, several research efforts have explicitly addressed the need to lower the technical barrier associated with the creation of VHEs. These works typically rely on VR authoring platforms specifically conceived to support the design and deployment of virtual exhibitions and visits. Such platforms are generally conceived with usability and accessibility in mind, and aim to provide end users with flexible tools to manage the structure, contents, and interaction mechanisms of the intended experiences with minimal technical effort.

Over time, different strategies have been explored to pursue this objective. An early example is represented by the VIRTUE system proposed by Giangreco et al. [79], a VR platform designed to support the creation of customizable and immersive virtual museum exhibitions. The system enabled museum curators to assemble exhibitions composed of both static and dynamic 2D and 3D artifacts, and to configure exhibition spaces through adjustable textures, lighting conditions, and soundscapes. End users could then explore the resulting VEs using VR devices. Despite these capabilities, the authors acknowledged the lack of automated mechanisms for artifact placement as a major limitation of the system.

More recently, Korkut and Surer [120] introduced a framework for the creation of PC, VR, and MR 3D environments grounded in the concept of *heterotopias*, defined as spaces in which multiple layers of reality coexist to support knowledge organization and exploration. Their work focused in particular on virtual museums, adopting procedural generation techniques to manage both spatial layouts and content layers. The framework supports multiple content types, including text, images, audio, video, and 3D models, and introduces object parametrization to identify core architectural elements (e.g., walls, floors, roofs) required for constructing the digital environment. Along similar lines, Zidianakis et al. proposed the *Invisible Museum* platform [230], a web-based solution aimed at democratizing the creation of virtual exhibitions through an accessible tool named *Exhibition Designer*. The platform is characterized by a high degree of flexibility, offering predefined templates, customizable floor plans, and tools for configuring spatial layouts, exhibits, and lighting conditions.

Other studies have investigated alternative paradigms for VHE creation, progressively shifting from user-driven authoring tools toward approaches characterized by a higher degree of automation. One of the earliest contributions in this direction is Minerva, proposed by Amigoni and Schiaffonati [8], a system conceived to assist museum curators by automating several phases of exhibition preparation and artifact allocation. During its evaluation, Minerva was employed by curators to classify and arrange artworks within a virtual reconstruction of the Caserma Napoleonica in Milan, Italy. While the system demonstrated its potential to support the generation of effective VHEs and received positive feedback from domain experts, the authors also acknowledged the need to extend its applicability to a broader range of users.

A comparable level of automation is found in the work by Goto et al. [82], who developed a system capable of automatically generating multiple virtual museum scenes. Their approach leveraged a database query language to retrieve 3D models and related metadata, while Unity was used as the underlying VR platform for rendering the generated environments. Hayashi et al. [93] proposed a system focused on the creation of personalized virtual museum visits, allowing users to curate exhibitions by selecting artifacts from online repositories such as Wikimedia Commons. However, this approach was limited to textual and image-based content.

An earlier framework, *Dynamus*, introduced by Kiourt et al. [116], adopted a user-centric design philosophy to simplify the creation of VHEs, and included features such as runtime retrieval of 2D and 3D assets, as well as basic management of lighting and shadows.

Beyond academic research, a number of commercial authoring platforms have emerged to support the creation and dissemination of VHEs. These solutions similarly aim to simplify planning and development processes, often targeting non-expert users [114]. Nevertheless, the quality and expressive power of the resulting experiences are strongly dependent on the specific features offered by each platform. The following platforms were selected for analysis based on their relevance within the CH domain, their feature sets, and their presence in prior comparative studies [114, 215, 205].

Kunstmatrix [124], for example, is an online platform dedicated to the creation and curation of professional 3D virtual galleries and exhibitions. It allows artists, curators, and galleries to present artworks, sculptures, and multimedia elements within customizable virtual spaces, and supports immersive exploration through VR headsets such as Meta Quest and HTC Vive. The platform also enables the definition of guided tours through the specification of guiding points. While this facilitates the rapid creation of VHEs, users are constrained to a set of predefined exhibition spaces, and content uploads appear to be limited primarily to image-based artworks, without native support for 3D models or other advanced media.

Another platform oriented toward virtual exhibition creation is *ArtPlacer* [11], which primarily targets artists interested in showcasing and selling their works. In addition to predefined environments, the platform offers the possibility to create custom virtual spaces, together with basic customization options such as wall, floor, and ceiling textures and colors. At present, however, the platform does not seem to support immersive exploration using VR headsets.

A broader range of features is provided by the web-based desktop VR platform *New Art City* [58]. This platform enables users to design immersive experiences within customizable VEs, supporting the upload of 3D models, audio, images, and video content. Users can also adjust several environmental parameters, such as sky color and floor type, and introduce interactive elements into the scene. The platform includes a feature for defining reflection behavior on reflective materials, although it does not offer full control over lighting settings. Unlike other solutions, it does not provide predefined environments, requiring users to import external 3D models to define the exhibition space.

One of the most widely adopted platforms for creating VHEs is *Artsteps* [51], which is specifically designed to support the construction, exploration, and sharing of interactive 3D virtual exhibitions. Users can start from an empty virtual space and progressively shape the exhibition area using built-in tools, as well as by importing 3D models, images, audio, and textual descriptions. Alternatively, predefined environments can be purchased directly from the platform, reducing the

effort required to build a VE from scratch. Artsteps also supports the definition of narrative structures through guiding points positioned at specific locations within the environment. Immersive exploration via VR headsets is available, along with a limited set of object interaction capabilities. However, advanced control over lighting and reflections is not supported, which may restrict the realism achievable within the VE.

A further alternative is *VR-All-Art* [216], a platform aimed at supporting virtual exhibition creation for galleries, artists, and museums. The platform allows the development of both public and private exhibitions, which can be accessed globally via VR headsets or mobile devices. VEs can be derived from real-world spaces and enriched with externally created 3D models. In contrast to several other platforms, VR-All-Art provides tools for managing lighting within the VE, although reflection handling is not supported. At present, the platform is not freely available and appears to be primarily targeted at professional or expert users.

Overall, the approaches and platforms discussed in this section share the common objective of enabling the creation of VR-based VHEs. Nevertheless, as summarized in Table 3.1, each solution is characterized by a specific combination of features and constraints. Moreover, none of the examined platforms is currently capable of generating a VR application that fully matches the functional richness and quality of bespoke VR experiences developed specifically for museum contexts.

		[116]	[120]	[8]	[79]	[93]	[230]	[82]	[51]	[11]	[124]	[216]	[58]	This work
VR type	Immersive VR	✓	✓	-	✓	-	✓	✓	✓	-	✓	✓	-	✓
	Desktop VR	✓	✓	✓	-	✓	✓	-	✓	✓	✓	✓	✓	✓*
External data retrieval	Text descriptions	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Images	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Videos	-	✓	-	✓	-	✓	-	✓	-	-	✓	✓	✓
	Audio descriptions	-	P	-	P	-	P	-	P	-	P	P	P	G
	3D models	✓	✓	✓	✓	-	✓	✓	✓	-	✓	✓	✓	✓
	BIM information	-	-	-	-	-	-	-	-	-	-	-	-	✓
Graphics and rendering	Lighting management	G	G	-	G	P	G	-	-	P	-	G	G	G/P
	Reflections management	-	-	-	-	P	-	-	-	-	-	P	-	G
	LODs handling	-	-	-	-	-	-	-	-	-	-	-	-	✓
Locomotion system	Teleport points	-	-	✓	✓	✓	✓	-	✓	✓	✓	-	-	✓
	Teleport areas	-	✓	-	✓	-	-	-	✓	-	✓	✓	✓	✓
	Continuous movement	✓	✓	✓	-	✓	✓	-	✓	✓	✓	✓	✓	✓
User Interface (UI)	UI panels	✓	✓	-	-	✓	✓	-	✓	✓	✓	-	✓	✓
	UI buttons	-	✓	-	-	✓	✓	-	-	-	✓	✓	✓	✓

Table 3.1: Characteristics of existing VHE platforms analyzed in Section 3.1.2, in comparison with the proposed solution. ✓ = Available, - = Not available, ✓* = Planned as future work, G = Generated, P = Predefined

Considerations

The analysis of the state of the art highlights that the integration of BIM, HBIM, and VR technologies holds considerable promise for the development of effective VHEs. Nevertheless, existing workflows are still largely characterized by manual procedures, the combined use of heterogeneous tools, and the requirement for advanced technical skills. These factors significantly hinder the widespread adoption of such solutions. At the same time, platforms designed to simplify or automate the creation of VHEs and make them accessible to less experienced users continue to present notable limitations, often producing results that are qualitatively inferior to the bespoke VR applications currently developed by professional teams within the museum sector.

One of the most critical challenges concerns the creation of visually realistic virtual environments. While many platforms provide basic customization options, such as texture assignment and simple lighting controls, these features are frequently insufficient to achieve a convincing level of realism. The absence of more advanced rendering capabilities, including dynamic lighting, reflections, and LOD management, can severely constrain both immersion and performance optimization. This issue is particularly evident in VHE authoring platforms, which tend to prioritize simplicity and ease of use at the expense of expressive power. Such limitations can directly affect user engagement, especially in virtual museum applications or experiences that rely heavily on narrative depth and visual authenticity.

Another relevant aspect relates to the dependence on predefined VEs or externally sourced assets. Many authoring platforms either provide a set of predefined templates or allow users to populate empty spaces by importing 3D models from external repositories. While this strategy lowers the entry barrier for non-expert users, it also introduces potential drawbacks. In particular, the reliance on proprietary formats and platform-specific workflows may generate interoperability issues for professional users. Moreover, cultural data associated with the VHE are often either embedded within the platform itself (e.g., predefined environments) or manually uploaded by users (e.g., 3D models, images, videos), which raises concerns from the perspective of data standardization and long-term reuse.

With respect to HBIM-based approaches, prior studies have identified additional challenges. Many works focusing on VHE creation require the import of HBIM models into game engines, a process that must handle both geometric and semantic information. To address this complexity, several approaches have moved toward automated data import pipelines, which facilitate model updates and allow modifications to geometry and semantics without repeated manual re-import operations. Despite these advances, the automation of the VHE deployment process itself remains largely underexplored, even though it represents a key opportunity to improve scalability and reduce development time and resource consumption.

Furthermore, although several platforms incorporate interaction and guidance

mechanisms to support user engagement and navigation, these systems are often relatively basic. Advanced features such as automated audio generation through TTS technologies or adaptive and context-aware user interfaces are rarely supported. In addition, cultural information, such as artwork descriptions, is frequently entered manually within the platform, which can further limit interoperability and hinder the adoption of standardized data representations.

Taken together, these observations underline the need for interoperable and standardized data formats capable of encompassing not only the information required to construct realistic and interactive VEs (e.g., geometry and material parameters), but also semantic content (e.g., cultural descriptions) and experience-related settings (e.g., guiding point locations). Such formats are essential to support the creation of engaging, immersive, and reusable VHEs.

In this context, the workflow proposed in the present investigation aims to address several of the identified limitations by substantially simplifying the generation of VHEs starting from HBIM-related IFC files. By leveraging IFC as a unified and extensible data container, the proposed approach enables the definition of structured data relationships and supports the reconstruction of high-fidelity VEs enriched with interaction capabilities. Moreover, the adoption of a dedicated VR platform allows the semi-automatic generation of VHEs directly from enriched IFC files, thereby streamlining the development pipeline and making it accessible also to non-expert users.

3.1.3 Materials and Methods

As discussed in the previous sections, existing approaches for generating VHEs from HBIM data are affected by several limitations, including restricted interoperability, highly manual workflows, and a lack of standardized data management strategies. To address these issues, the workflow proposed in this investigation introduces a streamlined approach for the creation of immersive and interactive VHEs, structured around two main phases: preparation and deployment. The preparation phase focuses on data export, enrichment of the HBIM model at the IFC level, and the organization of related external resources. The deployment phase, instead, relies on a VR platform that dynamically reconstructs and generates the virtual tour at runtime by leveraging the enriched IFC file together with its associated external data, ultimately delivering a fully immersive VHE.

The proposed VR platform has been developed using the Unity³ game engine and is designed to support a broad range of functionalities specifically tailored to VHE creation. In particular, the platform adopts a runtime-oriented approach for data retrieval, external file loading, and the reconstruction of both geometric and

³Unity Engine: <https://unity.com/>

semantic information encoded within the IFC file, enabling the procedural generation of the VE. The reconstruction of the HBIM model is performed through the use of *IfcConvert*⁴, a tool provided by the IfcOpenShell framework. In addition to model reconstruction, the platform supports several advanced features aimed at enhancing realism and performance, including space-adaptive lighting and reflections, LOD management for high-polygon meshes, and real-time audio generation through TTS services based on the ElevenLabs API⁵.

Preparation Phase

This subsection describes the data preparation and export procedures, including the enrichment of the HBIM model at the IFC level and the organization of the associated external resources. In particular, the external data considered within the workflow include environment textures and UV coordinates, as well as additional 3D meshes semantically linked to the HBIM model and required for VR visualization.

Multiple software tools are involved in this phase. Specifically, the Autodesk Revit⁶ suite is used for HBIM model development and IFC file export. Texturing operations and the export of model-related data are handled through the Blender 3D modeling software, supported by a custom Python script.

From the HBIM Model to the Enriched IFC File

The proposed workflow starts with the definition of the contents that need to be embedded within the HBIM model. A crucial step in this process is the identification and specification of the IFC entities, namely the objects to which geometries and parameters are associated. This step is fundamental for generating an enriched IFC file capable of storing all the information required for the creation of the VHE.

The VHE is conceived as a virtual visit structured around a set of POIs. These POIs represent conceptual observation points within the VE, corresponding to locations where users can access information related to the CH site [200]. In addition to POIs, another essential category of elements to be identified within the HBIM model is represented by the Look-At Objects (LAOs), such as decorations, artworks, and architectural components. Each LAO is associated with a specific POI through the definition of Property Sets (Psets), enabling the semantic linkage between observation points and related objects.

The specification of these elements is achieved through the enrichment of the HBIM model with a set of custom parameters, summarized in Table 3.2.

⁴IfcConvert: <https://docs.ifcopenshell.org/ifcconvert.html>

⁵ElevenLabs API: <https://elevenlabs.io/api>

⁶Autodesk Revit: <https://www.autodesk.com/it/products/revit/overview>

In detail, POIs are equipped with a custom Property Set named *Pset_POI*, which includes the following three parameters:

- *ID*: an integer value used to uniquely identify the POI during the virtual visit;
- *Sequential*: a boolean value indicating whether the POI follows a sequential visualization of the associated LAOs (alternatively, free exploration is allowed);
- *Description*: an optional textual description of the POI, which can be used to provide contextual information related to the specific spatial position (e.g., an introductory message welcoming the user at the beginning of the experience).

Similarly, LAOs are associated with a dedicated Property Set named *Pset_LookAt*, which contains the following parameters:

- *ID*: a string used to uniquely identify the LAO within the VE;
- *View_Point*: the identifier of the POI representing the observer's position; this parameter establishes the association between a POI and one or more LAOs;
- *ImageURI*: a text value specifying the path to an image associated with the LAO, which is displayed to the user upon activation;
- *VideoURI*: an optional text value specifying the path to a video associated with the LAO, which can be visualized through a dedicated interaction;
- *Order_Index*: an integer value defining the visualization order of the LAO when it belongs to a sequential POI;
- *Description*: a textual description of the LAO, presented to the user both visually and through audio narration.

As a final step of the preparation phase, the HBIM model is further enriched by assigning a Pset named *Pset_Material_Rendering_Properties* to each *IfcMaterial*. This Pset stores rendering-related parameters required to accurately reconstruct material appearance within the VR platform. Specifically, the following properties are defined:

- *IsTriplanar*: a boolean value indicating whether the material relies on triplanar mapping (true) or UV mapping (false);
- *Roughness*: a floating-point value representing surface roughness, ranging from 0.0 (perfectly smooth) to 1.0 (completely rough);

- *Metallic*: a floating-point value indicating the metallic nature of the material, where 0.0 corresponds to a non-metallic surface and 1.0 to a fully metallic one;
- *Transparency*: a floating-point value defining the material transparency, ranging from 0.0 (fully opaque) to 1.0 (fully transparent);
- *Tiling_X*: a floating-point value specifying the horizontal texture tiling factor, expressed as the number of texture repetitions per meter along the X axis;
- *Tiling_Y*: a floating-point value specifying the vertical texture tiling factor, expressed as the number of texture repetitions per meter along the Y axis.

Object	Class	Type	Pset	Parameter name	Parameter type
Point of Interest	<i>IfcBuildingElementProxy</i>	POI	Pset_POI	ID	Int
				Sequential	Bool
				Description	Text
Look-at Object	General	General	Pset_LookAt	ID	Text
				View_Point	Int
				ImageURI	Int
				Order_Index	Int
				Description	Text
Material	<i>IfcMaterial</i>	NA	Pset_MaterialRenderingProperties	IsTriplanar	Bool
				Roughness	Float
				Metallic	Float
				Transparency	Float
				Tiling_X	Float
				Tiling_Y	Float

Table 3.2: IFC entities and parameters defined within the HBIM model.

Texturing and UV Mapping

To enhance the geometric information provided by the HBIM model, the proposed workflow includes a dedicated texturing and UV mapping step carried out using Blender. This step is particularly relevant for real-time graphics applications,

where high visual fidelity cannot be achieved by relying exclusively on the geometric primitives and basic material attributes typically stored in BIM data (e.g., vertices and generic material properties).

During this phase, image-based textures (e.g., PNG files) are assigned to the 3D models, and UV coordinates are explicitly defined for each mesh. While textures convey the visual appearance of the surfaces, UV mapping determines how these textures are projected onto the geometry. This combination represents the primary technique for depicting fine-grained visual details that are characteristic of VHEs, such as paintings, photographic materials, and textual transcriptions.

The operations are performed using Blender version 4.1, together with the IfcOpenShell Blender-BIM add-on, which provides a set of tools for importing, editing, and exporting BIM models in the IFC format. Blender also enables manual UV unwrapping on individual meshes and supports the assignment of multiple materials to different portions of the same geometry. This capability is particularly important, as several IFC-related features, especially those concerning texture handling and UV coordinates, are not consistently or fully supported by conventional BIM authoring platforms. Nonetheless, these features are crucial to achieving a level of visual quality suitable for virtual visits.

Within the IFC schema, texture support is primarily managed through the *IfcSurfaceStyle* entity, which can include sub-entities such as *IfcSurfaceStyleWithTextures*. This entity references *IfcImageTexture*, which in turn points to an external image resource (e.g., JPG or PNG) via a URI or embeds the image directly within the IFC file as a pixel array. UV mapping information is similarly represented through entities such as *IfcTextureCoordinates* and *IfcTextureMap* [32]. However, texture-related entities are often not prioritized in BIM software implementations, making the export of fully textured IFC files a complex and error-prone process. Furthermore, workflows involving game engines typically rely on proprietary and sometimes costly plug-ins to correctly import textured IFC data.

To overcome these limitations, the proposed workflow adopts an alternative strategy that enables a custom export and import of texture and UV mapping data from Blender to Unity.

Once texturing and UV mapping are completed, texture-related information is exported to a local directory named “TextureData” by executing a custom Python script directly within Blender. Specifically, UV mapping data are stored in a JSON file that contains, for each mesh, the vertex coordinates and their corresponding UV coordinates, uniquely identified through the *GlobalId* associated with each element in the HBIM model. In addition, the script exports all texture assets used by the materials of the HBIM model as PNG files, placing them in the same directory for subsequent use during the deployment phase.

High-Definition Models

To preserve the lightweight nature of the enriched IFC file, the proposed workflow supports the integration of high-polygon 3D assets, hereafter referred to as High-Definition Models (HDMs), without embedding them directly into the HBIM model as geometric data. Instead, HDMs are referenced within the IFC file through textual identifiers, which can be resolved at runtime to retrieve the corresponding external assets when required (e.g., during the VHE execution).

Each HDM is linked to a specific LAO defined in the HBIM model by means of a unique identifier (*ID*) stored within the corresponding LAO Pset. Within the HBIM environment, these LAOs are represented by simplified placeholder geometries, referred to as *dummy meshes*, whose primary purpose is to act as spatial and semantic anchors for the high-resolution assets.

All HDMs are stored externally in a dedicated local directory named “HDModels”, where each model is placed in a subdirectory associated with the corresponding HBIM object. Given that the developed VR platform supports multiple LODs for each asset, these subdirectories may include several mesh files representing different resolutions of the same object.

In the use case addressed in the present work, all HDMs were authored using the GLB file format, an open and widely adopted standard for 3D content distribution. This choice facilitates efficient runtime loading and straightforward integration of the HDMs within the VR platform, allowing each high-definition asset to be dynamically instantiated in place of its corresponding dummy mesh in the VE.

Deployment Phase

The deployment phase leverages the developed VR platform to generate an immersive and visually coherent VHE starting from the enriched IFC file and the associated external data. As discussed in the review presented in Section 3.1.2, several of the functionalities required for such a platform, such as data parametrization, semantic data retrieval, and configuration management, naturally align with BIM-based approaches, as both configuration parameters and cultural information are explicitly encoded within the HBIM model.

The VHE generated by the platform is conceived as a guided virtual visit. Throughout the experience, visitors receive both visual and auditory guidance to navigate between predefined spatial locations within the VE. At each of these locations, the platform supports the exploration and understanding of specific visual and interactive elements present in the scene (e.g., museum artworks), providing contextual information and narrative cues. The individual components of this process are described in detail in the following subsections.

VR Deployment Platform

The deployment platform was implemented using the Unity game engine (version 2021.3). Nevertheless, the proposed workflow is not inherently tied to a specific game engine, as any development environment capable of automatically parsing and processing the information encoded within the enriched IFC file could, in principle, be adopted.

Immersive VR support was implemented through the SteamVR library, which provides the core functionalities required for user interaction within the VE. Alternative solutions, such as OpenXR, offering comparable levels of functionality, could also have been employed. Support for non-immersive VR modalities is currently under development.

To mitigate potential cybersickness effects [36], teleportation was selected as the primary locomotion technique, given its widespread adoption in immersive VR applications within the CH domain [201]. In addition, continuous movement, also referred to as smooth or joystick-based locomotion, was implemented, as it represents the de facto standard interaction paradigm in commercial VR applications [38].

In the current implementation, all data are retrieved from local directories. However, the platform architecture has been conceived to allow the substitution of local storage with cloud-based data retrieval mechanisms. Furthermore, the platform was deployed as a PC-based VR application rather than as a standalone solution to allow higher graphical fidelity and stable frame rates.

From an operational perspective, the platform execution consists of multiple stages, including IFC data acquisition, HBIM model reconstruction, management of graphical features (e.g., materials, textures, lighting), rendering, and the initialization and control of the resulting VHE.

Data Gathering

Once the user selects the environment to be generated, all required data are automatically retrieved from a local directory named “IFCFiles”, which contains a dedicated subfolder for each available HBIM model. Each subfolder includes the corresponding IFC file, along with the “TextureData” and “HDModels” directories generated during the preparation phase.

After the selection, the user can trigger the VHE generation process, which is fully automated and structured into three main stages: HBIM Model Reconstruction, Graphics and Rendering Management, and Visit Initialization. A schematic overview of these stages is illustrated in Figure 3.1.

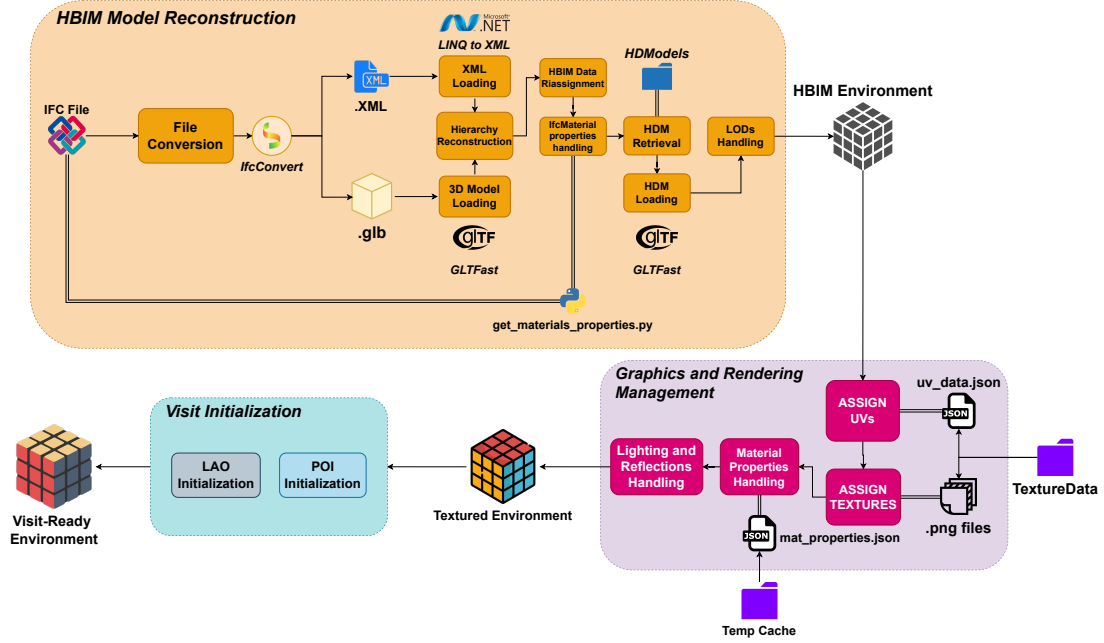


Figure 3.1: Schematic representation of the three automatic stages executed by the VR platform after the selection of the environment to be generated.

HBIM Model Reconstruction

The objective of this stage is to retrieve the HBIM-related data from the previously selected IFC file, together with the HDMs semantically associated with it, and to load them into the VE while managing the corresponding LODs.

As a first step, the platform automatically extracts both geometric and semantic information from the IFC file. This process exploits the capabilities of IfcConvert v.0.7.0, which supports the conversion of generic IFC files into multiple output formats [100]. In the proposed workflow, geometric data are converted into a GLB file, whereas semantic information is extracted and stored in an XML file. Upon completion of the conversion, both files are saved in a temporary local directory that acts as a cache.

Subsequently, the generated files are loaded into the Unity environment. In particular, the GLB model is imported into the active scene of the VR platform using the GLTFast v.6.7 module, which is natively available in Unity [212]. The XML file, instead, is parsed through LINQ to XML, a component of the Microsoft .NET framework that enables efficient manipulation of XML data structures [217]. Once both files are successfully loaded, the platform proceeds with the reconstruction of the object hierarchy and the reassignment of the HBIM semantic data.

After the BIM data are fully initialized and the hierarchy reconstruction is completed, the platform retrieves the HDMs associated with the HBIM model. At

the current stage, HDMs are loaded only for elements that are relevant for the virtual visit (e.g., decorations, statues, and other heritage-related objects), with the aim of enhancing their perceived cultural value through high visual fidelity. When available, multiple versions of each HDM with different levels of detail are loaded to enable LOD management within the virtual scene.

Upon completion of HDM retrieval, the HBIM Model Reconstruction stage concludes. The output of this phase consists of a fully reconstructed 3D representation of the environment.

Graphics and Rendering Management

Once the HBIM model has been reconstructed, the platform initiates the Graphics and Rendering Management stage. The goal of this phase is to achieve a realistic visual representation of the virtual environment by assigning UV coordinates and textures, configuring material parameters, and managing lighting and reflections.

The first operation concerns the assignment of UV coordinates to the meshes composing the HBIM model. UV data are retrieved from a local JSON file generated during the preparation phase and processed through a custom platform component. This component iterates over the meshes of the reconstructed model, matching their *GlobalId* values with those stored in the JSON file and reassigning the corresponding UV coordinates whenever a match is found.

The next step involves texture reassignment. Using the same local directory that contains the JSON file, a custom component retrieves the texture files associated with each mesh by analyzing their file naming conventions. In particular, it searches for matching *GlobalId* values within the “TextureData” folder. When a match is identified, the texture is loaded and assigned to the appropriate material. Each texture is further labeled according to its type, allowing it to be correctly mapped to the corresponding material property. For instance, a texture named “3dx9421_01_NORMAL” is associated with the normal map of the first material of the object whose *GlobalId* is “3dx9421”.

After textures have been applied, the platform assigns the material properties extracted from the IFC file during the HBIM Model Reconstruction phase. Specifically, values corresponding to parameters such as *Triplanar*, *Roughness*, *Metallic*, *Transparency*, *Tiling_X*, and *Tiling_Y* are set for each material, based on the information contained in the *Pset_Material_Rendering_Properties* property set defined within the HBIM model.

The final operation of this stage concerns the management of lighting and reflections. Lighting is handled through a combination of predefined and dynamically generated light sources. The virtual environment includes a directional light that acts as a sun-like illumination source, following a real-time approach to adapt to different spatial configurations. In addition, the platform generates a spotlight positioned in front of each LAO identified in the HBIM model. These real-time lights

are intended to guide the user’s attention toward relevant heritage elements and to emphasize their visual details.

With respect to reflections, the platform exploits the semantic structure of the HBIM model to identify all objects classified as *IfcWall*. A bounding box enclosing all *IfcWall* elements is then computed to define the spatial extent of reflection influence. Based on this bounding box, a Unity reflection probe component is added to the scene to enable reflective effects on materials that require them (e.g., mirrors or metallic surfaces). The position and dimensions of the reflection probe are automatically adjusted to match the center and size of the computed bounding box.

The output of this stage is a VE that is fully configured in terms of graphics and rendering, and ready for user interaction.

Visit Initialization

After completing the Graphics and Rendering Management stage, the VR platform proceeds with the initialization of the VHE. This phase is responsible for instantiating the virtual visit itself, as well as for configuring the components that enable user locomotion and interaction within the VE.

The initialization process revolves around two core entities: the POIs and the LAOs. As a first operation, the platform retrieves all POIs defined in the IFC model. Each POI is then augmented with a TP component provided by the SteamVR toolkit, enabling users to reach the corresponding location within the VE through teleportation. When this locomotion technique is employed, teleporting to a specific TP triggers the activation of the contents associated with that POI, such as the playback of audio narration or the visualization of UI panels containing textual and visual information.

When continuous movement (i.e., smooth locomotion) is enabled, the same triggering mechanism is activated through the physical interaction between the user and the collider associated with the TP. This strategy allows the platform to emphasize specific spatial locations within the VE, potentially improving user focus and perceptual engagement. At the same time, it may constrain the user’s freedom of movement.

To mitigate this limitation, the platform also supports the automatic generation of SteamVR Teleport Areas (TAs). These areas allow users to teleport freely to any position within predefined regions of the environment. TAs are generated by identifying all entities classified as *IfcFloor* in the HBIM model and using them to define navigable surfaces.

In parallel, the platform retrieves all LAOs and their associated properties. Each LAO is then assigned to its corresponding POI based on the information encoded in the *View_Point* parameter of the *Pset_LookAt* property set. This association enables the correct activation and presentation of cultural contents when a given

POI becomes active during the visit.

The outcome of this phase is a fully initialized environment, ready to support a structured and interactive virtual visit.

Visit Management

After the completion of all the previous phases, the VR platform presents the user with an initial preview of the reconstructed environment, which serves as an introduction to the VHE. In this preliminary stage, the environment is visualized as a reduced-scale 3D miniature of the real space, rendered using a neutral grey material. This visualization is intended to support the verification of the correct loading of the HBIM model and of all the associated external data. Additionally, the platform allows switching from a solid rendering mode to a wireframe visualization, enabling the inspection of internal elements, such as POIs, without occlusions caused by structural components (e.g., walls).

Once the generation process has been successfully validated, the user can initiate the virtual visit. At this point, the environment is displayed at full scale and the user is positioned at the first POI. The orchestration of the visit is managed by a custom component referred to as the *Visit Manager*, which is responsible for controlling the progression of the experience. In particular, the VM handles the activation of POIs and their associated TPs, as well as the enabling of the TAs available within the environment.

Throughout the experience, the VR user is guided by a synthetic voice generated through the ElevenLabs API, which performs the TTS conversion. The textual content used for speech synthesis is either predefined or dynamically retrieved from the semantic information encoded in the reconstructed HBIM model. The handling of TTS requests and the communication with the ElevenLabs API are delegated to a dedicated custom component, referred to as the *TTS Manager*.

At the beginning of the visit, the visit manager unlocks the first POI. When a POI becomes active, the following sequence of operations is executed (see Figure 3.2):

- A first TTS request is issued using the text contained in the *Description* parameter of the POI, providing a vocal introduction to the user about the current spatial location;
- After a short delay, a second TTS request announces the number of LAOs associated with the active POI;
- Depending on the value of the *Sequential* parameter defined for the POI, one of the following interaction strategies is adopted:
 - Sequential POI

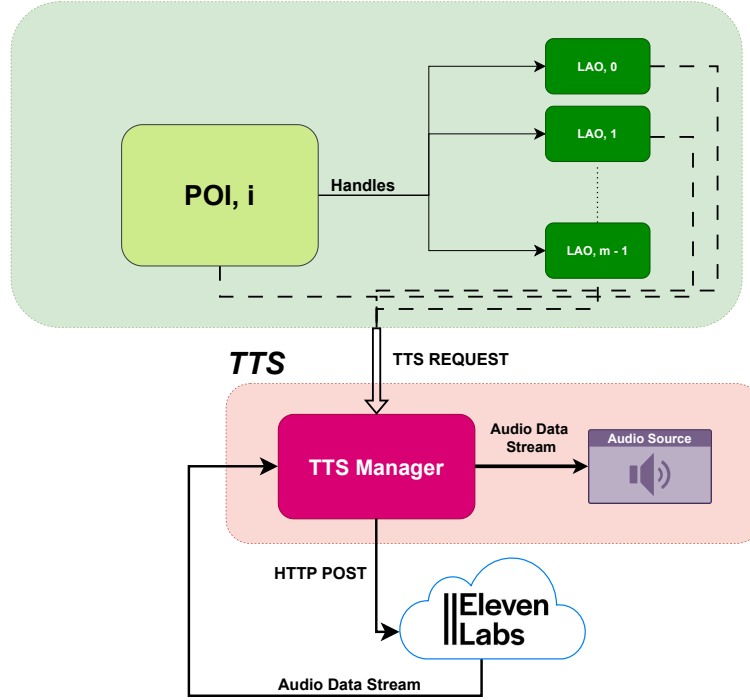


Figure 3.2: Schematic representation of the handling of POIs and LAOs. In this example, the i -th POI manages all the LAOs associated with it. As illustrated, both POIs and LAOs are capable of issuing TTS requests through the TTS Manager.

- The list of LAOs associated with the POI is ordered according to the *Order_Index* parameter defined for each LAO;
 - For each LAO, a UI panel is instantiated following the defined order, displaying relevant information such as title, descriptive text, and images;
 - In parallel, a TTS request is generated using the *Description* parameter of the LAO, providing a vocal explanation to the user;
 - Once the narration ends, a UI button becomes available, allowing the user to proceed to the next LAO in the sequence;
 - This procedure is iterated until all LAOs associated with the POI have been presented.
- Non-sequential POI
- The LAOs are not ordered according to any predefined sequence;
 - A UI panel corresponding to the first LAO is instantiated, and the user can navigate between LAOs through a dedicated UI button;
 - In this case, TTS requests are not automatically triggered, but are

instead activated by the user, allowing for self-paced exploration of the contents.

- Once the interaction with the current POI is completed, the subsequent POI is unlocked and a TTS notification informs the user that the next location is available.

This procedure is iteratively applied to each POI encountered by the user throughout the virtual visit.

UI Elements

As discussed in the previous subsections, user interaction within the VR environment is mediated through a set of UI elements, including UI buttons and UI panels (Figure 3.3a). These elements are designed to be interacted with via a laser-based pointing mechanism, which represents a widely adopted interaction technique in VR applications. This approach enables precise selection and activation of interface components while supporting a wide range of interaction paradigms within immersive environments.

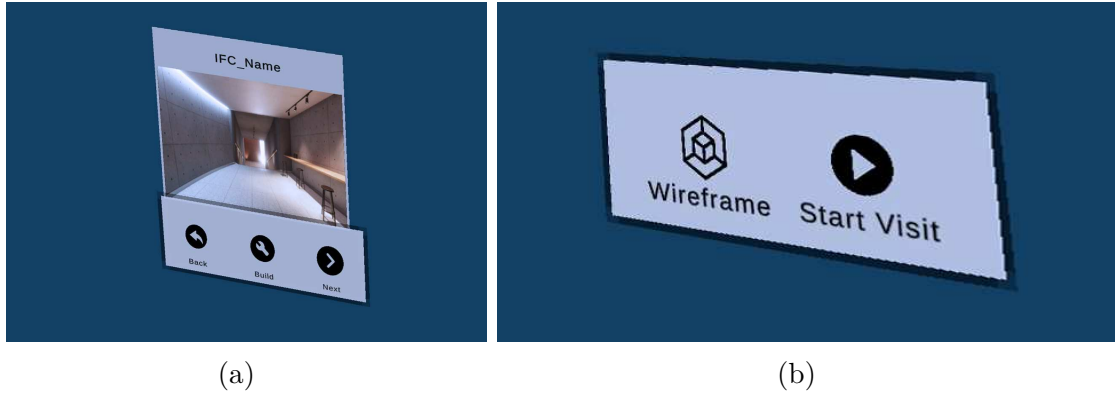


Figure 3.3: Examples of UI elements integrated within the application (main menu UI shown as reference): a) selection of the environment to be reconstructed and visualization of related information; b) option to make the 3D miniature of the environment transparent in order to display POIs or to initiate the virtual visit.

The cultural and informational contents associated with the visit are delivered to the VR user through UI panels that are instantiated at runtime whenever a LAO becomes active (Figure 3.4).

Each UI panel includes the following elements:

- A UI panel (denoted as a in the figure) equipped with a render texture component, displaying a video that visually indicates the position of the LAO



Figure 3.4: Example of a UI panel used to present a LAO within the VHE.

within the environment. The render texture shows the point of view of a virtual camera initially positioned at the center of the VE and framing the current POI; after a short interval, the camera shifts to frame the target LAO. This mechanism facilitates spatial understanding and allows the user to easily locate the LAO within the VE;

- Two textual UI elements (*b* and *c*) presenting the name and the descriptive text of the LAO, respectively;
- An additional UI panel (*d*) that dynamically loads and displays an image using the URI specified in the *ImageURI* parameter of the LAO;
- Three UI buttons (*e*, *f*, and *g*): one allows the user to dismiss the current LAO panel and proceed to the next available LAO; another triggers a TTS request to vocally reproduce the LAO description; the last one opens an additional UI panel that loads and plays a video using the URI stored in the *VideoURI* parameter of the LAO.

Overall, this UI configuration supports an informative and guided virtual visit, enabling users to explore the VE while seamlessly accessing and interacting with the associated cultural contents.

3.1.4 Use Case

To demonstrate the effectiveness and capabilities of the proposed workflow, it was applied to a use case focused on the creation of a virtual visit for a relevant

and iconic Italian CH site: Palazzo Carignano, located in Turin. The HBIM model of the entire palace was developed using Autodesk Revit and consisted of multiple federated models, coherently coordinated within a comprehensive reference model. However, not all the federated models composing the complete HBIM were considered for this investigation. The use case specifically focused on the “Hall of Seasons,” a room belonging to the Princely Apartments and a representative example of Piedmontese Baroque architecture [42] (Figure 3.5).

Notably, the same real-world environment selected for this use case was also employed in the investigation presented in Section 2.1. In the present one, however, the approach differs substantially: while the previous investigation relied on a fully manual design of both the environment and the experience, the current VE is generated from an enriched HBIM model, and the VX is automatically created within the VR platform.

Figure 3.5a shows a photograph of the Hall of Seasons, highlighting the decorative richness and architectural details typical of the Piedmontese Baroque style. Figure 3.5b, instead, illustrates the corresponding HBIM model reconstructed from survey data and historical documentation; in this representation, POIs and LAOs are explicitly highlighted in green and blue, respectively.

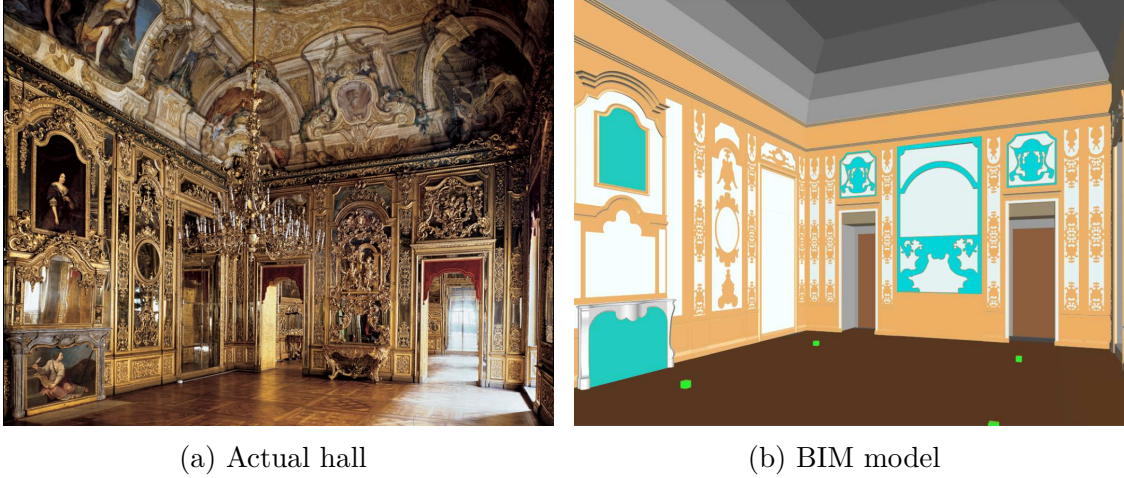


Figure 3.5: Picture of the real-world environment, i.e., the “Hall of Seasons” of Palazzo Carignano, in comparison to its HBIM model.

The proposed workflow was applied by using an IFC BIM file as input to generate a virtual visit of the hall, enabling the identification and management of the decorative elements typically included in a real guided visit. To this end, the HBIM model was enriched in Revit by introducing the parameters required to identify LAOs within the VE. Additionally, six POIs were placed within the hall to define the observer’s viewpoints in the exported IFC model.

The export of the enriched HBIM model in IFC format was performed through

the definition of a *Model View Definition* [33]. This element specified both the subset of the model to be exported and the Psets containing the custom parameters defined in Revit to be associated with the IFC entities. The export followed the IFC4 specification. In particular, the elements introduced to represent the observer's position within the hall were exported as *IfcBuildingElementProxy* entities of type POI and associated with a dedicated *Pset_POI*, containing the identifier and the additional parameters described in Section 3.1. The LAOs selected for the virtual visit included the frescoed vault, modeled as an *IfcCovering* of type Vault, as well as several wall decorations, exported as *IfcBuildingParts* of type Decoration.

The HDMs of the VE were defined from the meshes obtained through the reconstruction of the environment described in Section 2.1.4. Particularly, the meshes corresponding to the sculptures representing the four seasons (i.e., summer, fall, winter, spring) were considered as HDMs. The processing of the meshes was again performed using MeshLab version 2022.02, producing three LODs for each mesh. However, unlike the previous investigation which employed the FBX format, the generated meshes were imported into Blender and subsequently exported as GLB files. This operation was considered necessary, since MeshLab does not natively support the GLB format. Figure 3.6 presents an example of the LODs generated for one of the HDMs included in the use case.

To ensure a visually accurate and immersive representation of the environment in VR, the IFC model exported from Revit was further processed in Blender to define UV coordinates and texture mappings. This step involved assigning textures derived from photographs acquired during the on-site survey to the corresponding meshes. UV mapping was performed to correctly align textures with the 3D geometry, employing both manual UV unwrapping, where seams were explicitly defined to improve accuracy, and Blender's Smart UV Mapping tool to efficiently handle more complex geometries (Figure 3.7a and Figure 3.7b). Upon completion, texture files and UV data were exported locally following the proposed workflow, respectively as PNG images and a JSON file.

Finally, the prepared data were used by the VR platform to automatically generate the virtual visit. Figure 3.8 shows the miniature representation of the HBIM model generated by the platform, demonstrating its capability to correctly reconstruct the environment. This representation provides users with an initial overview of the space, supporting spatial orientation and allowing them to identify POIs before entering the full-scale VE.

Figure 3.9 illustrates the reconstructed VE during the virtual visit from multiple viewpoints. The results show that the platform is able to correctly manage graphical aspects (e.g., materials and textures) and rendering features (e.g., reflections), while also accurately identifying and handling the POIs defined in the HBIM model.

The POIs, positioned at floor level and visualized in blue, were equipped with TP components to enable user navigation between different viewpoints. Additionally, each POI features a dedicated and recognizable UI element to ensure clear

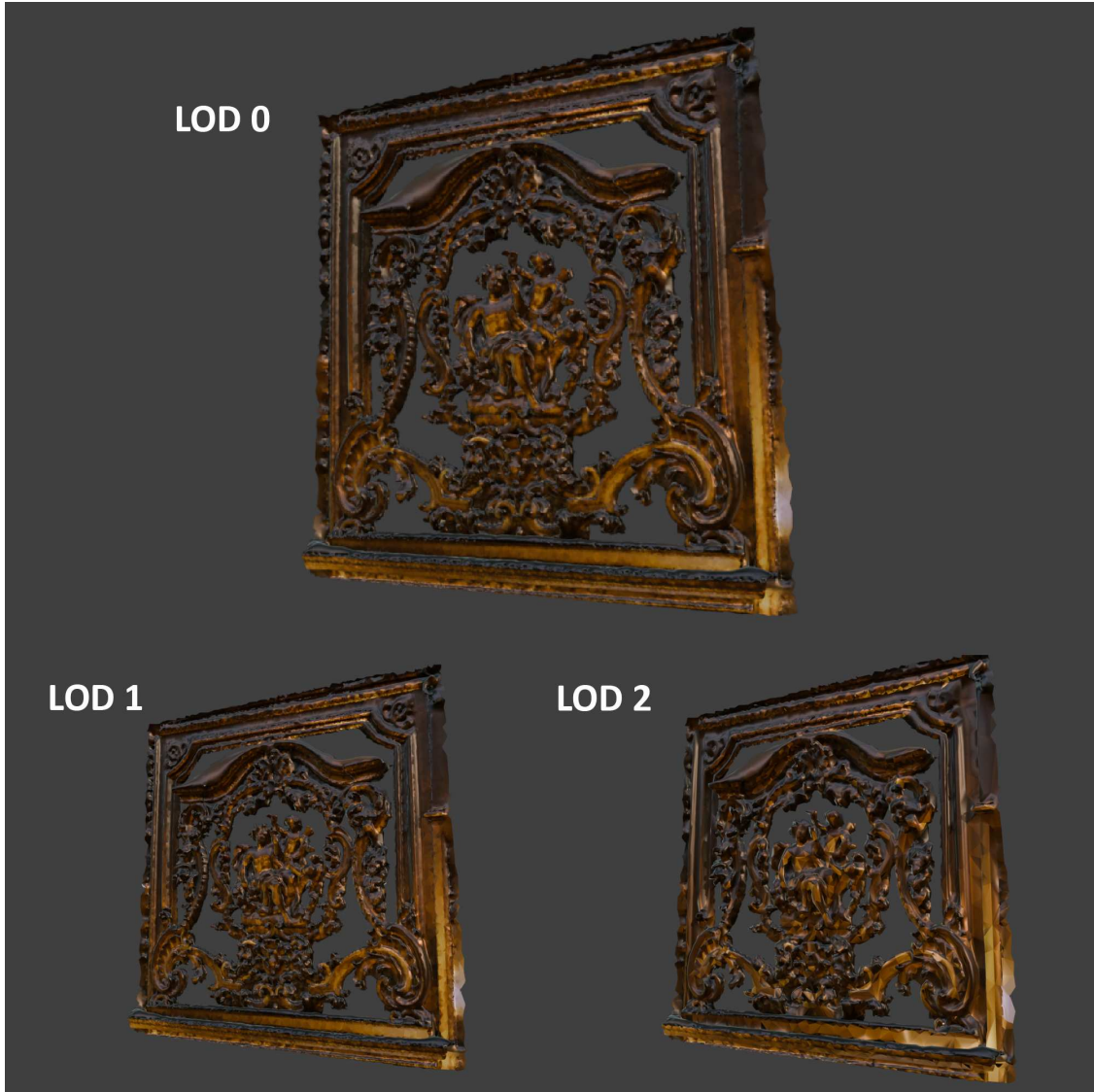


Figure 3.6: Picture of the three LODs for one of the meshes representing the “Fall” season. In this case, the number of vertices for the LODs is ~ 200 (LOD0), $\sim 80k$ (LOD1), and $\sim 15k$ (LOD2).

visibility and ease of interaction.

As shown in Figure 3.10a and Figure 3.10b, for each POI the platform activates the associated LAOs, which are dynamically highlighted to draw the user’s attention. Figure 3.10a depicts the first LAO associated with the initial POI, emphasized through a spotlight. Figure 3.10b, instead, presents the UI panel displayed whenever a LAO is activated, allowing the user to access textual descriptions, multimedia contents such as images and videos, and audio explanations generated in

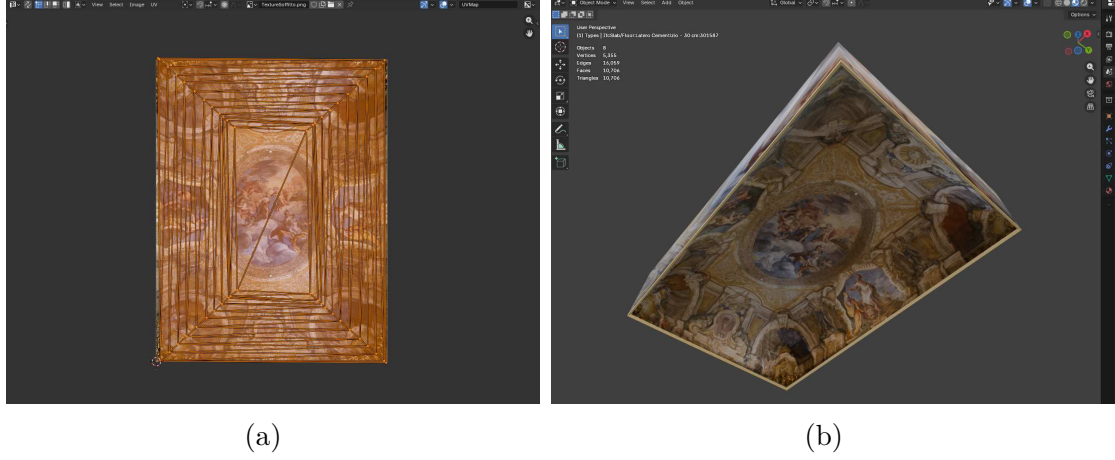


Figure 3.7: Frescoed vault texture UV mapping (a) and visual representation (b) in Blender.

real time through TTS.

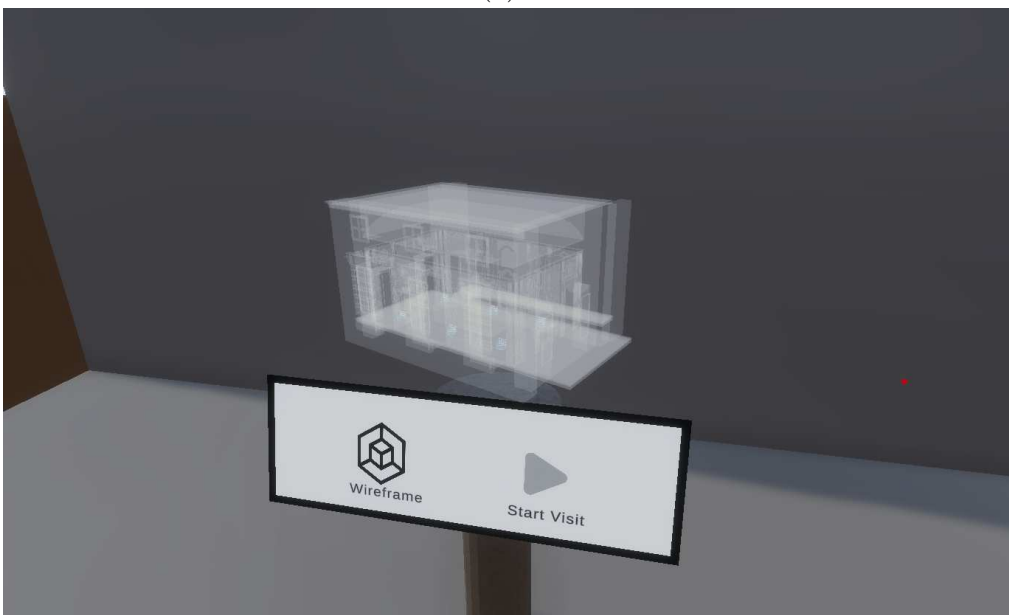
3.1.5 Results and Discussion

When compared with existing approaches in the literature, the proposed workflow exhibits several distinguishing characteristics. In contrast to many museum-integrated VHEs (Section 3.1.2), which require manual updates whenever geometric or semantic information changes, the adoption of HBIM models enables a more systematic and maintainable management of CH data. Additionally, the integration of a TTS-based audio guide allows auditory contents to be generated and updated dynamically, overcoming the limitations of static, pre-recorded narrations commonly used in existing solutions. These aspects contribute to the overall flexibility and extensibility of the workflow, facilitating the continuous evolution of both geometric and semantic contents. Despite a future comparative study and experiments should be conducted to ensure the full efficiency of the proposed approach, the devised workflow was able to reconstruct the VE of the selected use case (i.e., the “Hall of Seasons” of Palazzo Carignano) and generate a VHE (i.e., a virtual visit) starting from an enriched IFC file, thus positively answering the **RQ** formulated for this investigation.

At the same time, certain configurations explored in related work are not yet fully supported by the current implementation. For instance, some VHEs rely on 360° video to achieve highly realistic visual representations at the expense of interactivity, while others adopt detailed 3D environments that enable advanced interaction paradigms, such as the manipulation of virtual artifacts [178, 184]. Supporting these interaction modalities, as well as validating them with a user study, represents a natural direction for future development of the proposed workflow, with the



(a)



(b)

Figure 3.8: Miniature representation displayed at the end of the VE reconstruction displayed a) using a flat gray material, and b) in wireframe to enable the visualization of POIs.

objective of further broadening its applicability and strengthening its potential as



(a)



(b)



(c)

Figure 3.9: Representation of the state of the VE during the virtual visit: a) overview of the hall and of its features, both in terms of graphic and rendering (e.g. mirror reflections), as well as examples of automatic handling of textures and UVs for b) the frescoed vault and c) other meshes (e.g., wall paintings).

a general-purpose solution for the creation of immersive VHEs.

The proposed workflow currently imposes a set of requirements to enable the full reconstruction of a visually compelling environment starting from an HBIM model. In particular, textures and UV data are not directly embedded within the IFC file and therefore must be retrieved by the VR platform from external folders. This



(a)



(b)

Figure 3.10: a) The first LAO associated with the initial POI of the virtual tour, emphasized by the activation of a spotlight. In this instance, the LAO corresponds to the decorative element symbolizing the Summer season; b) UI panel shown for the first LAO linked to the initial POI of the virtual tour, presenting cultural information directly extracted from the HBIM model.

aspect may represent a limitation from the user’s perspective, as it requires additional manual processing steps carried out through external tools (e.g., Blender). However, this limitation largely reflects the current state of BIM platforms, which generally do not prioritize native support for advanced visual assets such as texture mappings and UV coordinates. As a consequence, BIM-to-VR workflows often remain dependent on external modeling software. A potential solution to this issue would involve extending the IFC schema to provide comprehensive support for textures and UV data, allowing their metadata to be directly integrated into the IFC

file itself.

Further considerations concern the management of lighting and reflections within the generated VEs. With respect to reflections, the current implementation computes them based on the presence of one or more *IfcWall* elements, an approach that may prove limiting in open or partially bounded environments. Regarding lighting, illumination is currently computed entirely at runtime, which imposes constraints on the number of active light sources due to performance considerations. A commonly adopted alternative would be to precompute lighting through a baking process, generating lightmaps that can be applied to the 3D models. However, light baking is not supported in Unity outside of the Editor environment, as it is a highly computationally intensive operation. Even if technically feasible at runtime, such a process would significantly increase application startup times, potentially by tens of minutes, making it impractical for the intended use case. Consequently, light baking would require either external preprocessing tools or an intermediate baking step within the Unity Editor. Nevertheless, the proposed platform was intentionally designed to operate entirely at runtime, relying on Unity’s real-time lighting system. While this choice may not be optimal from a performance standpoint, it provides several advantages: it avoids the need to re-bake lighting for every reconstructed environment, enables the management of dynamic VEs with varying lighting conditions, and simplifies the overall workflow. From an accessibility perspective, the reduction of preprocessing steps represents a relevant benefit.

Concerning the positioning of HDMs, the current implementation does not account for potential geometric clipping with adjacent meshes. As a result, visual artifacts may arise if the HDMs are not accurately positioned during the HBIM model creation phase. To mitigate this issue, future developments could focus on the integration of automated collision detection and positional adjustment mechanisms within the HDM placement process.

Despite these limitations, the proposed approach highlights an additional role for BIM models beyond traditional design and facility management tasks, emphasizing their potential for cultural dissemination and accessibility. By semi-automating the HBIM-to-VR pipeline and enriching IFC files with semantic metadata, the workflow substantially improves the scalability of VHE creation. This contributes not only to the understanding and appreciation of built heritage, but also to its digital enhancement and long-term preservation. The integration of BIM models into VR experiences opens new possibilities for educational and interactive applications within the CH domain.

From this perspective, future developments may include the simultaneous use of multiple HBIM models to illustrate the historical evolution of buildings across different time periods. While this scenario introduces additional challenges, such as ensuring consistent georeferencing and accurate spatial alignment, a well-structured

IFC configuration could facilitate the automation of these processes in the generation of VHEs. Further work could address the extension of the platform to support non-immersive VR experiences, including the management of currently unsupported elements such as screen-based UIs and mouse-driven interactions. Additional experiments involving diverse HBIM models could also be conducted to validate the generality of the proposed workflow across different CH contexts and types of VHEs. Moreover, new functionalities could be integrated, including virtual object manipulation, accessibility features tailored to users with physical or cognitive impairments, and advanced interaction mechanisms based on LLMs [211, 213]. It is also worth noting that, although the current implementation retrieves data from local folders, the platform has been explicitly designed to support future integration with cloud-based data retrieval systems, improving scalability and accessibility.

Finally, user studies involving different categories of participants could be conducted to collect feedback and further refine the usability and effectiveness of the workflow. Moreover, it could be useful to compare a non-immersive VR experience with an immersive one generated by the proposed workflow to evaluate the changes in terms of users' perception. This evaluation phase could possibly be essential to ensure that the generated VHEs are not only technically robust, but also engaging, accessible, and educational for a broad and diverse audience.

3.2 Closing Remarks

This chapter explored the use of ER technologies for the creation of VXs. Specifically, it described an investigation regarding the usage of HBIM to generate VHEs.

The presented investigation introduced a semi-automated HBIM-to-VR workflow aimed at supporting the generation of VHEs through the use of enriched IFC files and a dedicated VR deployment platform. The proposed approach was articulated into two complementary phases. The first phase focused on data preparation, where the IFC file was augmented with additional information stored externally, including texture data, UV coordinates, and HDMs. The second phase consisted of an automated deployment process, in which the developed VR platform reconstructed the environment at runtime and generated a structured virtual visit.

The workflow was applied to a real-world use case involving the creation of a virtual visit based on the HBIM model of the Hall of Seasons at Palazzo Carignano, as in the first investigation. The resulting VR application demonstrated the feasibility of leveraging IFC-based HBIM models to produce immersive and educational VHEs, highlighting the potential of the proposed workflow to improve both the scalability and accessibility of digital CH valorization. The generated experience incorporated core features commonly found in VHEs, including guided

navigation through teleportation mechanics, attention-directing mechanisms, multimedia cultural contents, and the integration of both environment-scale geometry and high-fidelity 3D assets. Furthermore, the platform supported advanced graphics and rendering functionalities, such as dynamic lighting and reflections, as well as LOD management to ensure real-time performance.

Despite the limitations and future enhancements described in Section 3.1.5, the workflow demonstrated the potential of integrating BIM-based DTTs within VXs, positively contributing to the **MRQ2** of this thesis. To further enhance the value of the carried out investigation, the proposed workflow could be possibly expanded to other domains beyond CH, employing BIM models of buildings for creating various types of VXs (e.g., educational platforms, social VR scenarios). In terms of effort and time invested to build the experience, the described VR platform enables the automation of several aspects in relation to the VE and the VX itself, such as the management of rendering features (e.g., textures, lights, reflections), meshes coordinates, and UI elements. Concurrently, it allows to ease the import of BIM models and their processing within a virtual application, which can be beneficial for both the VR and BIM domains. A future evolution of the research could aim to address the limitations emerged, possibly involving other DTTs in the process. For instance, the current workflow leverages external software applications to enrich the IFC file and managing the LODs of the most high-poly meshes (i.e., HDMs). Possible solutions could include the use of generative AI-based DTTs to automatically enrich the files, and manage the aspects related to the meshes. Furthermore, user studies involving experts from different domains could be conducted to further enhance the value of the proposed workflow.

Overall, the investigation described in this chapter highlighted how DTTs for ER can ease VXs creation through the automation of the workflows, partially addressing the effort and interoperability issues pointed out in Chapter 2 (Section 2.2). To tackle the challenges in relation to user's guidance, the next chapter is focused on exploring the convergence of another type of DTTs with VR, namely those based on generative AI. Specifically, the chapter faces the challenges of employing these technologies for the creation of conversational agents in VXs.

Chapter 4

Generative AI for Conversational Agents

The work described in this chapter has been formerly published in [16, 73].

Following the disruptive revolution brought with the introduction of ChatGPT in 2022, generative AI-based DTTs have attracted worldwide attention [75]. By definition, generative AI generally describes computational methods that can produce apparently new and meaningful content, such as text, images, or audio, based on the data they were trained on [70]. Taking advantage of these capabilities, this type of DTTs (e.g., LLMs) has been demonstrated to be useful across several industrial sectors and research fields, including education, advertising and entertainment [83]. This trend is still on-going and evolves along with the computational power of these technologies, continuously opening doors to new research opportunities.

Recently, the advancements in this kind of DTT have allowed the creation of new paradigms and methodologies within Vxs. One of the most useful applications of these technologies within this context is the creation of conversational agents, which often are defined as dialogue systems that can respond to the natural language of users [225]. In many Vxs, these agents are usually represented with a physical appearance (i.e., embodied), enabling the definition of ECAs [39]. While traditional approaches for creating these agents relied on conversational systems with several limitations (e.g., limited responses), the integration of LLMs has substantially enhanced their capabilities. Thus, ECAs have been used for various purposes in Vxs, ranging from providing simple user guidance within the VE, to simulating complex human-behavior in real-world situations. However, numerous challenges and research opportunities still remain. The evolution of these DTTs requires constantly novel methods to integrate ECAs within Vxs in several domains. Additionally, while text-generation has been widely explored in literature to create high quality talkative ECAs, other types of content generation (e.g., image-to-text) are mostly underexplored.

This chapter contains two exploratory studies regarding the integration of ECAs, empowered with generative AI-based DTTs, within VXs. The first investigation is placed in the context of UX in the CH domain. Specifically, it proposes guidelines for creating the VE and integrating an LLM-based ECA providing user’s guidance across multiple VMEs, with the specific objective of designing a unified and effective UX. Subsequently, the second investigation proposes an approach to integrate vision capabilities within traditional ECA architectures. Specifically, it describes a module leveraging VLMs to empower ECAs with the ability of perceiving the surrounding VE. The devised approach was validated through an exploratory user study in an existing VR game, comparing the perception of users towards agents with and without environment-aware capabilities.

4.1 Unified UX Design Across Multiple Virtual Museum Experiences

This section describes an investigation conducted in the context of UX across multiple VHEs employing LLM-based ECAs in combination with VR. The following subsections provide an overview of the topic, the related research within the context, the devised methodology, and a preliminary evaluation with users.

4.1.1 Overview

In recent years, the rapid advancement of VR technologies has opened new avenues for research and experimentation in the field of UX. The latter term broadly refers to the set of feelings, perceptions, and responses that arise from the use, or anticipated use, of a product, system, or service [154]. An extensively explored application domain in terms of UX is that of CH, where VR has enabled the development of experiences inspired by traditional museum practices. These experiences include virtual visits, virtual exhibitions, and virtual workshops, which are often grouped within the concept of “virtual museum”. One of the earliest formalizations of this idea dates back to the early 2000s, when Mase et al. [146] introduced the concept of the “Meta-Museum.” In their work, the Meta-Museum was defined as a space that “*blends VR and AI technologies with conventional museums to maximize the utilization of the museum’s knowledge base and to provide an interactive, exciting and educational experience for visitors.*” Key characteristics of this approach included the presence of personal artificial guidance agents and a strong emphasis on “*visitor-oriented and personalized access, experience, and learning of exhibitions.*”

Over time, this definition has been progressively extended through numerous studies focusing on the design and evaluation of meaningful learning experiences within virtual museums and related CH environments, with particular attention

to both educational outcomes and UX aspects [129, 150, 56]. More recently, the emergence of metaverse-oriented platforms has further increased the complexity of these VMEs by introducing shared virtual environments and social interaction dynamics. In parallel, the adoption of chatbots powered by LLMs, such as ChatGPT, has represented a significant paradigm shift in the development of ECAs. Unlike traditional rule-based or scripted approaches, these models enable the generation of responses dynamically, leveraging knowledge acquired during training.

Despite these advances, UX evaluation in VR-based cultural experiences is still frequently approached in a fragmented manner, with individual experiences being assessed in isolation. This represents a limitation, as a VME may encompass multiple sub-activities, such as guided visits or interactive workshops, each characterized by different interaction requirements and engagement patterns.

This investigation aims to explore UX within the domain of virtual museums by considering multiple VMEs as part of a unified framework. To this end, two distinct VMEs were developed: “Piazza Memorabilia”, implemented within a VR metaverse platform, and a Renaissance-era learning laboratory situated in a virtual workshop. Drawing on guidelines identified in the state of the art, this preliminary investigation focused on two core elements as foundational components for a coherent UX across VMEs: the VE and the ECA. The VE was designed to foster user engagement and curiosity through exploration, while maintaining conceptual and experiential coherence between the two environments. The ECA, on the other hand, was implemented by integrating a LLM (GPT-4) into the system, with the goal of enabling effective and natural conversational interaction, as well as providing guidance to users throughout the experience. Finally, a preliminary user study was conducted by evaluating the usability of the experience, as well as the users’ perception towards the developed ECA.

4.1.2 Background

This section reviews the current related works in the field of UX in museums and methodological approaches in the context of VMEs.

UX in museums

A substantial body of research has investigated the nature of museum experiences. Among the most influential contributions in this field are Falk and Dierking’s *Museum Experience* framework [65] and Packer and Ballantyne’s visitor experience model [168]. These works identify several factors that shape the overall museum experience, including the visitor’s personal context, the immediate experience, and the remembered experience. While early studies primarily focused on the visitor’s individual dimension, King et al. [115] introduced an additional perspective by emphasizing the relationship between the subjective visitor experience and the actions

undertaken by the museum, gallery, or heritage site hosting the visit. From this standpoint, the concept of an “environmental response” emerges as a key element to be considered in the design of cultural experiences.

Building on the foundations laid by Falk and Dierking [65], Chang et al. [43] proposed a set of guidelines for designing engaging museum experiences. Among these recommendations, it was argued that museum experiences should be designed to “*offer open and free choices and put learners in control of their own learning*,” while still accounting for the visitor’s personal context. At the same time, they highlighted the importance of addressing the social dimension of museums, encouraging interactions among visitors as well as between visitors and museum staff to support the learning process. Furthermore, they suggested that museum educators should create opportunities to integrate knowledge and information through digital tools and facilitators, and to “*design learning environments where visitors can navigate more spontaneously from one experience to the next*.”

The principles identified in these studies are highly relevant for the present investigation, as they are not only compatible with the affordances of VR technologies and metaverse platforms, but also aligned with the objective of connecting multiple museum sub-activities into a coherent and unified VME.

UX in Virtual Museums

With respect to VMEs, a central aspect to consider is the design of the VE in which users are immersed. As discussed by Chung et al. [49], it is crucial to carefully balance the creation of environments that faithfully resemble real-world spaces with the possibility of designing unrealistic or imaginative environments that cannot be experienced in physical settings. Their work investigated the impact of realistic (i.e., accurately representing the physical world, high fidelity) and unrealistic (i.e., surreal environments, low fidelity) VEs on UX, showing that both approaches influence the overall experience in different ways. While realistic representations provide users with a sense of familiarity and spatial coherence, unrealistic VEs tend to stimulate curiosity and foster exploratory behavior within the virtual space.

Another relevant line of research concerns the enhancement of user engagement and learning through the gamification of virtual museum experiences. For instance, Pietroszek et al. [175] developed a Unity-based prototype for a virtual exhibit of the Museum of Peace Corps Experiences. The results of a pilot user study indicated positive feedback in terms of engagement, suggesting that game-like mechanics can effectively support user involvement in VMEs.

In order to establish a coherent UX that integrates the aforementioned elements, the inclusion of ECAs for user guidance is widely regarded as a fundamental component. A recent study by Trichopoulos et al. [211] explored the integration of LLMs within an ECA designed to guide users inside a museum environment. Their proposed architecture included features such as TTS, Speech-To-Text (STT), and

GPT-4, enabling multimodal interaction. In this approach, the ECA also acted as the primary content provider, as the LLM was pre-loaded with predefined responses related to specific topics. On the one hand, this strategy may be seen as limiting, since it primarily exploits the Natural Language Processing (NLP) capabilities of the LLM without fully leveraging its generative potential. Such an approach is often preferred when using traditional chatbot frameworks (e.g., RASA or DialogFlow), which rely on predefined responses and offer greater control over conversational flows. On the other hand, traditional chatbots are inherently constrained in flexibility, as highlighted in several studies [62, 92], and their development process can be relatively time-consuming.

4.1.3 Materials and Methods

The investigation was conducted within the context of the project “CTE - Musei nel Metaverso” promoted by LINKS Foundation, in collaboration with Casa delle Tecnologie Emergenti (CTE NEXT) and Palazzo Madama, a historic building and museum located in Turin, Italy.

The investigation involved the development of two distinct VMEs using the Unity¹ game engine. The first is a meta-museum experience hosted within a metaverse environment and developed for the Spatial platform through its dedicated Software Development Kit (SDK). The second is a virtual workshop specifically designed for Meta Quest HDMs. The 3D assets for both VEs were modeled using Blender, while 2D assets and textures were produced with Adobe Illustrator and Substance 3D Painter. Additionally, several tools were employed for the creation of the ECA, including the OpenAI API and various TTS and STT modules.

Meta-museum experience

The meta-museum experience, accessible both via desktop devices and HDMs, offers users a virtual visit to an imaginary square named “Piazza Memorabilia” (Fig. 4.1a), inspired by the Renaissance concept of the ideal city. The experience begins with the 3D reconstruction of an existing real-world room located within Palazzo Madama, where users are introduced to Virgilio, an ECA designed to guide them throughout the VME.

Subsequently, a time machine appears inside the room, allowing users to travel back in time and reach “Piazza Memorabilia”. Upon arrival in the square, Virgilio assigns the users a task consisting in finding and collecting five historical medals scattered across the environment. At this stage, users are free to explore the virtual space autonomously in order to locate the assigned items.

¹Unity Engine: <https://unity.com/>

Whenever a medal is found, the historical character represented on it materializes as a 3D model rendered with a spectral-like shader and narrates its own story. The information is conveyed both visually, through a textual UI, and audibly, through the audio output of the user's device. After all the medals have been collected, users are encouraged to continue the VME by accessing the Renaissance virtual workshop. The VE also includes a showroom area, where the historical medals created by users during the workshop are exhibited.

Virtual workshop experience

In the virtual workshop experience (Fig. 4.1b), users are invited to create a customized historical medal by leveraging hand-tracking features provided by the Oculus VR Toolkit. At any time, users can alternatively switch to traditional controllers according to their preferences. Upon launching the application, a dedicated menu allows users to choose between accessing a tutorial scene or directly entering the virtual workshop experience.

Within the tutorial, users are introduced to the fundamental interaction mechanics required to navigate and operate within the VE. These include teleportation, object manipulation (e.g., grabbing), and interaction with UI elements. The corresponding hand gestures and interactions are illustrated through UI panels and representative 3D models. Additionally, in the case the integrated instructions resulted as insufficient, users can request real-time assistance vocally from the ECA.

The workshop experience is structured around four sequential working stations:

- Customization of the historical medal;
- Creation of the mold using a wax model;
- Selection and melting of the metal for the casting process;
- Refinement and finishing of the final medal.

Each working station is organized in steps and associated with a predefined time constraint for each one of them. In cases where users exceed the allotted time, the system automatically completes the required actions to complete the step through predefined animations. This design choice aims to support users during the activity while ensuring a consistent overall duration of the experience.

UX Design

The unified UX that guides users seamlessly from the meta-museum experience to the virtual workshop is grounded in two core design elements: the VE and the ECA.

The VE acts as a connective narrative and spatial framework between the two experiences, relying on shared semantic characteristics and a coherent visual style.



(a)



(b)

Figure 4.1: Visual representation of the two VEs: the meta-museum experience (a) and the virtual workshop (b).

The overall experience immerses users in a historical and educational journey in which an idealized Renaissance square initially serves as the setting for the exploration and interpretation of historical medals, contextualizing them within their respective cultural and temporal backgrounds.

The metaverse experience concludes at the entrance of what corresponds, in practice, to the VE of the virtual workshop. This transition is considered a fundamental design choice, as it functions not only as a practical gateway to the workshop experience while preserving environmental continuity, but also as a narrative bridge that reinforces the coherence of the overall VME.

The idealized Renaissance VE deliberately avoids assigning a specific historical or geographical connotation to the medals, which originate from diverse regions. At

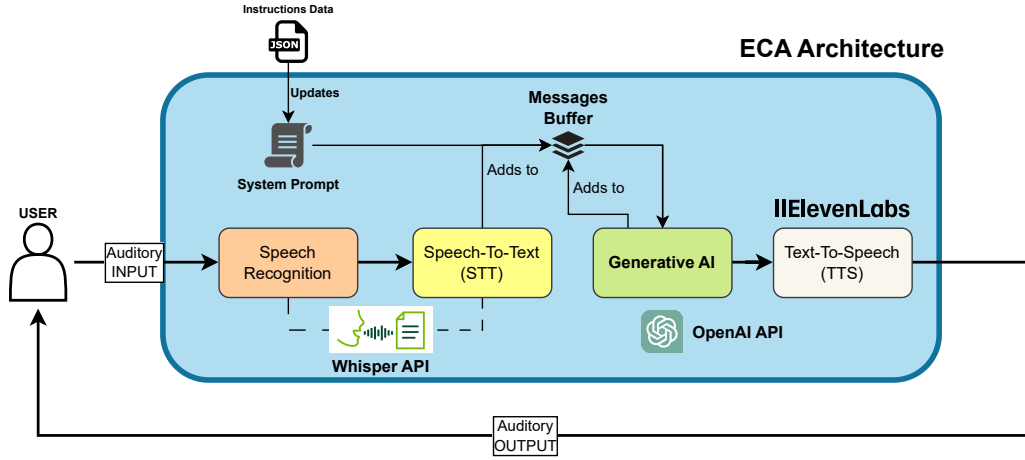


Figure 4.2: Schematic representation of the ECA architecture;

the same time, it enables users to acquire architectural notions characteristic of the Renaissance period. Moreover, in line with the analysis of the state of the art, both VR applications adopt a cartoon-style visual aesthetic, aimed at gently stimulating users' curiosity and fostering engagement within a familiar and non-intimidating VE.

The ECA, named “Virgilio”, was designed to enable seamless interaction with users through voice-based communication. Throughout the experience, the ECA is responsible for two primary functions: supporting users with gameplay-related guidance (e.g., teleportation and object manipulation) and delivering instructions and information related to the cultural contents of the VMEs. Interaction with the ECA is initiated by the users through the pronunciation of a predefined activation keyword (i.e., calling the agent by name), after which they can pose questions related to the task currently being performed, which are processed by the underlying LLM-based system.

As shown in Fig. 4.2, the system architecture is composed of four main components. First, user speech is captured and processed by AI-based Speech Recognition (SR) modules, relying on the Whisper API to convert vocal input into text with high accuracy. The resulting transcription is then appended to a conversational buffer that stores the interaction history between the user and the ECA. Subsequently, the complete buffer is forwarded to a Generative AI module, implemented through the OpenAI API for Unity, which enables the generation of coherent and context-aware responses to user queries.

For the virtual workshop, response generation is driven by carefully crafted prompts, designed according to the guidelines provided by OpenAI [166]. These prompts define the behavioral constraints and instructional role of the ECA throughout the experience. More specifically, the ECA is dynamically prompted by the

system using a set of instructions retrieved from a JSON file, which are selected based on the current workstation and the specific step being performed by the user. Each instruction is composed of the following fields:

- *Workstation ID*: an integer value identifying the current workstation;
- *Step ID*: an integer value indicating the current step within the workstation;
- *Guideline*: a textual description of the instructions required to complete the step;
- *Rules*: a textual field defining constraints that the ECA must adhere to when generating responses (e.g., “Do not mention that you are an AI”);
- *Environment*: a textual description containing contextual information about the surrounding VE.

Finally, the architecture integrates the ElevenLabs API ² to provide high-quality TTS output. This overall design supports immersive, responsive, and context-aware interactions between users and the ECA, ensuring that guidance and information are consistently aligned with the state of the VR experience.

Experiments

A preliminary evaluation on usability and users’ perception of the ECA was conducted during a three days experimental phase at Palazzo Madama, involving LINKS Foundation, CTE NEXT, and 53 high-school students from three different institutions located in Turin, Italy.

A series of portable PCs and Meta Quest HMDs were employed, respectively for allowing participants to test the metaverse experience and the virtual workshop. Each session lasted approximately 50 minutes, and included an introduction to the experience (≈ 10 min.), the testing of the metaverse experience (≈ 15 min.) and the virtual workshop (≈ 15 min.), as well as the anonymous compilation of two questionnaires (≈ 10 min.).

The evaluation on usability was performed by employing the SUS questionnaire [28], as in many other investigations within the state-of-the-art, which provided a subjective measure of usability. The 10 items of the questionnaire were rated on a 5-point Likert-type scale, with 1 representing “Strongly Disagree” and 5 for “Strongly Agree”. In turn, the ECA present within the experience was evaluated in terms of users’ perception considering various dimensions, namely Hedonic Motivation (HM) (i.e., the pleasure derived from interacting with the ECA), Perceived

²ElevenLabs API: <https://elevenlabs.io/api>

Humanness (PH) (i.e., the degree to which the ECA has human-like traits, Performance Expectancy (PE) (i.e., the extent to which users believe that the ECA will help them in completing a task), and Cognitive Attitudes (CA) (i.e., users' perception of how effectively the ECA can provide information). All these dimensions were selected from the approach proposed by Ma et al.[142] and adapted for this considered user study. The items for each dimension of the questionnaire were rated on a 7-point Likert-type scale, with 1 representing "Strongly Disagree" and 7 for "Strongly Agree".

4.1.4 Results and Discussion

In terms of usability, the results derived from the first questionnaire showed a mean SUS score equal to "Good" ($M = 72.97$, $SD = 14.75$), in accordance with the rating rationale developed by Bangor et al. [14]. The SUS score was computed in accordance with its original formulation [28]. Specifically, the scores for the odd-numbered items were calculated by subtracting 1 from the scale position, while the scores for the even-numbered items were obtained by subtracting the scale position from 5. The overall SUS score was then derived by multiplying the resulting sum by 2.5. From an individual standpoint, among the 53 users involved, seven scored equal to "Excellent" (min = 87.50, max = 95.00); twenty-two scored equal to "Good" (min = 75.00; max = 85.00); twenty scored equal to "OK" (min = 52.50, max = 72.50); and four scored equal to "Poor" (min = 37.50; max = 50.00). Since usability is strictly related to UX, this result may indicate a slightly positive trend towards an effective and usable VX, while at the same time highlights the need of future enhancements for the experience.

Regarding users' perception of the ECA, results from the second questionnaire (Figure 4.3) showed aggregated ratings between 4 and 5, with the exception of one dimension (PH) which was below 4. The highest rated dimension was CA ($M = 4.55$, $SD = 1.70$), while the lowest was PH ($M = 3.94$, $SD = 1.45$).

Although preliminary and based on a limited analysis in terms of sample size, these results partially suggest a slightly positive trend towards an effective UX across multiple experiences. To improve usability, future developments could involve a deeper analysis of the modalities to enhance UX, as well as expanding and diversifying the sample of the users involved. Regarding the ECA, all the users were able to complete the experience following the guidance of the devised agent. However, results did not highlight particularly low or high values for the evaluated dimensions. Interestingly, most of the users evaluated usability positively, while the agent was generally perceived as neutral to them. This consideration could suggest that the selected scenario for the developed VXs was not entirely suitable to fairly evaluate the ECA. This result may be explained by the fact that users were more focused on completing the tasks within the VR experience than on interacting with the ECA. Future research in this directions should involve additional testing in

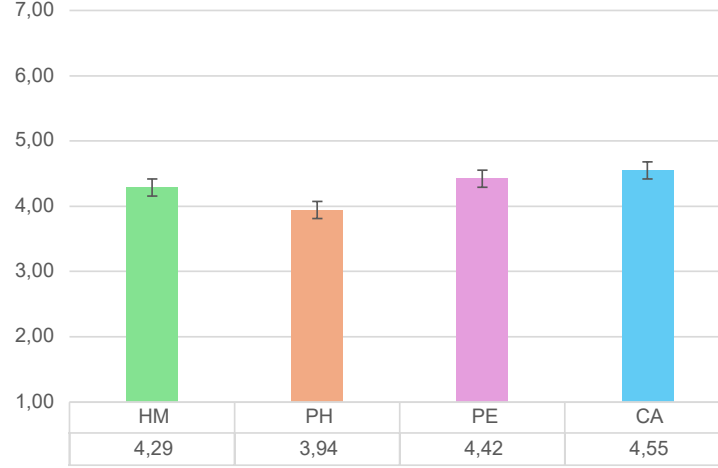


Figure 4.3: Preliminary data for the users' perception of the ECA

different use cases, as well as a statistical analysis to confirm the obtained results. Additionally, it could be interesting to perform a comparative evaluation with traditional approaches for the design of ECAs. The latter aspect could possibly be fundamental for asserting whether or not LLM-based ECAs are more suitable than traditional ones (e.g., RASA) within VMEs.

4.2 Vision-Language Models for Environment Aware Agents

This section describes an investigation carried out in the context of environment-aware ECAs employing VLMs in combination with VR. The following subsections provide an overview of the topic, the related research within the context, the devised methodology and its application in an existing VX (i.e., a videogame), and an exploratory user study.

4.2.1 Overview

The rapid growth of generative AI-based DTTs has enabled researchers and developers worldwide to significantly enhance interactive VXs. In particular, the integration of LLMs, such as GPT-4 [29] and Llama [210], within VEs has recently emerged as a highly influential factor [206]. The growing adoption of LLMs can be attributed to the numerous advantages offered by these models, including their ability to understand and generate human-like language, perform reasoning-based

tasks that traditionally require human intelligence, as well as process and generate multimodal content (e.g., images, videos, and speech) [105].

With respect to the use of generative AI for the creation of ECAs, several challenges remain open. A critical issue concerns the ability of such agents to effectively perceive and reason about their surrounding VE. In traditional gaming scenarios, ECAs have often lacked awareness of the environment, which may negatively affect immersion and the overall player experience [103, 181]. When LLMs are employed, the integration of accurate environmental knowledge into the agents' responses becomes more crucial, as it directly influences the relevance and coherence of their interactions [180]. Moreover, the degree of environmental awareness exhibited by an LLM-ECA has been shown to significantly impact conversational quality and human-agent interaction [224].

To tackle these challenges, existing approaches typically rely on extracting environmental information directly from the development platform or game engine (e.g., Unity or Unreal Engine). Such solutions are commonly based on custom pipelines [131, 132] or on Scene Graph (SG) representations [227, 128] to structure environmental knowledge. While these methods allow LLM-ECAs to access precise contextual information, they generally require developers to manually define and integrate object descriptions, spatial relationships, and visual attributes within the development environment. For large-scale VEs, this process can become extremely time-consuming and labour-intensive.

An alternative solution to mitigate these limitations is represented by VLMs. By jointly leveraging computer vision and NLP, VLMs have proven effective for deployment within interactive virtual applications [112, 134]. A key advantage of VLMs lies in their capability to directly extract contextual information from visual inputs, such as images or video streams, and transform it into structured textual representations [153]. Nevertheless, the use of VLMs to provide environmental knowledge to LLM-powered ECAs in gaming and VR scenarios, as well as their impact on users' perception and conversational experience, remains largely underexplored.

Building upon these considerations, this investigation explores the use of VLMs to enhance conversational LLM-ECAs with accurate and contextually grounded environmental information, and examines how environment-aware ECAs influence players' perception within a VX. To address the former objective, a software module was developed that integrates a VLM into an existing virtual application through multi-image inference. By means of prompt engineering, visual inputs are converted into a structured SG representation of the VE, enabling LLM-ECAs to perceive and reason about their surroundings. To address the latter objective, an exploratory within-subject user study was conducted using a state-of-the-art video game populated with multiple LLM-powered conversational NPCs. These ECAs were augmented with the proposed VLM-based module to ensure environmental awareness. The evaluation focused on players' perception of the ECAs with

and without environment-aware capabilities, considering the following dimensions: perceived knowledge and specific environmental knowledge (i.e., degree to which players perceive the ECAs as knowledgeable about the environment and its specific elements), perceived intelligence (i.e., degree to which the players perceive the ECAs as intelligent entities), environment factuality (i.e., the extent to which players perceive the agents' responses to be truthful), willingness for future interaction (i.e., players' willingness to interact again with the ECAs in the future), and conversational capabilities (i.e., the perceived communication abilities of the ECAs, such as flexibility and human-likeness). A within-subject design was preferred over a between-subject approach to facilitate the comparison of individual users' perceptions of the two different ECA configurations.

By defining environmental awareness as the degree to which an ECA is able to recognize the elements present within a given virtual scenario (e.g., objects, figures) and their characteristics (e.g., materials, colors), the objectives of this investigation are summarized by the following research questions:

- **RQ1:** How does environment knowledge enabled by a VLM affect players' perception of an LLM-ECA in terms of perceived knowledge and specific knowledge of the environment?
- **RQ2:** How does environment knowledge enabled by a VLM affect players' perception of an LLM-ECA in terms of perceived intelligence, environment factuality, and willingness for future interaction?
- **RQ3:** How does environment knowledge enabled by a VLM affect players' perception of an LLM-ECA in terms of conversational capabilities, namely communicative ability, flexibility, and human-likeness?

4.2.2 Background

In recent years, a growing body of research has focused on the development and enhancement of agents based on LLMs. This section reviews the methodologies, tools, and approaches proposed in the state of the art to improve LLM-based agents, including ECAs, and their interaction capabilities. Particular emphasis is placed on approaches aimed at endowing such agents with perceptual abilities, such as vision, as well as on VLMs and their application within VEs.

LLM-based Agents

With the disruptive emergence of LLMs, research interest has increasingly shifted toward the design and deployment of LLM-based agents operating within VEs. Depending on the target application and the interaction paradigm, these agents can generally be categorized into three groups: single-agent systems, multi-agent systems, and Human-Agent Interaction (HAI) scenarios.

Single-agent studies typically investigate how a single LLM-based agent can autonomously perform tasks within a system. Representative examples include the evaluation of player performance in serious games [102] and decision-making processes while autonomously playing video games [106]. Multi-agent approaches, on the other hand, focus on interactions among multiple agents, often aiming to simulate complex, human-like social behaviors. Notable contributions in this area include the works of Park et al. [172] and Lin et al. [133], which explore emergent behaviors arising from agent-to-agent interactions.

In contrast to the previous categories, studies in the HAI domain explicitly involve direct interaction between LLM-based agents and human users. While single-agent and multi-agent systems often emphasize internal capabilities such as memory, reasoning, or personality modeling, the presence of a human user introduces additional design constraints related to usability, responsiveness, and perceived intelligence. For example, Pan et al. [170] proposed ELLMA-T, an ECA designed for language learning in social VR environments. The agent engages users in meaningful conversations by leveraging an LLM (GPT-4), combined with STT for processing user input and TTS for delivering spoken responses. Similarly, Christiansen et al. [48] integrated multiple ECAs as conversational NPCs within a VR video game, enabling immersive role-playing interactions through both speech-based and dialogue-driven communication. Another relevant contribution is provided by Yang et al. [224], who adopted a comparable ECA architecture and employed prompt engineering techniques to investigate how different levels of agent knowledge influence users' perception.

In summary, these studies establish a solid foundation for the development of engaging and knowledgeable ECAs. However, recent trends in HAI research increasingly highlight the necessity of extending existing architectures with additional capabilities, not only to improve agents' conversational and knowledge-related skills, but also to enhance their awareness of the surrounding context in which interactions take place.

Environment-aware ECAs

The architecture of ECAs typically relies on a set of core modules responsible for specific functionalities. This modular structure is clearly articulated in the work of Xi et al. [223], who conceptualize LLM-based agents as composed of three primary components: perception, brain, and action. Within this framework, the perception module plays a crucial role by enabling agents to sense and interpret stimuli from their surroundings, including auditory and visual cues.

In the context of ECAs operating within VEs, several approaches have been proposed to provide agents with environmental awareness. One common strategy consists of supplying contextual information through in-game events. This is often

implemented via dedicated game objects, such as trigger volumes, which dynamically inject contextual data into the agent’s prompt or internal state [185]. An example of environment information extraction from the game system is presented by Li et al. [131], who focused on role-play scenarios. By leveraging an LLM (GPT-4) and carefully structured prompts, their approach enabled ECAs to access reliable environmental information within the VE. A similar methodology was adopted by Li et al. [128], who introduced a world-state module within the prompt engineering pipeline. This module was dynamically updated using a SG representation of the environment extracted from the underlying game system.

Across these approaches, textual representations emerge as the predominant means for enabling environmental awareness in ECAs. However, since such information is typically derived from the game system, visual properties of the environment (e.g., geometry, materials, spatial relationships) must be explicitly predefined and manually integrated. This process can be costly in terms of development effort and often requires re-implementation for each new application or virtual scenario.

Given the multimodal capabilities of contemporary LLMs, an alternative and promising solution is the adoption of VLMs. These models have already demonstrated their effectiveness in a variety of scenarios involving VEs. For instance, Hebri et al. [94] introduced SmartSage, an ECA designed for cognitive assessment in immersive VR. In this system, VLMs were employed alongside LLMs to interpret contextual information directly from the VE, enabling the agent to reason about its surroundings. Similarly, Konenkov et al. [119] proposed a VLM-based agent for medical training applications. Through a fine-tuning strategy, the VLM inferred environmental information from images captured from multiple viewpoints within the VE, demonstrating the feasibility of providing agents with contextual awareness while minimizing manual annotation efforts.

Despite the advantages offered by VLM-based approaches, their application within VEs remains relatively underexplored. Moreover, to the best of the author’s knowledge, no prior study has systematically investigated user perception of environment-aware ECAs empowered through VLMs. As discussed in the previous subsection, this aspect represents a critical dimension when designing ECAs intended for meaningful and engaging human-agent interactions.

An overview of the analyzed works is reported in Table 4.1.

4.2.3 Materials and Methods

This section introduces the proposed approach aimed at empowering ECAs with environmental perception capabilities. In particular, it describes the integration of an Environment Perception Module (EPM) within a conversational LLM-ECA architecture. The EPM leverages a VLM to extract a structured SG representation of the 3D VE through a multi-image inference process. Furthermore, this section

Study	Platform	Type	Tools	Capabilities	EK
[102]	ChatGPT	Single-Agent	GPT-3.5	MM	-
[106]	NetHack	Single-Agent	GPT-4-Turbo	MM AP	GSE
[172]	Phaser	Multi-Agent	GPT-3.5-turbo	MM PSLTY AP	GSE
[170]	VR-Chat	HAI	GPT-4 Whisper-STT OpenAI-TTS	MM BEX	-
[48]	Unity	HAI	GPT-4 Oculus SDK-STT Wit.ai-TTS	MM PSLTY BEX	-
[224]	Unity	HAI	GPT-4 Azure-STT Azure-TTS	MM	-
[131]	Unity	HAI	GPT-4-turbo STT Wit.ai-TTS	MM BEX FEX	GSE
[128]	Unity	HAI	GPT-4-turbo	MM PSLTY AP BEX	GSE
[94]	Unity	HAI	Phi3 PaliGemma	MM UG TM	VLMG
[119]	Unity	HAI	Qwen-VL Whisper-STT JETS-TTS	MM UG	VLMG

Table 4.1: Detailed list of current agents employed in the state-of-the-art. The not yet defined acronyms are reported as follows: EK, Environmental Knowledge; GSE, Game System Extracted; VLMG, VLM-Generated; MM, Memory Management; AP, Action Planning; PSLTY, Personality; BEX, Body Expressions; FEX, Facial Expressions; UG, User Guidance; TM, Task Management;

details the prompt-engineering strategies adopted to obtain such a structured representation in a reliable and reproducible manner.

System Architecture

As highlighted in the state of the art, the design of conversational LLM-based ECAs typically relies on modular architectures, where distinct components are responsible for specific functionalities [16][229]. To enable natural interaction between players and the system, these architectures commonly include SR and STT modules, which convert spoken input into textual form. The resulting text is then forwarded to a conversational module, usually composed of an LLM, a system prompt containing instructions and contextual information, and a memory buffer that preserves the dialogue history. The textual output generated by the LLM is subsequently transformed into audio through a TTS module, allowing the ECA to deliver spoken responses to the player.

To introduce environment-aware capabilities into this pipeline, the proposed approach incorporates an additional component, namely the EPM. As illustrated in Fig. 4.4, the process begins with the virtual platform (e.g., a game engine) extracting visual information from the VE. This is achieved by exploiting one or more Virtual Cameras (VCs) positioned within the 3D environment. These cameras asynchronously capture a set of N frames, each representing the scene from a different viewpoint.

Once acquired, the extracted frames are provided as input to a VLM module, which operates under the guidance of a dedicated system prompt. The VLM processes the visual data and produces a structured SG representation of the VE, encoded in the form of a JSON schema. This SG is then injected into the system prompt of the conversational module, effectively augmenting the LLM-ECA with explicit knowledge about its surrounding environment.

The resulting architecture enables the conversational module to generate contextually grounded responses, thereby allowing the ECA to exhibit awareness of its surrounding VE. Depending on the characteristics of the environment, whether static or dynamic (i.e., subject to changes over time), and on the inference latency of the selected VLM, the SG generation process and the subsequent update of the system prompt can be executed repeatedly at configurable time intervals during runtime.

Prompt Design

Drawing inspiration from state-of-the-art approaches [27][91][227], the task assigned to the VLM module is to generate a SG describing all the objects present in the VE based on the previously extracted frames. This objective is achieved through prompt engineering, by defining a dedicated system prompt that explicitly instructs the VLM on how to process the visual inputs and structure the resulting representation.

The system prompt is organized into four main sections: *context*, *input*, *output*, and *rules*. The context section introduces the general task to be performed by the

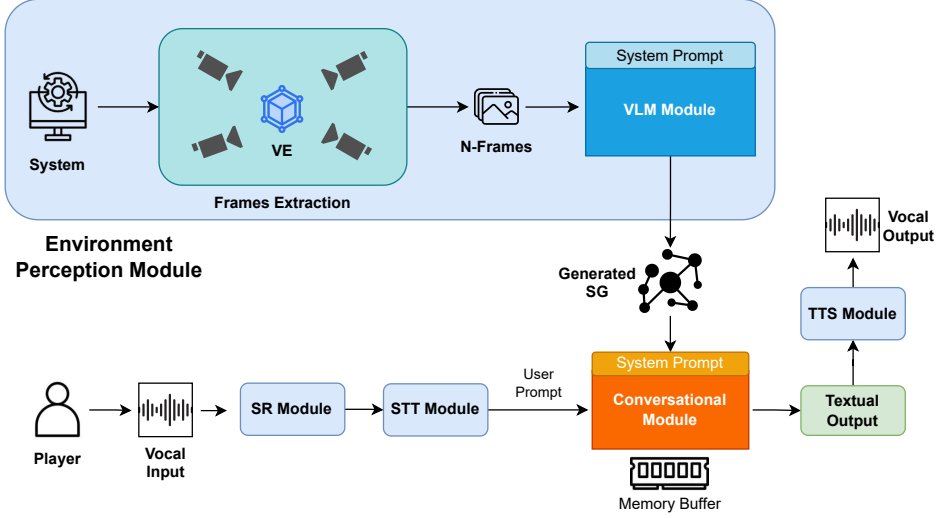


Figure 4.4: Devised VLM-based environment perception module integrated in a traditional architecture of an LLM-based ECA. Over both the VLM module and the Conversational Module, their respective system prompts can be observed.

VLM and specifies the application scenario, which can be adapted depending on the target virtual experience. The input section clarifies the nature of the data provided to the model, namely a set of frames captured from the VE. The output section specifies the expected response format, namely a SG representation. Finally, the rules section defines a set of constraints that the VLM must comply with while processing the input and generating the output. An example of a template system prompt adopted for the VLM module is reported in Table 4.2.

The instructions conveyed through the system prompt enable the generation of a SG composed of nodes and edges. For each object detected within the scene, the VLM generates a node containing the object name along with a set of attributes describing its visual properties. These attributes include features such as color, material, and size, which play a crucial role in defining the visual identity of an object. Even minor variations in these attributes may significantly affect the perception and interpretation of the object. For this reason, the prompt design adopts a structured attribute set derived from the Open-vocabulary Attribute Detection (OvAD) dataset [27], which defines 19 attribute classes (e.g., color, pattern, material) and a total of 117 distinct attribute categories.

In addition to object attributes, the VLM module also generates a set of edges encoding spatial relationships between objects, with the goal of enhancing the environmental awareness of the target LLM-ECA. These relationships are expressed

using subject–predicate–object triplets and are constructed according to three categories commonly employed in traditional spatial reasoning approaches [91]: proximity (e.g., *near*), directionality (e.g., *on*, *under*), and assembly (e.g., *inside*, *hanging from*). Since the extracted frames depict the same VE from multiple viewpoints, relational categories such as adjacency (e.g., *left side*, *front side*), orientation (e.g., *aligned*), and geographic references (e.g., *north*, *south*) are deliberately excluded from the prompt specification.

Following this approach, the output of the EPM consists of a SG representation of the VE that encodes a rich set of environmental information. This representation is subsequently injected into the system prompt of the conversational module, thereby equipping the LLM-ECA with explicit environment-aware capabilities.

4.2.4 Use Case

To validate the proposed system, the investigation adopted an open-source murder-mystery video game featuring ECAs, originally presented in a state-of-the-art work [48]. Since the video game included a highly detailed virtual environment populated by various ECAs, it was deemed suitable for this investigation. The current implementation adapted the application from an immersive VR setup to a desktop-based VR configuration, ensuring accessibility without requiring dedicated immersive hardware. Since the level of immersion was not a relevant factor for the objectives of the investigation, this choice allowed for a broader participation while preserving the integrity of the interaction mechanisms.

The proposed EPM was integrated into the architecture of the ECAs populating the game, thereby equipping them with VLM-based environmental perception capabilities. To evaluate players’ perceptions of the ECAs, a within-subject pilot user investigation was conducted, comparing interactions with ECAs endowed with environmental knowledge against ECAs lacking such capabilities.

Game Application

Within the game, players interact with multiple ECAs in order to collect information and ultimately solve a homicide. The VE is set inside a haunted mansion composed of a diverse set of rooms populated by a wide variety of objects and visual elements, such as furniture, paintings, and narrative props.

The ECAs are embodied as NPCs distributed across different rooms of the VE, each empowered with conversational abilities based on LLMs (GPT-4), along with STT and TTS components. Following prompt-engineering design principles, the authors of the original study [48] defined a structured system prompt for each NPC, conveying information about the game setup (e.g., narrative context and location), the NPC’s personality traits, and their alibi related to the homicide. Additionally, an event-based mechanism was employed to update the system prompt whenever

Section	System prompt instructions
Context	“You are a frame analyzer of a certain virtual application ...”
Input	“You’ll receive one or more images depicting a 3D virtual environment. Assume that all the frames represent the environment state at the same time ...”
Output	“Identify all the visible objects , their attributes (e.g., color, material, size), and their spatial relations to generate a scene graph representation of the virtual environment. Provide the output exclusively in JSON format following this scheme:“ <code>{"objects": [{"name": string, "attributes": [string]}], "relations": [{"subject": string, "predicate": string, "object": string}]}</code> ”.
Rules	1: Label objects of the same category as follows: obj, obj_1, obj_2... (e.g., two windows: window, window_1). 2: The attributes must be listed using their category followed by their value (e.g., “color: red”, “size: tiny/small” ...). You can use the same attribute multiple times, if necessary, but each one must contain only one value. The list of attributes that you are allowed to use is the following ... 3: You can use only the following values for the “predicate” parameter when you generate the spatial relations: [on, under, near, hanging from, inside]. 4: Do not include which frames depict what, just describe what you see. 5: The graphic style does not matter, describe things as it were real. 6: Do not provide additional text, except for the scene graph.

Table 4.2: Example of template system prompt for the VLM module

the player discovered specific interactive objects or clues within the VE.

The system prompt also included a set of behavioral rules to ensure smooth and coherent interactions (e.g., “*Do not mention that you are an NPC*”). Apart from these initialization elements, the original NPC architecture did not incorporate any explicit form of environmental knowledge.

Using the same Unity³ version as the original application (Unity 2022.3.9f1), the proposed EPM was embedded into the NPC architecture. Since each NPC occupies a single room, four VCs were placed within each room at different angles. The cameras were positioned and oriented to maximize coverage of the surrounding environment and its visual features. Each VC captures four frames representing different viewpoints of the room (Figure 4.5). As both the VE and the NPCs are static in the considered game—meaning no spatial changes occur during gameplay—the frames are captured once during game initialization.

The captured frames are encoded in JPG format and forwarded to the VLM module to generate a SG representation in JSON format. In the current implementation, GPT-4.1 is employed as the VLM for the EPM, both for visual processing and SG generation. The resulting SG is then injected into the NPC’s conversational module by appending a dedicated textual rule to its system prompt, instructing the LLM on how to exploit environmental information during player interactions.



(a) Frame 1



(b) Frame 2



(c) Frame 3



(d) Frame 4

Figure 4.5: One of the rooms part of the VE within the video game. The four images represent the frames extracted before passing them to the VLM module to generate the SG. An NPC is standing within the room, surrounded by different objects.

³Unity Engine: <https://unity.com/>

To assess players’ perceptions of environment-aware NPCs, a three-step interaction protocol was defined (Figure 4.6):

- **S1:** Upon encountering an NPC for the first time, the player is instructed via a textual UI to ask three open-ended questions aimed at understanding the NPC’s identity and their stance regarding the homicide.
- **S2:** After completing the first step, the objective shifts to asking questions related to the surrounding VE. This interaction is facilitated through a UI featuring buttons corresponding to four predefined questions about the environment. The questions were designed to align with the narrative context in order to preserve player immersion (e.g., “*If you were the murderer, where would you hide your weapon in here?*”).
- **S3:** In the final step, players are encouraged to freely explore the room in search of clues, while retaining the possibility to interrogate the NPC again if needed. Since the original game provided a limited number of clues, additional interactive element, such as object-based interactions or UI components, were introduced in each room to further stimulate environmental exploration.

Additional adjustments were introduced to improve interaction quality and accessibility. GPT-4.1 was employed for dialogue generation, while the Whisper API⁴ and ElevenLabs Flash V2.5⁵ were utilized respectively for STT and TTS. The changes regarding TTS were made in relation to ElevenLabs faster response times compared to Wit.ai. In addition to the original voice-based interaction, a chat-based interface was introduced, allowing players to type questions as an alternative to spoken interaction.

Experimental Procedure

The evaluation was carried out through an exploratory within-subject user study aimed at comparing players’ perceptions of LLM-based NPCs without environmental knowledge (Version A) and with environmental knowledge (Version B). Specifically, Version A corresponds to the original LLM-NPCs implemented in the video game, whereas Version B includes NPCs augmented with the proposed EPM. To mitigate potential carryover effects between conditions, a Latin Square ordering method [111] was adopted to alternate the presentation order of the two versions across participants.

The exploratory study involved nine university students with prior experience in playing desktop-based VR video games. Personal data was not collected, as it

⁴Whisper STT: <https://platform.openai.com/docs/guides/speech-to-text>

⁵ElevenLabs API: <https://elevenlabs.io/api>



Figure 4.6: An example of the devised interaction modality with one of the NPCs within the VE. In the first step (a,b) the player can freely have a conversation by using a chat interface, while within the second step (c,d) it can select questions from a UI showing dialogue options. In the third step (e,f,g), the player can explore the environment in search of clues, freely interacting again with the NPCs.

was beyond the scope of this exploratory study. All experimental procedures were conducted in accordance with the ethical regulations of Ritsumeikan University, Ibaraki, Japan.

The study took place in a laboratory setting using a desktop PC running the game application. After a brief introduction explaining the goals and procedure of the study, participants were asked to complete an online informed consent form (approximately 10 minutes). Subsequently, each participant was instructed to play one of the two game versions (A or B). Upon completion of the gameplay session

(approximately 20 minutes), participants were asked to complete two questionnaires designed to assess their perceptions of the encountered LLM-NPCs (approximately 10 minutes).

Based on the research questions of the study, questionnaire items assessing Perceived Knowledge, Specific Perceived Knowledge, Perceived Intelligence, Factuality, and Willingness for Future Interaction were adapted from prior work [224]. In addition, the Partner Modeling Questionnaire (PMQ) [59] was employed to evaluate the conversational capabilities of the NPCs (e.g., the ability to effectively communicate, use flexible and human-like language). The two questionnaires comprised 17 and 18 items respectively, all measured on a 7-point Likert scale, resulting in a total of 35 items.

After completing the questionnaires, participants were asked to replay the game using the alternative NPC version (B or A) and subsequently complete the same set of questionnaires again (approximately 20 minutes for gameplay and 10 minutes for questionnaires). Overall, each experimental session lasted approximately 1 hour and 10 minutes.

4.2.5 Results and Discussion

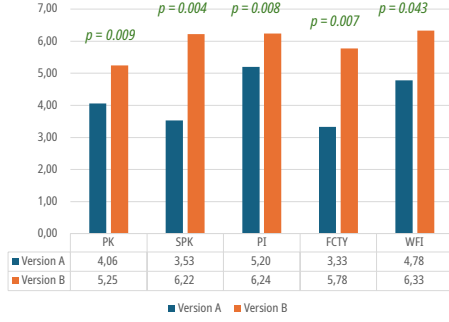
The analysis of the results was conducted across seven dimensions derived from the two employed questionnaires. The first questionnaire was used to assess Perceived Knowledge (PK), Specific Perceived Knowledge (SPK), Perceived Intelligence (PI), Factuality (FCTY), and Willingness of Future Interaction (WFI). The second questionnaire, namely the PMQ, was analyzed to evaluate Communicative Competence and Dependability (CCD), Human-Likeness in Communication (HLC), and Communicative Flexibility (CF).

For each dimension, average participant scores were computed separately for Version A and Version B. The analysis was performed both at the level of individual questionnaire items and by considering the aggregated scores for each dimension. To assess data normality, the Shapiro–Wilk test was applied to each dimension. Based on the outcome of this test, either a Paired Samples t-test (for normally distributed data) or a Wilcoxon Signed-Rank test (for non-normally distributed data) was used. In all cases, a significance level of 5% ($p < .050$) was adopted to identify statistically significant differences between the two versions.

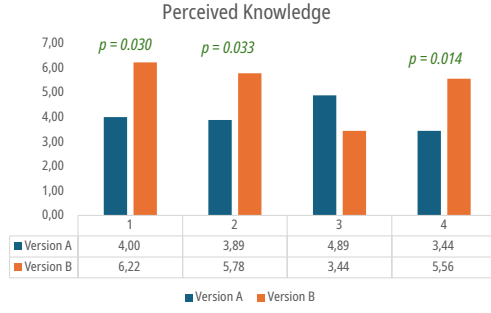
Version A vs. Version B: First Questionnaire

The analysis of the first questionnaire highlights consistent differences between the two experimental conditions. As shown in the overall comparison (Figure 4.7a), Version B, which included NPCs empowered with environmental knowledge, received higher ratings than Version A across all evaluated dimensions. A comparison of mean scores reveals systematically higher values for Version B ($PK = 5.25$, $\sigma =$

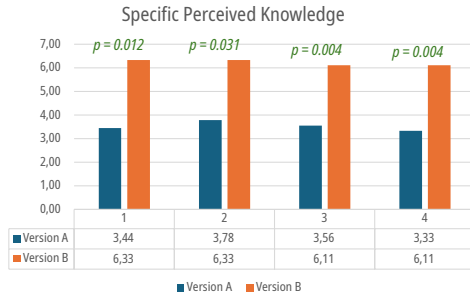
0.57 ; $SPK = 6.22$, $\sigma = 0.47$; $PI = 6.24$, $\sigma = 0.58$; $FCTY = 5.78$, $\sigma = 0.87$; $WFI = 6.33$, $\sigma = 0.86$) compared to Version A ($PK = 4.06$, $\sigma = 1.02$; $SPK = 3.53$, $\sigma = 1.54$; $PI = 5.20$, $\sigma = 0.77$; $FCTY = 3.33$, $\sigma = 1.41$; $WFI = 4.78$, $\sigma = 1.92$). These differences were statistically significant for all five dimensions ($p = 0.009$; 0.004 ; 0.008 ; 0.007 ; 0.043).



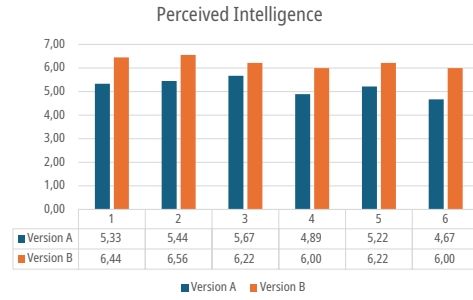
(a) Overall Comparison



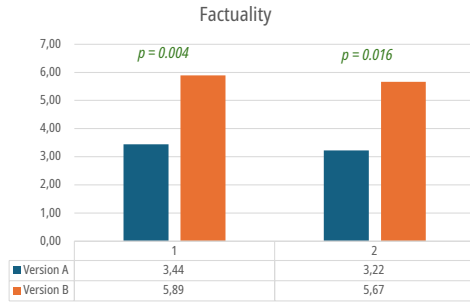
(b) Perceived Knowledge



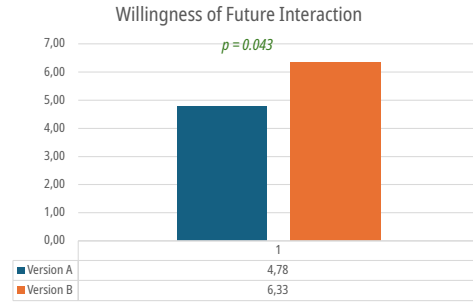
(c) Specific Perceived Knowledge



(d) Perceived Intelligence



(e) Factuality



(f) Willingness of Future Interaction

Figure 4.7: The graphs depict the results obtained from the experimental phase. Specifically, an overall comparison (a) and a more detailed analysis on the individual dimensions: PK (b), SPK (c), PI (d), FCTY (e), and WFI.

A more detailed analysis of individual dimensions shows that participants attributed higher levels of PK to Version B ($mean = 6.22$, $\sigma = 0.83$; $mean = 5.78$, σ

$= 0.83$; $mean = 3.44$, $\sigma = 1.74$; $mean = 5.56$, $\sigma = 1.01$) compared to Version A, with statistically significant differences for most sub-items (PK $p = 0.030$, 0.033 , 0.014), with the exception of the third PK sub-item (Figure 4.7b). Similarly, SPK was consistently rated higher for Version B ($mean = 6.33$, $\sigma = 0.50$; $mean = 6.33$, $\sigma = 0.71$; $mean = 6.11$, $\sigma = 0.78$; $mean = 6.11$, $\sigma = 0.33$), with all but one sub-item showing statistically significant differences (SPK $p = 0.012$, 0.031 , 0.004 , 0.004 ; Figure 4.7c).

WFI (Figure 4.7f) was also rated significantly higher for Version B ($mean = 6.33$, $\sigma = 0.86$) compared to Version A ($mean = 4.78$, $\sigma = 1.92$), yielding a statistically significant difference ($p = 0.043$). Likewise, FCTY (Figure 4.7e) received more favorable evaluations for Version B ($mean = 5.89$, $\sigma = 1.05$; $mean = 5.67$, $\sigma = 0.86$), with both sub-items showing significant differences in favor of NPCs equipped with environmental knowledge ($p = 0.004$; 0.016). In contrast, no statistically significant differences were observed at the level of individual sub-items for PI (Figure 4.7d).

Overall, in relation to the first two research questions (**RQ1** and **RQ2**), these findings indicate that the integration of environmental knowledge substantially enhanced participants' perception of NPCs' knowledge, factuality, and willingness to engage in future interactions. Furthermore, although individual PI sub-items did not reach statistical significance, the aggregated results (Figure 4.7a) indicate a significant improvement in perceived intelligence for NPCs empowered with environment-aware capabilities.

Version A vs. Version B: Second Questionnaire (PMQ)

For the second questionnaire (PMQ), the overall comparison of mean scores (Figure 4.8a) again favored Version B; however, the statistical significance varied across the analyzed dimensions.

CCD exhibited the most pronounced differences between the two conditions. As shown in Figure 4.8b, Version B was consistently rated higher across all CCD sub-items (*mean values ranging from 5.44 to 6.33*), with sub-items 2, 4, and 6 reaching statistical significance ($p = 0.012$; 0.030 ; 0.008). Moreover, the aggregated CCD score ($mean = 5.75$, $\sigma = 1.54$) in the overall comparison was also significantly higher for Version B ($p = 0.033$).

These results suggest that participants perceived NPCs empowered with environmental knowledge as more competent and dependable. A plausible explanation is that the integration of environmental information reduced the likelihood of misleading or inconsistent responses, thereby strengthening users' trust in the NPCs' conversational behavior.

In contrast, no statistically significant differences were observed for HLC (Figure 4.8c) or CF (Figure 4.8d). Although Version B achieved slightly higher mean

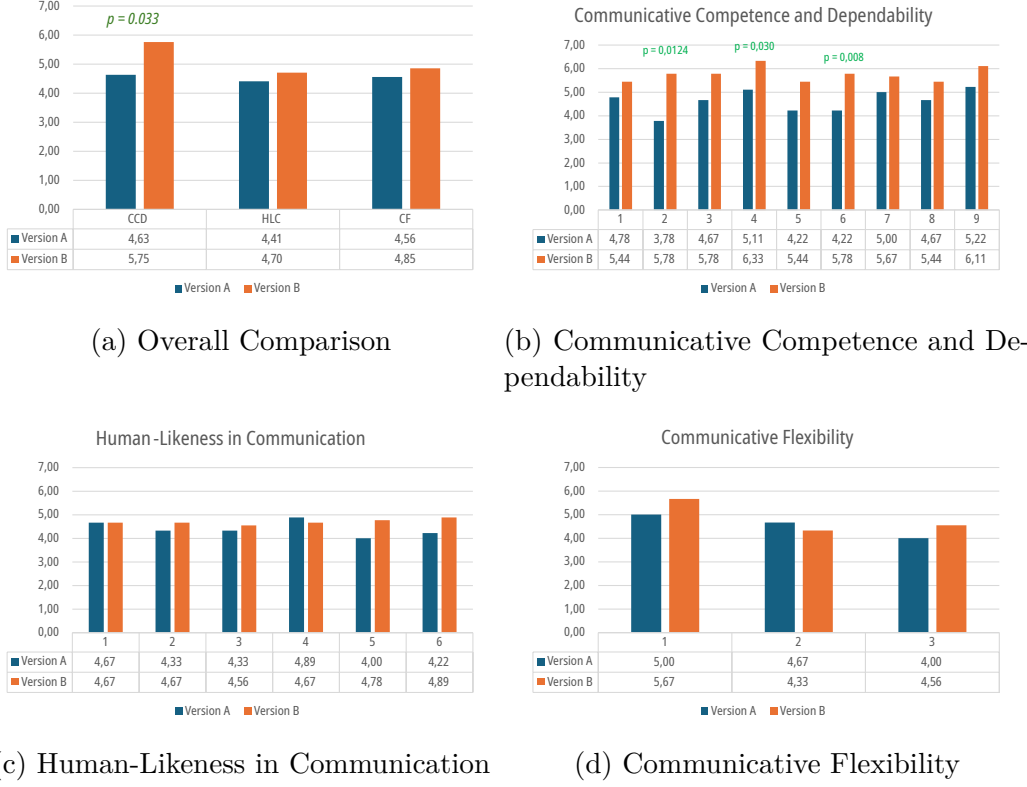


Figure 4.8: The graphs show the results obtained from the experimental phase. Specifically, an overall comparison (a) and a more detailed analysis on the individual dimensions of the PMQ: CCD (b), HLC (c), and CF(d).

scores than Version A in both dimensions, these differences did not reach significance, indicating that environmental awareness did not substantially influence participants' perception of NPCs' human-likeness or conversational adaptability. Given the strong dialogue generation capabilities of the LLM employed in both conditions (GPT-4.1), regardless of the presence of environmental knowledge, this outcome appears reasonable.

In relation to the third research question (**RQ3**), these findings indicate that the integration of environmental knowledge primarily enhanced players' perception of NPCs' communicative competence and dependability, while leaving perceived human-likeness and communicative flexibility largely unaffected.

Limitations and Future Developments

Given the exploratory nature of the study, the proposed approach presents various limitations and challenges. From the perspective of the devised EPM, the awareness capabilities of the ECA are currently possible for static VEs, i.e.,

that do not change their state during gameplay. The SG representation of VE is currently computed during the initialization phase of the VX (i.e., when the game scene is loading). Despite this approach may empower ECAs with environmental knowledge from the beginning of the experience, it doesn't provide any updates to the SG during gameplay, limiting their potential when dynamic environments are involved. These considerations highlight two main concerns:

1. When considering dynamic environments, the employed VLM should rapidly re-compute the VE and updating the SG. This could be fundamental for not hindering the game experience and breaking players' immersion within the VX.
2. Consequently, the SG should be enough accurate to ensure contextual and truthful responses in terms of the surrounding environment.

These two concerns respectively assume the future need of implementing a more sophisticated system to generate and/or update the SG on-the-fly, as well as an additional investigation aiming at comparing the generated SG with a ground-truth SG of the environment.

Another limitation regards the awareness capabilities of the ECA in terms of orientation. Although this kind of information could possibly be inserted within the SG by prompting the VLM with more predicates for the spatial relations (e.g., leftside, frontside), the current approach of extracting frames from multiple view-points of the VE doesn't seem suitable for ensuring an accurate response from the ECA. A possible solution to this concern could be to adopt an hybrid approach, leveraging VLMs for generated the attributes of the objects within the VE, while utilizing traditional approaches (e.g., ray-casting and object tags) to generate the spatial relations and orientation of the objects from the ECA perspective. On the one hand, this approach could accurately ensure the awareness of the ECA in terms of spatial orientation. On the other hand, it would require to embed information in the objects within the VE for each developed application, possibly hindering scalability.

Finally, the current investigation relied on a limited sample size, and performed a analysis exclusively based on users' perception. Future works, considering the described limitations, could involve a deeper analysis on the correctness of the generated SG and the ECA's responses, as well as a wider sample of users, a various set of user populations, and the testing across different application domains (e.g., training, education) for strengthening the generalization and impact of the findings. Moreover, given the within-subject nature of the exploratory study, a complementary between-subject investigation may be required to minimize learning effects and potential carryover across conditions.

4.3 Closing Remarks

This chapter described two investigations regarding the integration of generative AI-based DTTs within VXs for creating ECAs.

The first investigation was placed in the context of CH, exploring principles for a unified UX design across multiple VMEs. Specifically, it described the creation of two different VMEs: a metaverse experience and a virtual workshop. To build a unified UX, the approach focused on creating a contextual and shared VE between the experiences, as well as on the design of a LLM-based ECA. In particular, the latter was based on an architecture leveraging several modules to empower the agent with conversational capabilities, namely SR and STT modules, a generative AI module based on an LLM (GPT-4), and a TTS module to translate AI responses into audio. Additionally, the system relied on carefully designed prompt, establishing conversational rules, providing information about the state of the users and the surrounding VE, as well as communicating instructions to the ECA throughout the entire VX. A preliminary evaluation with users showed positive results in terms of usability of the experience and perception of the ECA, partially suggesting the effectiveness of the approach when it comes to UX across multiple VMEs. Future research is planned to explore the UX aspect in a deeper manner, also considering its correlation with various types of ECAs (e.g., traditional vs LLM-based).

The previous investigation showed a preliminary integration of generative AI within a VX for designing a conversational agent, as well as a prompt-engineering approach to make the ECA functional to the experience. However, the evaluation was carried out within a VX scenario where the interaction with the ECA wasn't the main focus for the users. During the testing phase, users were observed to be more concentrated in following the ECA's instructions and performing the tasks within the VE, than actively trying to establish a conversation. This could possibly explain the positive results for usability and medium-low ratings for all the dimensions (≈ 4) obtained when evaluating the perception of users towards the ECA. Although further research is needed on this topic, these elements provide some indication that the ECA was able to provide guidance to users throughout the experience, while the virtual scenario didn't allow an effective evaluation of users' perception toward the agent. Another consideration is the prompt-engineering modalities, which relied on manually inserting the VE information within the prompts for all the environments. This aspect emerged as a considerable challenge during the VX design process, given the presence of several objects and particular visual traits within the VE.

Based on the previous considerations, the research shifted on environment-aware ECAs. Unlike the prompt-engineering method described in Section 4.1.3, the second investigation proposed an alternative approach leveraging VLMs. Thus, it described the integration of an EPM within traditional ECA architectures, whose pipeline involved the extraction of image-frames from the 3D environment through

virtual cameras, the processing of the frames, and the generation of an SG representation of the VE through a VLM. The devised module was applied to a use case derived from the state of the art, empowering the ECAs of a videogame application with environmental-awareness. Differently from the first investigation, the selected scenario assumed that the player had to actively interact with the ECAs to accomplish the tasks within the VX (i.e., complete the game). Subsequently, an exploratory within-subject study was conducted, evaluating users' perception of ECAs with and without environment-aware capabilities.

The results, although preliminary and exploratory, indicated a substantial enhancement of users' perception towards ECAs empowered with the devised system in terms of various dimensions. Specifically, in comparison to their counterpart, users perceived these agents as more knowledgeable (PK, SPK), more intelligent (PI) and truthful in their responses (FCTY), as well as more competent in their communication skills (CCD) and engaging for future interactions (WFI). The previous results answered to the first two research questions (**RQ1**, **RQ2**) by validating the effectiveness of VLMs to enhance user's perception towards ECAs. Conversely, the third research question (**RQ3**) was not fully answered, as no significative differences between the compared modalities were found in terms of human-likeness (HLC) and flexibility in communication (CF). The latter outcome could possibly due to the fact that both the dimensions (HLC and CF) are not just dependent on the knowledge of the ECA about the environment, but also on its modalities of interaction and conversation (e.g., human-like body movement, eye-gaze, way of speaking).

From the outcomes of this second investigation, some limitations and challenges can be discussed. The proposed EPM currently supports only static VEs, as the SG is generated during the initialization phase and is not updated during gameplay. This limits the applicability of the approach in dynamic scenarios, where re-computation of the SG would be required to preserve environmental accuracy and user immersion. Moreover, ensuring both real-time performance and the correctness of the generated SG remains an open challenge, highlighting the need for future investigations comparing automatically generated representations with ground-truth data. A second limitation concerns spatial orientation awareness. While additional spatial predicates could theoretically be included in the SG, the current multi-view image-based approach does not guarantee reliable orientation-related information from the ECA's perspective. Hybrid solutions combining VLM-based attribute extraction with traditional engine-level techniques (e.g., ray-casting or object tagging) may improve spatial accuracy, but at the cost of reduced scalability and increased integration effort. Finally, the evaluation relied on a limited sample size and focused exclusively on users' subjective perceptions. Although suitable for an exploratory study, this restricts the generalization of the findings.

Overall, the described investigations concluded the three years of the Ph.D.,

highlighting the potential of generative AI-based DTTs when employed for designing ECAs in VXs. In relation to the users' guidance issue described in Chapter 2 (Section 2.2), the first investigation showed how ECAs can be integrated within a VX to direct users throughout the experience. Subsequently, the research focus shifted toward the investigation of VLMs as a means to improve agents' environmental awareness, thus strengthening users' perception of their capabilities.

The described investigations contributed to the **MRQ3** of this thesis by exploring the use of LLMs and VLMs to create and enhance ECAs in two different virtual scenarios. Additionally, although based on preliminary results, the outcomes of the two carried out user studies showed a positive impact on usability and users' perception across several dimensions. A natural continuation of the research is to employ a user-centric design approach to adapt ECAs for specific target users. This element could be particularly fundamental for the suitability of such agents to support fragile or impaired users (e.g., ASD, ADHD) in specific scenarios. Future studies are already planned to further extend the exploration on the matter, as it is also considered to be a key area of interest within the global research community.

Chapter 5

Conclusions

This chapter concludes the thesis by presenting a comprehensive overview of the investigations undertaken and the main results obtained, as well as a discussion of the limitations inherent in the current research and possible future developments.

5.1 Summary of Contributions

The research activities conducted throughout the Ph.D. enabled a comprehensive understanding of the current open challenges in the VR domain for the creation of VXs. From the conceptual issues related to VX design, UX, HCI, and VE management, the research moved toward the convergence of VR with other DTTs. This convergence revealed the potential of such DTTs to tackle existing limitations in VXs, highlighting both their strengths and weaknesses across different application scenarios.

As a first step of the research, the definition of design principles explicitly tailored to the target users was identified as a fundamental aspect in the creation of VXs. Adopting a user-centered design approach allows the main features of a VX, such as the VE, the interaction modalities, and the integrated DTTs, to be systematically defined according to the specific needs and objectives of the intended users. Although user-centered design has been widely investigated in the literature, certain user groups remain underrepresented in current studies and applications.

In Chapter 2, the previous aspects were explored in the context of CH accessibility. Specifically, the chapter presented an investigation encompassing a methodology for the creation of immersive VXs in museums tailored to a specific overlooked category of users, namely those with ASD. The devised methodology was based on core VX design principles that were derived from a profound review of the literature, as well as from a thorough analysis of the challenges typically encountered by ASD individuals in museums (e.g., sensory overload, social discomfort). The primary aim of the investigation was to promote the creation of engaging and accessible

VMEs, while maintaining the cultural value gathered during real-world museum visits.

The methodology was applied for the creation of a VME inspired by a real-world museum hall, namely the Hall of Seasons at Palazzo Carignano (Turin, Italy). The design of the experience was defined through a close cooperation with the curators of the museum and several associations supporting individuals with ASD. The VE development phase, beyond the presence of manual operations, included the use of ER technology, namely 3D photogrammetry, for mirroring elements of the museum environment into their digital replicas. Additionally, the VX creation was complemented by co-design sessions with museum curators, ASD experts and users to gradually validate and enhance the VME. After this refinement phase, the resulting VR application was evaluated through an experimental user study involving individuals with ASD. The study aimed to compare the devised virtual visit with a real-world visit in terms of museum experience, as well as to detect any possible correlation between the effectiveness of the experience and the ASD traits levels of participants. From the analysis of the data obtained during the experiments, several findings emerged. The VX created with the proposed methodology was positively rated by the participants of the study, who perceived the virtual visit as more engaging, inclusive, and pleasant than the real-world visit. Furthermore, those with higher ASD traits were noticed as more positively affected by the VME than the rest of the participants.

Beyond the presence of limitations regarding the methodology and the design principles, two main concerns emerged from this first investigation. Firstly, the technical complexity and knowledge required to reconstruct the VE and create the VX resulted in a substantial investment in terms of time and effort. Secondly, the lack of presence of any embodied entity within the environment limited the effectiveness of guide the users through the experience, and was perceived as a source of discomfort by some participants during the experiments. Thus, the research of the thesis continued towards two directions, both of them considering the convergence of other DTTs with VR to address the noted concerns: the reconstruction of digital environments employing ER technologies, and the creation of conversational agents based on generative AI.

Building on these considerations, Chapter 3 discussed the topic of VE reconstruction. In particular, it described an investigation exploring the use of HBIM to foster the interoperability and the automation of workflows designated for the creation of VHEs. In order to achieve these objectives, the investigation proposed an HBIM-to-VR workflow leveraging an enriched file based on an open BIM format and a complementary VR platform. More specifically, the workflow was structured into two main phases. The first phase focused on enriching the BIM file with geometrical, semantical, and rendering-related information. The second phase involved the VR platform, which leveraged the enriched file for the automated reconstruction of the VE and the creation of the VHE. The devised workflow was applied to

the same museum environment employed for the investigation described in Chapter 1 (i.e., Hall of Seasons of Palazzo Carignano, Turin, Italy). The VR platform was fully able to reconstruct the environment starting from the enriched BIM file, and to generate the VME. Furthermore, it was successful in managing all the rendering features within the VE, including LODs, dynamic lighting and reflections. Overall, the platform showcased the potential of converging this type of ER technology with VR for creating VXs.

The latter investigation provided a deeper understanding of the processes that handle the VE and its characteristics when BIM technologies and VR applications are concurrently employed. The use of the same museum environment examined in Chapter 2 allowed this research to evaluate the efficiency of the proposed semi-automated workflow with respect to both time and effort. At the same time, it enabled to ease the transfer of BIM models in VR applications leveraging an open BIM file format. Altogether, the approach of adopting specific ER technologies in combination with VR showcased how the convergence of DTTs can be useful to support the creation of VXs. This is possibly true also when the considered DTTs (e.g., BIM for VXs) are usually employed for other purposes.

Subsequently, the research of the thesis moved towards generative AI-based DTTs for supporting the creation of conversational agents in VXs. Particularly, Chapter 4 included two investigations respectively employing LLMs and VLMs for enhancing ECA architectures. The aim of the first investigation consisted of designing a unified UX across multiple VMEs, proposing guidelines for the creation of the VE and the inclusion of ECAs. The latter aspect was described through the definition of an ECA architecture and a careful prompt engineering process, with the main purpose of providing guidance to the user throughout the VR experience. A preliminary evaluation with users was also carried out, taking into account two VMEs: a metaverse experience and virtual workshop. The experiments aimed to gather results in terms of usability and users' perception towards the ECA present in both the experiences. The results indicated a positive outcome in terms of usability, while users' perception of the ECA was observed to be prevalently neutral.

The preliminary investigation proved useful for the thesis research, enabling a better understanding of the nature of ECA architectures and supporting the adoption of alternative evaluation scenarios. From a technical perspective, the design of the ECA used in the latter investigation encompassed various types of information, including the context and the state of the application. However, information such as the characteristics of the VE had to be manually inserted for each involved scenario to preserve the ECA context awareness. Additionally, although the ECA was capable of providing instructions throughout the experience, the experimental phase highlighted a lack of interaction between the agent and the users.

The above considerations suggested focusing the thesis research onto specific features of ECAs, possibly leveraging suitable scenarios for their evaluation. Hence, the second investigation in Chapter 4 addressed these concerns by concentrating

on the environmental-awareness of ECAs, proposing an alternative approach based on VLMs. Particularly, the approach included a module for the environmental perception of ECAs (i.e., EPM), which was designed to be inserted into traditional architectures. The module pipeline encompassed the extraction and elaboration of multiple image frames from a VR application, as well as the generation of an SG describing the VE. An exploratory within-subject study was conducted to evaluate the devised approach within the context of a videogame application. Unlike the first investigation, the scenario utilized in this study encouraged active user participation in interacting with the ECA. The evaluation was performed comparing two different modalities in terms of users' perception, considering ECAs with and without environment-awareness. The results highlighted a meaningful enhancement of users' perception towards the environment-aware ECAs in different aspects, including environment knowledge, intelligence, answers truthfulness, and communication skills. On the contrary, no significant enhancements were observed in terms of human-likeness and flexibility in their answers, suggesting the other influential factors (e.g., body language, conversation responsiveness) should be taken into account for conducting further research on ECAs.

In conclusion, the main contributions to the **MRQs** of this doctoral thesis related to VXs and DTTs that emerged from the conducted investigations can be described as follows:

1. **MRQ1:** In VR applications, adopting a user-centered design based on key principles can be crucial for allowing the creation of effective VXs for specific target users (e.g., ASD individuals) and purposes. However, employing VR as a standalone DTT to craft such experiences may result in both conceptual and technical difficulties during the phases of design and development. A solution to these challenges can rely on converging other DTTs with VR, also when they are not usually adopted in the considered context.
2. **MRQ2:** The technical development of VEs can be supported by alternative ER technologies (e.g., BIM) for automating the workflows and easing the VX design process. Although a manual intervention is still currently required to fully enable this convergence of DTTs, the adoption of ER technologies in VR applications represents a first step towards optimizing workflows for creating VXs.
3. **MRQ3:** The aspects related to UX, HCI, and user guidance in VXs can possibly be sustained by using generative AI-based DTTs for creating ECAs. This objective can be achieved by carefully designing the prompt engineering modalities and architectures of the involved ECAs. However, the use cases in which such ECAs are placed should be cautiously selected to allow a fair evaluation of their potential benefits within VXs. Additionally, ECAs can be empowered with specific capabilities that can enhance their architecture

and improve users' perception. Although based on preliminary results, the latter aspect proved to be truthful for certain perceptual dimensions (e.g., PK, SPK), while others remained unaffected (e.g., HLC). Thus, future research should aim to explore in a deeper way the unaffected dimensions to further enhance ECAs in VXs.

To conclude the thesis, the following section describes the current limitations of the conducted research, and reports possible future developments.

5.2 Limitations and Future Works

A primary limitation of the research presented in the thesis relates to its evaluation scenarios. With the exception of the one regarding VLMs for ECAs environment-awareness (Section 4.2), the other investigations were mostly evaluated within the CH domain. Although the devised methodologies could be easily adapted for other fields, this lack of evaluation in alternative scenarios may decrease their scientific value. To address this issue, future work should aim to consider other application domains to confirm the effectiveness of such methodologies. For instance, social VR in the metaverse could be a good fit to further explore users' perception towards ECAs; training and educational platform for industry and manufacturing could be a suitable choice for evaluating VX design and accessibility for various categories of users; ER technologies could be also evaluated in other settings, such as in VR serious games for simulating real-world scenarios (e.g., earthquakes, wildfire emergencies, etc.).

Another limitation is the exploratory nature of some of the proposed investigations, as well as in the reduced sample size involved during the experiment phases. This aspect is mostly referred to the investigations that were focused on generative AI-based ECAs, for which only a preliminary evaluation of the proposed approaches was carried out. Further work is already planned to extend the exploration on this topic, involving a larger and more diverse sample size and ensuring the value of this type DTTs in VXs. With respect to the investigation related to VX design for ASD, the sample size of the between-subject study was influenced by the long timescales involved in recruiting the ASD participants. Furthermore, some of the initially recruited participants were eventually excluded, since their disability level was considered too high for them to enjoy the VX and/or they were not comfortable in wearing an HMD. This aspect, together with other challenges encountered during the experimental phase when working with ASD users, such as concurrent disabilities and/or individual needs, may hinder the initial goal of ensuring accessibility for this population. In this context, future work is planned to extend the research by involving a larger number of ASD individuals, as well as considering alternative modalities of use (e.g., desktop VR) for those who cannot sustain a VX through immersive devices.

Beyond the adjustments needed to address the previous limitations, future research could be directed toward different paths. For instance, the convergence between VR and DTTs based on IoT could be explored. One of the common advantages for this integration is the remote control of IoT devices (e.g., smart systems, appliances) through a VR application. This aspect is especially useful when the considered devices are located in unsafe places, for which remote piloting in a VE could be more practicable [193]. Conversely, using IoT devices can be helpful within VXs for various purposes, such as enabling the synchronization of the VE with the real-world. A natural future development is to leverage the latter aspect to explore the creation of VX-based digital twins, which can be useful in scenarios where the real-time simulation of physical environments is a core element.

Another future development is represented by considering other XR technologies for VXs creation, such as AR and MR. The convergence between these technologies and other DTTs is currently a relevant topic within the research community, such as ER methods for MR applications [122], and the use of LLMs within AR environments [123]. Exploring such integrations for creating VXs could represent an additional opportunity for research and innovation.

Overall, the investigations carried out during the three years of the Ph.D. studies highlighted valuable insights regarding VR and other DTTs, including the importance of adopting an effective VX design, the complexity of creating and reconstructing VEs, as well as the strengths and weaknesses given by employing ECAs. With a deeper exploration of VXs nature and the adoption of other DTTs, future applications of these technologies could be beneficial across various domains. Furthermore, the evolution of DT is still ongoing, continuously introducing new DTTs. It is hoped that the convergence of such technologies with VR will bring further value within VXs, expanding their potential and opening new research opportunities.

Acronyms

AAC Augmentative and Alternative Communication. 15, 19, 24, 33, 36, 42

ADHD Attention-Deficit/Hyperactivity Disorder. 10, 127

AEC Architectural, Engineering, and Construction. 60

AI Artificial Intelligence. 1, 4, 5, 6, 97, 98, 104, 107, 125, 127, 130, 131, 132, 133

AQ-Short Autism-Spectrum Quotient. xii, xiii, 41, 45, 46, 51, 52

AR Augmented Reality. 2, 134

ASD Autism Spectrum Disorder. xii, xiii, 2, 4, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 27, 28, 29, 30, 31, 32, 33, 38, 40, 41, 42, 45, 46, 47, 48, 51, 52, 53, 54, 55, 56, 57, 127, 129, 130, 133

BIM Building Information Modeling. xi, 3, 5, 6, 58, 59, 60, 61, 62, 63, 64, 65, 67, 71, 76, 77, 79, 86, 92, 93, 95, 130, 131, 132

CA Cognitive Attitudes. 106

CCD Communicative Competence and Dependability. xiv, 120, 122, 123, 126

CF Communicative Flexibility. xiv, 120, 122, 123, 126

CG Control Group. 42, 43, 45, 46, 47, 49, 51, 53, 55

CH Cultural Heritage. 6, 9, 10, 11, 12, 13, 17, 18, 19, 23, 25, 43, 54, 56, 57, 59, 60, 61, 62, 63, 67, 69, 73, 78, 86, 89, 93, 94, 95, 98, 125, 129, 133

DOF Degrees of Freedom. 66

DT Digital Transformation. 1, 2, 134

DTT Digital Transformation Technologies. xi, 1, 2, 3, 4, 5, 6, 9, 10, 15, 28, 58, 59, 60, 95, 97, 98, 107, 125, 127, 129, 130, 131, 132, 133, 134

- ECA** Embodied Conversational Agent. xiii, xiv, 4, 5, 6, 7, 97, 98, 99, 100, 101, 102, 104, 105, 106, 107, 108, 109, 110, 111, 113, 114, 115, 123, 124, 125, 126, 127, 131, 132, 133, 134
- EPM** Environment Perception Module. 111, 113, 115, 117, 118, 123, 125, 126, 132
- ER** Environment Reconstruction. 4, 5, 6, 10, 57, 58, 59, 94, 95, 130, 131, 132, 133, 134
- FCTY** Factuality. xiv, 120, 121, 122, 126
- HAI** Human-Agent Interaction. 109, 110
- HBIM** Heritage Building Information Modeling. x, xiii, 59, 60, 61, 62, 63, 64, 65, 67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 86, 87, 89, 91, 92, 93, 94, 130
- HCI** Human-Computer Interaction. 2, 4, 129, 132
- HDM** High-Definition Model. 77, 79, 80, 87, 93, 94, 95, 101
- HLC** Human-Likeness in Communication. xiv, 120, 122, 123, 126, 133
- HM** Hedonic Motivation. 105
- HMD** Head-Mounted Display. 2, 15, 26, 32, 33, 35, 38, 42, 53, 66, 133
- ID** Intellectual Disability. xi, 10, 15, 20, 21
- IFC** Industry Foundation Classes. x, 61, 62, 63, 64, 65, 72, 73, 75, 76, 77, 78, 79, 80, 81, 86, 87, 89, 91, 92, 93, 94, 95
- IoT** Internet of Things. xi, 1, 3, 134
- IP** Interaction Pattern. xi, 24, 25
- IT** Information Technology. 1, 2
- LAO** Look-At Object. xiii, 73, 74, 77, 80, 81, 82, 83, 84, 85, 86, 87, 88, 92
- LLM** Large Language Model. xiv, 4, 5, 6, 94, 97, 98, 99, 100, 101, 104, 107, 108, 109, 110, 111, 113, 114, 115, 117, 118, 120, 123, 125, 127, 131, 134
- LOD** Level of Detail. xiii, 29, 32, 64, 71, 73, 77, 79, 80, 87, 88, 95, 131
- MEEQ** Museum Experience Evaluation Questionnaire. xii, xiii, 43, 46, 48, 49, 50, 51, 52

- MR** Mixed Reality. 2, 134
- MRQ** Macro Research Question. 5, 57, 95, 127, 132
- NLP** Natural Language Processing. 101, 108
- NPC** Non-Player Character. xiv, 4, 108, 110, 115, 116, 117, 118, 119, 120, 122, 123
- PE** Performance Expectancy. 106
- PECS** Picture Exchange Communication System. 14, 18, 19, 24
- PH** Perceived Humanness. 105, 106
- PI** Perceived Intelligence. xiv, 120, 121, 122, 126
- PK** Perceived Knowledge. xiv, 120, 121, 122, 126, 133
- PMQ** Partner Modeling Questionnaire. xiv, 120, 122, 123
- POI** Point of Interest. x, xii, xiii, 18, 28, 31, 32, 35, 36, 37, 38, 39, 40, 42, 53, 54, 55, 73, 74, 81, 82, 83, 84, 85, 86, 87, 88, 90, 92
- Pset** Property Set. 73, 74
- ROI** Region of Interest. 17, 20, 25
- SDK** Software Development Kit. 101
- SG** Scene Graph. xiv, 108, 111, 113, 114, 115, 117, 124, 126, 132
- SPK** Specific Perceived Knowledge. xiv, 120, 121, 122, 126, 133
- SR** Speech Recognition. 104, 113, 125
- STEM** Science, Technology, Engineering, and Mathematics. 16
- STT** Speech-To-Text. 100, 101, 110, 113, 115, 118, 125
- SUS** System Usability Scale. 29, 38, 40, 105, 106
- TA** Teleport Area. 81, 82
- TP** Teleport Point. xii, 34, 35, 37, 81, 82, 87
- TTS** Text-To-Speech. xiii, 61, 66, 67, 72, 73, 82, 83, 84, 85, 89, 100, 101, 105, 110, 113, 115, 118, 125

- UDL** Universal Design for Learning. 14, 18
- UI** User Interface. xiii, xiv, 21, 22, 23, 24, 25, 26, 27, 33, 35, 36, 37, 40, 41, 47, 53, 55, 61, 66, 81, 83, 84, 85, 87, 88, 92, 94, 95, 102, 118, 119
- UX** User eXperience. 2, 4, 5, 6, 11, 43, 66, 98, 99, 100, 102, 106, 125, 129, 131, 132
- VC** Virtual Camera. 113, 117
- VE** Virtual Environment. xi, xii, xiii, xiv, 2, 3, 4, 5, 6, 9, 10, 11, 12, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 28, 32, 34, 35, 36, 37, 38, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 66, 67, 68, 69, 70, 71, 72, 73, 74, 77, 78, 79, 81, 85, 86, 87, 89, 90, 91, 93, 95, 97, 98, 99, 100, 101, 102, 103, 104, 105, 107, 108, 109, 110, 111, 113, 114, 115, 116, 117, 118, 119, 123, 124, 125, 126, 129, 130, 131, 132, 134
- VG** VR Group. 42, 43, 45, 46, 47, 49, 51, 55
- VHE** Virtual Heritage Experience. x, xiii, 59, 60, 61, 62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 76, 77, 78, 81, 82, 85, 89, 91, 93, 94, 98, 130
- VLE** Virtual Learning Environment. 16
- VLM** Vision Language Model. x, xiv, 4, 5, 7, 98, 107, 108, 109, 111, 112, 113, 114, 115, 116, 117, 124, 125, 126, 127, 131, 132, 133
- VME** Virtual Museum Experience. xii, 12, 13, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 33, 35, 38, 41, 42, 43, 48, 51, 53, 55, 56, 57, 98, 99, 100, 101, 102, 103, 104, 107, 125, 130, 131
- VR** Virtual Reality. xi, xii, xiii, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 32, 33, 35, 36, 37, 38, 40, 41, 42, 43, 44, 46, 47, 48, 49, 51, 53, 55, 56, 57, 59, 60, 61, 62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 77, 78, 79, 81, 82, 84, 86, 87, 91, 92, 93, 94, 95, 98, 99, 100, 102, 104, 105, 106, 107, 108, 110, 111, 115, 118, 129, 130, 131, 132, 133, 134
- VX** Virtual eXperience. 2, 3, 4, 5, 6, 9, 10, 11, 16, 17, 20, 22, 43, 55, 56, 57, 58, 59, 86, 94, 95, 97, 98, 106, 107, 108, 124, 125, 126, 127, 129, 130, 131, 132, 133, 134
- WFI** Willingness of Future Interaction. xiv, 120, 121, 122, 126
- XR** eXtended Reality. xi, 2, 3, 134

Bibliography

- [1] C. Achille, N. Lombardini, and C. Tommasi. “BIM and cultural heritage: Compatibility tests in an archaeological site”. In: *Proc. of International Conference on Building Information Modelling (BIM 2015)*. 2015, pp. 593–604. DOI: [10.2495/BIM150481](https://doi.org/10.2495/BIM150481).
- [2] Marcus P. Adams. “Modularity, Theory of Mind, and Autism Spectrum Disorder”. In: *Philosophy of Science* 78.5 (2011), pp. 763–773. DOI: [10.1086/662269](https://doi.org/10.1086/662269).
- [3] Ali Adjorlu et al. “Daily Living Skills Training in Virtual Reality to Help Children with Autism Spectrum Disorder in a Real Shopping Scenario”. In: *Proc. 2017 IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR-Adjunct)*. 2017, pp. 294–302. DOI: [10.1109/ISMAR-Adjunct.2017.93](https://doi.org/10.1109/ISMAR-Adjunct.2017.93).
- [4] Gilles Albeaino et al. “DroneSim: a VR-based flight training simulator for drone-mediated building inspections”. In: *Construction Innovation* 22.4 (2021), pp. 831–848. DOI: [10.1108/CI-03-2021-0049](https://doi.org/10.1108/CI-03-2021-0049).
- [5] S. Alizadehsalehi, A. Hadavi, and J.C. Huang. “From BIM to extended reality in AEC industry”. In: *Automation in Construction* 116 (2020), p. 103254. DOI: [10.1016/j.autcon.2020.103254](https://doi.org/10.1016/j.autcon.2020.103254).
- [6] Ashwaq Z. Amat et al. “Design of an Interactive Virtual Reality System, InViRS, for Joint Attention Practice in Autistic Children”. In: *IEEE Trans. on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 1866–1876. DOI: [10.1109/TNSRE.2021.3108351](https://doi.org/10.1109/TNSRE.2021.3108351).
- [7] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TRTM*. 5th. American Psychiatric Association Publishing, 2013.
- [8] F. Amigoni and V. Schiaffonati. “The Minerva System: A Step toward Automatically Created Virtual Museums”. In: *Applied Artificial Intelligence* 23 (2009), pp. 204–232. DOI: [10.1080/08839510802715159](https://doi.org/10.1080/08839510802715159).

- [9] Elisabeth André and Catherine Pelachaud. “Interacting with Embodied Conversational Agents”. en. In: *Speech Technology: Theory and Applications*. Ed. by Fang Chen and Kristiina Jokinen. New York, NY: Springer US, 2010, pp. 123–149. ISBN: 978-0-387-73819-2. DOI: [10.1007/978-0-387-73819-2_8](https://doi.org/10.1007/978-0-387-73819-2_8). URL: https://doi.org/10.1007/978-0-387-73819-2_8 (visited on 04/13/2026).
- [10] Abdullah M. Al-Ansi et al. “Analyzing augmented reality (AR) and virtual reality (VR) recent development in education”. In: *Social Sciences & Humanities Open* 8.1 (2023), p. 100532. DOI: [10.1016/j.ssaho.2023.100532](https://doi.org/10.1016/j.ssaho.2023.100532).
- [11] ArtPlacer. *ArtPlacer*. <https://www.artplacer.com/>. Accessed: [28-October-2024]. 2024.
- [12] A. Babalola et al. “A bibliometric review of advances in building information modeling (BIM) research”. In: *Journal of Engineering Design and Technology* 21 (2021). DOI: [10.1108/JEDT-01-2021-0013](https://doi.org/10.1108/JEDT-01-2021-0013).
- [13] F. Banfi. “The Evolution of Interactivity, Immersion and Interoperability in HBIM: Digital Model Uses, VR and AR for Built Cultural Heritage”. In: *ISPRS International Journal of Geo-Information* 10.10 (2021). DOI: [10.3390/ijgi10100685](https://doi.org/10.3390/ijgi10100685).
- [14] Aaron Bangor, Philip Kortum, and James Miller. “Determining what individual SUS scores mean: adding an adjective rating scale”. In: *J. Usability Studies* 4.3 (2009), pp. 114–123. DOI: [10.5555/2835587.2835589](https://doi.org/10.5555/2835587.2835589).
- [15] Luca Barbieri, Marcello M. Mariani, and Ilenia Bregoli. “Immersive technologies in museum experiences: a bibliometric analysis”. In: *Front. in Virtual Reality* 4 (2023), p. 1240562. DOI: [10.3389/frvir.2023.1240562](https://doi.org/10.3389/frvir.2023.1240562).
- [16] Rojin Bayat et al. “Exploring Methodologies to Create a Unified VR User-Experience in the Field of Virtual Museum Experiences”. In: *Proc. of 2024 IEEE Gaming, Entertainment, and Media Conference (GEM 2024)*. 2024, pp. 1–4. DOI: [10.1109/GEM61861.2024.10585452](https://doi.org/10.1109/GEM61861.2024.10585452).
- [17] Esubalew Bekele et al. “Understanding How Adolescents with Autism Respond to Facial Expressions in Virtual Reality Environments”. In: *IEEE Trans. on Visualization and Computer Graphics* 19.4 (2013), pp. 711–720. DOI: [10.1109/TVCG.2013.42](https://doi.org/10.1109/TVCG.2013.42).
- [18] Matthew K. Belmonte and Deborah A. Yurgelun-Todd. “Functional anatomy of impaired selective attention and compensatory processing in autism”. In: *Cognitive Brain Research* 17.3 (2003), pp. 651–664. DOI: [10.1016/s0926-6410\(03\)00189-7](https://doi.org/10.1016/s0926-6410(03)00189-7).

- [19] Carlo Bianchini and Giorgia Potestà. “BIM for Built Cultural Heritage: Semantic Segmentation, Architectural Stratification and LOD of the Baptistery of San Giovanni in Florence”. In: *From Building Information Modelling to Mixed Reality*. Ed. by Cecilia Bolognesi and Daniele Villa. Cham: Springer International Publishing, 2021, pp. 1–15. DOI: [10.1007/978-3-030-49278-6](https://doi.org/10.1007/978-3-030-49278-6).
- [20] Alethea Blackler, Vesna Popovic, and Doug Mahar. “Intuitive Interaction Applied to Interface Design”. In: *Proc. Int. Design Congress (IASDR 2005)*. 2005, pp. 1–10. URL: <https://eprints.qut.edu.au/3638/>.
- [21] C.M. Bolognesi and F. Fiorillo. “Virtual Representations of Cultural Heritage: Sharable and Implementable Case Study to Be Enjoyed and Maintained by the Community”. In: *Buildings* 13.2 (2023). DOI: [10.3390/buildings13020410](https://doi.org/10.3390/buildings13020410).
- [22] Andy Bondy and Lori Frost. “The Picture Exchange Communication System”. In: *Behavior Modification* 25.5 (2001), pp. 725–744. DOI: [10.1177/0145445501255004](https://doi.org/10.1177/0145445501255004).
- [23] C. Borés, C. Saurina, and R. Torres. “Technological convergence: a strategic perspective”. In: *Technovation* 23.1 (2003), pp. 1–13. DOI: [10.1016/S0166-4972\(01\)00094-3](https://doi.org/10.1016/S0166-4972(01)00094-3).
- [24] Rineke Bossenbroek et al. “Efficacy of a Virtual Reality Biofeedback Game (DEEP) to Reduce Anxiety and Disruptive Classroom Behavior: Single-Case Study”. In: *JMIR mental health* 7.3 (2020), pp. 1–18. DOI: [10.2196/16066](https://doi.org/10.2196/16066).
- [25] Evren Bozgeyikli et al. “Point & Teleport Locomotion Technique for Virtual Reality”. In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY ’16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 205–216. ISBN: 978-1-4503-4456-2. DOI: [10.1145/2967934.2968105](https://doi.org/10.1145/2967934.2968105). (Visited on 04/15/2026).
- [26] Lal Bozgeyikli et al. “A Survey on Virtual Reality for Individuals with Autism Spectrum Disorder: Design Considerations”. In: *IEEE Trans. on Learning Technologies* 11.2 (2018), pp. 133–151. DOI: [10.1109/TLT.2017.2739747](https://doi.org/10.1109/TLT.2017.2739747).
- [27] María A. Bravo et al. “Open-vocabulary Attribute Detection”. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 7041–7050. DOI: [10.1109/CVPR52729.2023.00680](https://doi.org/10.1109/CVPR52729.2023.00680).
- [28] John Brooke. “SUS: A ‘Quick and Dirty’ Usability Scale”. In: *Usability Evaluation In Industry*. Ed. by Patrick W. Jordan et al. 0th ed. Taylor & Francis, 1996, pp. 207–212. DOI: [10.1201/9781498710411-35](https://doi.org/10.1201/9781498710411-35).
- [29] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. DOI: [10.48550/arXiv.2005.14165](https://doi.org/10.48550/arXiv.2005.14165).

- [30] Building Smart. *Industry Foundation Classes (IFC)*. <https://technical.buildingsmart.org/standards/ifc/>. Accessed: [23-August-2024]. 2010.
- [31] Building Smart. *openBIM*. <https://www.buildingsmart.org/about/openbim/>. Accessed: [05-August-2024]. 2010.
- [32] BuildingSmart. *IfcTextureMap*. <https://ifc43-docs.standards.buildingsmart.org/IFC/RELEASE/IFC4x3/HTML/lexical/IfcTextureMap.htm>. Accessed: [05-August-2024]. 2020.
- [33] BuildingSmart. *Model View Definition*. <https://technical.buildingsmart.org/standards/ifc/mvd/>. Accessed: [23-August-2024]. 2020.
- [34] Gürcan Büyüksalih et al. “Preserving the Knowledge of the Past Through Virtual Visits: From 3D Laser Scanning to Virtual Reality Visualisation at the Istanbul Çatalca İnceğiz Caves”. en. In: *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 88.2 (2020), pp. 133–146. DOI: [10.1007/s41064-020-00091-3](https://doi.org/10.1007/s41064-020-00091-3).
- [35] Yiyu Cai et al. “Design and development of a Virtual Dolphinarium for children with autism”. In: *IEEE Trans. on Neural Systems and Rehabilitation Engineering* 21.2 (2013), pp. 208–217. DOI: [10.1109/TNSRE.2013.2240700](https://doi.org/10.1109/TNSRE.2013.2240700).
- [36] Davide Calandra and Fabrizio Lamberti. “A Testbed for Studying Cybersickness and its Mitigation in Immersive Virtual Reality”. In: *IEEE Transactions on Visualization and Computer Graphics* (2024), pp. 1–18. DOI: [10.1109/TVCG.2024.3448203](https://doi.org/10.1109/TVCG.2024.3448203).
- [37] Aldo Caldarelli et al. “Co-designing Immersive and Inclusive Virtual Museum with children and People with Disabilities: A Pilot Study”. In: *Proc. 2022 Int. Conf. on Computational Science and Computational Intelligence (CSCI)*. 2022, pp. 1964–1969. DOI: [10.1109/CSCI58124.2022.00353](https://doi.org/10.1109/CSCI58124.2022.00353).
- [38] Alberto Cannavò et al. “An Evaluation Testbed for Locomotion in Virtual Reality”. In: *IEEE Transactions on Visualization and Computer Graphics* 27.3 (2021), pp. 1871–1889. DOI: [10.1109/TVCG.2020.3032440](https://doi.org/10.1109/TVCG.2020.3032440).
- [39] Justine Cassell. “Embodied conversational interface agents”. In: *Commun. ACM* 43.4 (2000), pp. 70–78. DOI: [10.1145/332051.332075](https://doi.org/10.1145/332051.332075).
- [40] Joe Cecil, Mary Sweet-Darter, and Avinash Gupta. “Design and Assessment of Virtual Learning Environments to Support STEM Learning for Autistic Students”. In: *Proc. 2020 IEEE Frontiers in Education Conf. (FIE)*. 2020, pp. 1–9. DOI: [10.1109/FIE44824.2020.9274031](https://doi.org/10.1109/FIE44824.2020.9274031).
- [41] Centro Aragonés para la Comunicación Aumentativa y Alternativa - ARASAAC. *simboli CAA - ARASAAC*. <https://beta.arasaac.org/pictograms/search/stagioni>. [Accessed: 14 Jan 2025]. 2007.

- [42] Maria Grazia Cerri. *Palazzo Carignano: Tre secoli di idee, progetti e realizzazioni*. Umberto Allemandi, 1990, p. 262.
- [43] EunJung Chang. “Interactive Experiences and Contextual Learning in Museums”. In: *Studies in Art Education* 47.2 (2006), pp. 170–186.
- [44] Hailin Chen, Jonathan Hale, and Laura Hanks. “Archive or exhibition?: A comparative case study of the real and virtual Pitt Rivers Museum”. In: *Digital Applications in Archaeology and Cultural Heritage* 32 (2024), e00305. DOI: [10.1016/j.daach.2023.e00305](https://doi.org/10.1016/j.daach.2023.e00305).
- [45] Yufang Cheng and Ruowen Huang. “Using virtual reality environment to improve joint attention associated with pervasive developmental disorder”. In: *Research in Developmental Disabilities* 33.6 (2012), pp. 2141–2152. DOI: [10.1016/j.ridd.2012.05.023](https://doi.org/10.1016/j.ridd.2012.05.023).
- [46] James Cho. “Building bridges: Librarians and autism spectrum disorder”. In: *Reference Services Review* 46.3 (2018), pp. 325–339. DOI: [10.1108/RSR-04-2018-0045](https://doi.org/10.1108/RSR-04-2018-0045).
- [47] Nakhoon Choi and Heeyoul Kim. “Technological Convergence of Blockchain and Artificial Intelligence: A Review and Challenges”. en. In: *Electronics* 14.1 (2025), p. 84. DOI: [10.3390/electronics14010084](https://doi.org/10.3390/electronics14010084).
- [48] Frederik Roland Christiansen et al. “Exploring Presence in Interactions with LLM-Driven NPCs: A Comparative Study of Speech Recognition and Dialogue Options”. In: VRST ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 1–11. DOI: [10.1145/3641825.3687716](https://doi.org/10.1145/3641825.3687716).
- [49] Seung Jin Chung, So Yeon Kim, and Ki Han Kim. “Comparison of visitor experiences of virtual reality exhibitions by spatial environment”. In: *International Journal of Human-Computer Studies* 181 (2024), p. 103145. DOI: <https://doi.org/10.1016/j.ijhcs.2023.103145>.
- [50] J. Corey et al. “Digital motor intervention effects on motor performance of individuals with developmental disabilities: A systematic review”. In: *J. of intellectual disability research: JIDR* 68.11 (2024), pp. 1221–1252. DOI: [10.1111/jir.13169](https://doi.org/10.1111/jir.13169).
- [51] Dataverse Ltd. *Artsteps*. <https://www.artsteps.com/>. Accessed: [27-October-2024]. 2024.
- [52] Anais Dawson et al. “Lessons Learned from the Design and Evaluation of InterViewR: A Mixed-Reality Based Interview Training Simulation Platform for Individuals with Autism”. In: *Proc. 2022 IEEE 46th Ann. Computers, Software, and Applications Conf. (COMPSAC)*. 2022, pp. 1324–1330. DOI: [10.1109/COMPSAC54236.2022.00209](https://doi.org/10.1109/COMPSAC54236.2022.00209).

- [53] Valerio De Luca et al. “Virtual Reality and Spatial Augmented Reality for Social Inclusion: The “Includiamoci” Project”. In: *Information* 14.1 (2023), pp. 1–8. DOI: [10.3390/info14010038](https://doi.org/10.3390/info14010038).
- [54] Alex Deffner et al. “Accessibility to culture and heritage: Designing for all”. In: *Proc. Ann. AESOP Congress, Definite Space -- Fuzzy Responsibility*. 2015, pp. 830–844. URL: <https://hdl.handle.net/20.500.14235/1090>.
- [55] Chiara Di Lello. “Guggenheim for All: Museum Education for Students on the Spectrum”. In: *Occasional Paper Series* 2015.33 (2016), pp. 1–14. DOI: [10.58295/2375-3668.1014](https://doi.org/10.58295/2375-3668.1014).
- [56] Jason Diakoumakos, Anastasios Theodoropoulos, and George Lepouras. “Species’ Citadel’s Timeline: A Virtual Reality Experience in a peripheral Archaeological Museum”. In: *Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter*. CHIGREECE ’23. Athens, Greece: Association for Computing Machinery, 2023. DOI: [10.1145/3609987.3610006](https://doi.org/10.1145/3609987.3610006).
- [57] Carlos Alejandro Diaz Schery et al. “Paths to BIM-based digital transformation: a bibliometric and systematic review of critical factors”. In: *Engineering, Construction and Architectural Management* 32.7 (2024), pp. 4357–4383. DOI: [10.1108/ECAM-12-2023-1230](https://doi.org/10.1108/ECAM-12-2023-1230).
- [58] Don Hanson. *New Art City*. <https://newart.city/>. Accessed: [27-October-2024]. 2024.
- [59] Philip R. Doyle et al. “The Partner Modelling Questionnaire: A Validated Self-Report Measure of Perceptions toward Machines as Dialogue Partners”. In: *ACM Trans. Comput.-Hum. Interact.* 32.4 (2025), 39:1–39:33. DOI: [10.1145/3729170](https://doi.org/10.1145/3729170).
- [60] Darby Drageset et al. “Promoting Inclusive Visits to a Natural History Museum with a Pre-Visit VR Tour for Autistic Families”. In: *Research in Science Education* (2024), pp. 1–18. DOI: [10.1007/s11165-024-10207-z](https://doi.org/10.1007/s11165-024-10207-z).
- [61] John Dudley et al. “Inclusive Immersion: A review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse”. In: *Virtual Reality* 27.4 (2023), pp. 2989–3020. DOI: [10.1007/s10055-023-00850-8](https://doi.org/10.1007/s10055-023-00850-8).
- [62] Mihai Duguleană et al. “A Virtual Assistant for Natural Interactions in Museums”. In: *Sustainability* 12.17 (2020). DOI: [10.3390/su12176958](https://doi.org/10.3390/su12176958).
- [63] Sebastian Duque Mahecha et al. “Texture mapping and IFC material retrieval for virtual reality applications”. In: *Proc. of Building Simulation 2021: 17th Conference of IBPSA (BS2021)*. Vol. 17. 2021, pp. 3129–3135. DOI: [10.26868/25222708.2021.30541](https://doi.org/10.26868/25222708.2021.30541).

- [64] European Committee for Standardization. *EN ISO 7817-1:2024: Building Information Modelling - Level of Information Need - Part 1 Concepts and principles (ISO 7817-1:2024)*. European Standard. 2024.
- [65] J. H. Falk and L. D. Dierking. *The Museum Experience Revisited (1st ed.)* Routledge, 2013.
- [66] Silvia Farias-Gaytan, Ignacio Aguaded, and Maria-Soledad Ramirez-Montoya. “Digital transformation and digital literacy in the context of complexity within higher education institutions: a systematic literature review”. en. In: *Humanities and Social Sciences Communications* 10.1 (2023), p. 386. DOI: [10.1057/s41599-023-01875-9](https://doi.org/10.1057/s41599-023-01875-9).
- [67] Zhenan Feng et al. *Rapid 3D Reconstruction of Indoor Environments to Generate Virtual Reality Serious Games Scenarios*. arXiv:1812.01706 [cs]. Dec. 2018. DOI: [10.48550/arXiv.1812.01706](https://doi.org/10.48550/arXiv.1812.01706). URL: <http://arxiv.org/abs/1812.01706> (visited on 04/13/2026).
- [68] U. Ferretti, R. Quattrini, and M. D’Alessio. “A Comprehensive HBIM to XR Framework for Museum Management and User Experience in Ducal Palace at Urbino”. In: *Heritage* 5.3 (2022), pp. 1551–1571. DOI: [10.3390/heritage5030081](https://doi.org/10.3390/heritage5030081).
- [69] Fabio Ferrucci. “Chapter 8 – The Problem of Museum Accessibility: A New Perspective from Relational Sociology and Communicative Gesture”. In: *Gestures*. Ed. by Giovanni Maddalena et al. De Gruyter, 2024, pp. 135–154. DOI: [10.1515/9783110785845-009](https://doi.org/10.1515/9783110785845-009).
- [70] Stefan Feuerriegel et al. “Generative AI”. en. In: *Business & Information Systems Engineering* 66.1 (2024), pp. 111–126. DOI: [10.1007/s12599-023-00834-7](https://doi.org/10.1007/s12599-023-00834-7).
- [71] Jacopo Fiorenza et al. “Designing accessible museum experiences in virtual reality for individuals with Autism Spectrum Disorder”. In: *Digital Applications in Archaeology and Cultural Heritage* 39 (2025), e00471. DOI: [10.1016/j.daach.2025.e00471](https://doi.org/10.1016/j.daach.2025.e00471).
- [72] Jacopo Fiorenza et al. “Enhancing HBIM-to-VR workflows: Semi-automatic generation of virtual heritage experiences using enriched IFC files”. In: *Digital Applications in Archaeology and Cultural Heritage* 36 (2025), e00391. DOI: [10.1016/j.daach.2024.e00391](https://doi.org/10.1016/j.daach.2024.e00391).
- [73] Jacopo Fiorenza et al. “Investigating player perception of environment-aware embodied conversational agents enabled by vision language models”. In: *In Press*. 2026.
- [74] Jennifer K. Fortuna et al. “Exploring the Usefulness of Pre-Visit Materials for Children with Autism at a Public Museum”. In: *Translational Science in Occupation* 1.1 (2024), pp. 1–18. DOI: [10.32873/unmc.dc.tso.1.1.04](https://doi.org/10.32873/unmc.dc.tso.1.1.04).

- [75] Fiona Fui-Hoon Nah et al. “Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration”. In: *Journal of Information Technology Case and Application Research* 25.3 (2023), pp. 277–304. DOI: [10.1080/15228053.2023.2233814](https://doi.org/10.1080/15228053.2023.2233814).
- [76] H. B. Garretson, D. Fein, and L. Waterhouse. “Sustained attention in children with autism”. In: *J. of Autism and Developmental Disorders* 20.1 (1990), pp. 101–114. DOI: [10.1007/BF02206860](https://doi.org/10.1007/BF02206860).
- [77] Franca Garzotto et al. “Improving Museum Accessibility through Storytelling in Wearable Immersive Virtual Reality”. In: *Proc. 2018 3rd DigitalHeritage Int. Congress (DigitalHERITAGE) held jointly with 2018 24th Int. Conf. on Virtual Systems & Multimedia (VSMM)*. 2018, pp. 1–8. DOI: [10.1109/DigitalHeritage.2018.8810097](https://doi.org/10.1109/DigitalHeritage.2018.8810097).
- [78] Catia Giaconi et al. “Virtual and Augmented Reality for the Cultural Accessibility of People with Autism Spectrum Disorders. A Pilot Study”. In: *The Int. J. of the Inclusive Museum* 14.1 (2021), pp. 95–106. DOI: [10.18848/1835-2014/CGP/v14i01/95-106](https://doi.org/10.18848/1835-2014/CGP/v14i01/95-106).
- [79] I. Giangreco et al. “VIRTUE: A virtual reality museum Experience”. In: *Companion Proc. of 24th International Conference on Intelligent User Interfaces (IUI '19 Companion)* (2019), pp. 119–120. DOI: [10.1145/3308557.3308706](https://doi.org/10.1145/3308557.3308706).
- [80] Abhrajit Giri et al. “Lived experience and the social model of disability: Conflicted and inter-dependent ambitions for employment of people with a learning disability and their family carers”. In: *British J. of Learning Disabilities* 50.1 (2022), pp. 98–106. DOI: [10.1111/bld.12378](https://doi.org/10.1111/bld.12378).
- [81] Ana Isabel González-Herrera et al. “Educational and accessible museums and cultural spaces”. In: *Humanities and Social Sciences Communications* 10.1 (2023), pp. 1–8. DOI: [10.1057/s41599-023-01563-8](https://doi.org/10.1057/s41599-023-01563-8).
- [82] K. Goto, M. Kotani, and M. Toyama. “3D Virtual museum generation by using SuperSQL”. In: *International Journal of Pervasive Computing and Communications* 14.2 (2018), pp. 124–146. DOI: [10.1108/IJPCC-D-18-00026](https://doi.org/10.1108/IJPCC-D-18-00026).
- [83] Roberto Gozalo-Brizuela and Eduardo C. Garrido-Merchán. *A survey of Generative AI Applications*. 2023. DOI: [10.48550/arXiv.2306.02781](https://doi.org/10.48550/arXiv.2306.02781).
- [84] K. Graham, L. Chow, and S. Fai. “From BIM to VR: Defining a level of detail to guide virtual reality narratives”. In: *Journal of Information Technology in Construction* 24 (2019), pp. 553–568. DOI: [10.36680/j.itcon.2019.031](https://doi.org/10.36680/j.itcon.2019.031).
- [85] Leandro Soares Guedes. *Accessibility by design: Designing Inclusive Technologies with and for People with Intellectual Disabilities*. Doctoral dissertation. 2024. URL: <https://n2t.net/ark:/12658/srd1328319>.

- [86] Leandro Soares Guedes et al. “Co-designing a Museum Application with People with Intellectual Disabilities: Findings and Accessible Redesign”. In: *Proc. European Conf. on Cognitive Ergonomics 2023*. ECCE '23. Association for Computing Machinery. 2023, pp. 1–8. DOI: [10.1145/3605655.3605687](https://doi.org/10.1145/3605655.3605687).
- [87] Pascal Guitton, Hélène Sauzéon, and Pierre-Antoine Cinquin. “Accessibility of Immersive Serious Games for Persons with Cognitive Disabilities”. In: *Proc. 2019 IEEE Int. Symp. on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2019, pp. 443–447. DOI: [10.1109/ISMAR-Adjunct.2019.00051](https://doi.org/10.1109/ISMAR-Adjunct.2019.00051).
- [88] Yangyang Guo et al. “Virtual human pose estimation in a fire education system for children with autism spectrum disorders”. In: *Multimedia Systems* 30.2 (2024), pp. 1–15. DOI: [10.1007/s00530-024-01274-3](https://doi.org/10.1007/s00530-024-01274-3).
- [89] Ives Gutierrez, João J Ferreira, and Paula O Fernandes. “Digital transformation and the new combinations in tourism: A systematic literature review”. EN. In: *Tourism and Hospitality Research* 25.2 (2025), pp. 194–213. DOI: [10.1177/14673584231198414](https://doi.org/10.1177/14673584231198414).
- [90] Erwin Halim et al. “Virtual Reality As A Social Learning Tool For Individuals With Autism”. In: *Proc. 2022 5th Int. Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. 2022, pp. 805–810. DOI: [10.1109/ISRITI56927.2022.10053065](https://doi.org/10.1109/ISRITI56927.2022.10053065).
- [91] Steven Häsler and Philipp Ackermann. *Spatial Reasoner: A 3D Inference Pipeline for XR Applications*. 2025. DOI: [10.48550/arXiv.2504.18380](https://doi.org/10.48550/arXiv.2504.18380).
- [92] Syed Zohaib Hassan et al. “Towards an AI-driven talking avatar in virtual reality for investigative interviews of children”. In: *Proceedings of the 2nd Workshop on Games Systems*. GameSys '22. Athlone, Ireland: Association for Computing Machinery, 2022, pp. 9–15. DOI: [10.1145/3534085.3534340](https://doi.org/10.1145/3534085.3534340).
- [93] M. Hayashi, S. Bachelder, and M. Nakajima. “Automatic Generation of Art Exhibitions from the Internet Resources”. In: *Proc. of 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE 2020)*. 2020, pp. 642–643. DOI: [10.1109/GCCE50665.2020.9291722](https://doi.org/10.1109/GCCE50665.2020.9291722).
- [94] Aref Hebri et al. “Can Virtual AI Agents Improve Cognitive Assessment? A Virtual Reality Perspective”. In: *Proceedings of the 18th ACM International Conference on Pervasive Technologies Related to Assistive Environments*. PETRA '25. New York, NY, USA: Association for Computing Machinery, 2025, pp. 424–428. DOI: [10.1145/3733155.3734917](https://doi.org/10.1145/3733155.3734917).
- [95] Fiona Heilemann, Gottfried Zimmermann, and Patrick Münster. “Accessibility Guidelines for VR Games - A Comparison and Synthesis of a Comprehensive Set”. English. In: *Frontiers in Virtual Reality* 2 (Oct. 2021). ISSN: 2673-4192. DOI: [10.3389/frvir.2021.697504](https://doi.org/10.3389/frvir.2021.697504).

- [96] Libby Hladik et al. “Accessibility and Inclusion for Families with Children with Autism Spectrum Disorders in Cultural Institutions”. In: *Curator: The Museum J.* 65.2 (2022), pp. 435–449. DOI: [10.1111/cura.12468](https://doi.org/10.1111/cura.12468).
- [97] Rosa A. Hoekstra et al. “The Construction and Validation of an Abridged Version of the Autism-Spectrum Quotient (AQ-Short)”. In: *J. of Autism and Developmental Disorders* 41.5 (2011), pp. 589–596. DOI: [10.1007/s10803-010-1073-0](https://doi.org/10.1007/s10803-010-1073-0).
- [98] James Hutson and Piper Hutson. “Perspective Chapter: Museums and the Metaverse – Emerging Technologies to Promote Inclusivity and Engagement”. In: *Application of Modern Trends in Museums*. Ed. by Ladislav Župčán. IntechOpen, 2023, pp. 1–20. DOI: [10.5772/intechopen.110044](https://doi.org/10.5772/intechopen.110044).
- [99] Piper Hutson and James Hutson. “Neurodivergence and Inclusivity in Cultural Institutions: A Review of Theories and Best Practices”. In: *Creative Education* 13.09 (2022), pp. 3069–3080. DOI: [10.4236/ce.2022.139193](https://doi.org/10.4236/ce.2022.139193).
- [100] IfcOpenShell. *IfcConvert*. <https://docs.ifcopenshell.org/ifcconvert.html>. Accessed: [02-August-2024]. 2022.
- [101] Julie E. N. Irish. “An Exploratory Study Testing Environmental Wayfinding Aids as an Intervention for Children With Autism”. In: *Health Environments Research & Design J.* 15.4 (2022), pp. 114–130. DOI: [10.1177/19375867221111467](https://doi.org/10.1177/19375867221111467).
- [102] Andrés Isaza-Giraldo et al. “Prompt-Gaming: A Pilot Study on LLM-Evaluating Agent in a Meaningful Energy Game”. In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 1–12. DOI: [10.1145/3613905.3650774](https://doi.org/10.1145/3613905.3650774).
- [103] Katherine Isbister. *Better Game Characters by Design*. en. 0th ed. CRC Press, 2006. DOI: [10.1201/9780367807641](https://doi.org/10.1201/9780367807641).
- [104] International Organization for Standardization. *ISO/TS 16739-1:2024 Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries*. Standard. Edition 2. 2024.
- [105] Md Asaduzzaman Jabin et al. *AdCare-VLM: Leveraging Large Vision Language Model (LVLM) to Monitor Long-Term Medication Adherence and Care*. 2025. DOI: [10.48550/arXiv.2505.00275](https://doi.org/10.48550/arXiv.2505.00275).
- [106] Dominik Jeurissen et al. *Playing NetHack with LLMs: Potential & Limitations as Zero-Shot Agents*. 2024. DOI: [10.48550/arXiv.2403.00690](https://doi.org/10.48550/arXiv.2403.00690).
- [107] Yunshui Jin, Minhua Ma, and Yun Liu. “Comparative Study of HMD-based Virtual and Augmented Realities for Immersive Museums: User Acceptance, Medium, and Learning”. In: *J. Comput. Cult. Herit.* 17.1 (2024). DOI: [10.1145/3627164](https://doi.org/10.1145/3627164).

- [108] Timothy Jung et al. “Effects of Virtual Reality and Augmented Reality on Visitor Experiences in Museum”. In: *Proc. of Information and Communication Technologies in Tourism 2016*. Ed. by Alessandro Inversini and Roland Schegg. 2016, pp. 621–635. DOI: [10.1007/978-3-319-28231-2_45](https://doi.org/10.1007/978-3-319-28231-2_45).
- [109] A. Kamari, A. Paari, and H.Ø. Torvund. “BIM-enabled Virtual Reality (VR) for Sustainability Life Cycle and Cost Assessment”. In: *Sustainability* 13.1 (2021). DOI: [10.3390/su13010249](https://doi.org/10.3390/su13010249).
- [110] M. Kazhdan and H. Hoppe. “Screened poisson surface reconstruction”. In: *ACM Transactions on Graphics* 32.3 (2013). DOI: [10.1145/2487228.2487237](https://doi.org/10.1145/2487228.2487237).
- [111] “Front matter”. In: *Latin Squares and their Applications (Second Edition)*. Ed. by A. Donald Keedwell and József Dénes. Boston: North-Holland, 2015, pp. i–iii. DOI: [10.1016/B978-0-444-63555-6.50016-7](https://doi.org/10.1016/B978-0-444-63555-6.50016-7).
- [112] Muhammad Haris Khan et al. *HapticVLM: VLM-Driven Texture Recognition Aimed at Intelligent Haptic Interaction*. 2025. DOI: [10.48550/arXiv.2505.02569](https://doi.org/10.48550/arXiv.2505.02569).
- [113] A. Khorchi and C. Botton. “An OpenBIM-based 4D approach to support coordination meetings in virtual reality environments”. In: *Journal of Building Engineering* 85 (2024), p. 108647. DOI: [10.1016/j.jobe.2024.108647](https://doi.org/10.1016/j.jobe.2024.108647).
- [114] H. Kim et al. “Space-adaptive Artwork Placement Based on Content Similarities for Curating Thematic Spaces in a Virtual Museum”. In: *ACM Journal on Computing and Cultural Heritage* 17.1 (2024). DOI: [10.1145/3631134](https://doi.org/10.1145/3631134).
- [115] Ellie King et al. “Creating Meaningful Museums: A Model for Museum Exhibition User Experience”. In: *Visitor Studies* 26.1 (2023), pp. 59–81. DOI: [10.1080/10645578.2022.2129944](https://doi.org/10.1080/10645578.2022.2129944).
- [116] C. Kiourt, A. Koutsoudis, and G. Pavlidis. “DynaMus: A fully dynamic 3D virtual museum framework”. In: *Journal of Cultural Heritage* 22 (2016), pp. 984–991. DOI: [10.1016/j.culher.2016.06.007](https://doi.org/10.1016/j.culher.2016.06.007).
- [117] Joanna Kisker, Thomas Gruber, and Benjamin Schöne. “Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure”. eng. In: *Psychological Research* 85.1 (2021), pp. 68–81. DOI: [10.1007/s00426-019-01244-9](https://doi.org/10.1007/s00426-019-01244-9).
- [118] Rajiv Kohli and Shawn Johnson. “Digital Transformation in Latecomer Industries: CIO and CEO Leadership Lessons from Encana Oil & Gas (USA) Inc.” In: *MIS Quarterly Executive* 10.4 (2011). URL: <https://aisel.aisnet.org/misqe/vol10/iss4/3>.
- [119] Mikhail Konenkov et al. *VR-GPT: Visual Language Model for Intelligent Virtual Reality Applications*. 2024. DOI: [10.48550/arXiv.2405.11537](https://doi.org/10.48550/arXiv.2405.11537).

- [120] E.H. Korkut and E. Surer. “Developing a framework for heterotopias as discursive playgrounds: A comparative analysis of non-immersive and immersive technologies”. In: *Virtual Reality* 28.1 (2024), p. 16. DOI: [10.1007/s10055-023-00905-w](https://doi.org/10.1007/s10055-023-00905-w).
- [121] Pradeep Raj Krishnappa Babu, Poojan Oza, and Uttama Lahiri. “Gaze-Sensitive Virtual Reality Based Social Communication Platform for Individuals with Autism”. In: *IEEE Trans. on Affective Computing* 9.4 (2018), pp. 450–462. DOI: [10.1109/TAFFC.2016.2641422](https://doi.org/10.1109/TAFFC.2016.2641422).
- [122] Aman Kukreja et al. “A knowledge framework of environment reconstruction methods for mixed reality prototype applications”. en. In: *Proceedings of the Design Society* 5 (2025), pp. 141–150. DOI: [10.1017/pds.2025.10028](https://doi.org/10.1017/pds.2025.10028).
- [123] Santosh Kumar et al. “Future of Human-AI Interaction: Bridging the Gap with LLMs and AR Integration”. In: *2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS)*. 2025, pp. 1–6. DOI: [10.1109/WorldSUAS66815.2025.11199115](https://doi.org/10.1109/WorldSUAS66815.2025.11199115).
- [124] Kunstmatrix. *Kunstmatrix*. <https://www.kunstmatrix.com/en>. Accessed: [27-October-2024]. 2024.
- [125] Harold J. Leavitt and Thomas L. Whisler. “Management in the 1980’s”. en. In: *Harvard Business Review* (1958). URL: <https://hbr.org/1958/11/management-in-the-1980s>.
- [126] Geonsun Lee et al. ““May I Speak?”: Multi-Modal Attention Guidance in Social VR Group Conversations”. In: *IEEE Trans. on Visualization and Computer Graphics* 30.5 (2024), pp. 2287–2297. DOI: [10.1109/TVCG.2024.3372119](https://doi.org/10.1109/TVCG.2024.3372119).
- [127] Hyunae Lee et al. “Experiencing immersive virtual reality in museums”. In: *Information & Management* 57.5 (2020), pp. 1–9. DOI: [10.1016/j.im.2019.103229](https://doi.org/10.1016/j.im.2019.103229).
- [128] Haoyang Li et al. “X’s Day: Personality-Driven Virtual Human Behavior Generation”. In: *IEEE Transactions on Visualization and Computer Graphics* 31.5 (2025), pp. 3514–3524. DOI: [10.1109/TVCG.2025.3549574](https://doi.org/10.1109/TVCG.2025.3549574).
- [129] Jiajia Li et al. “DianTea: Designing and Evaluating an Immersive Virtual Reality Game to Enhance Youth Tea Culture Learning”. In: *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction. MobileHCI ’23 Companion*. Athens, Greece: Association for Computing Machinery, 2023. DOI: [10.1145/3565066.3608707](https://doi.org/10.1145/3565066.3608707).
- [130] Jingjing Li et al. “A systematic review of digital transformation technologies in museum exhibition”. In: *Computers in Human Behavior* 161 (2024), p. 108407. DOI: [10.1016/j.chb.2024.108407](https://doi.org/10.1016/j.chb.2024.108407).

- [131] Ziming Li et al. “Exploring Large Language Model-Driven Agents for Environment-Aware Spatial Interactions and Conversations in Virtual Reality Role-Play Scenarios”. In: *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 2025. DOI: [10.1109/VR59515.2025.00025](https://doi.org/10.1109/VR59515.2025.00025).
- [132] Bill Yuchen Lin et al. *SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks*. 2023. DOI: [10.48550/arXiv.2305.17390](https://doi.org/10.48550/arXiv.2305.17390).
- [133] Jiaju Lin et al. *AgentSims: An Open-Source Sandbox for Large Language Model Evaluation*. 2023. DOI: [10.48550/arXiv.2308.04026](https://doi.org/10.48550/arXiv.2308.04026).
- [134] Chao Liu et al. “Toward Facilitating Search in VR With the Assistance of Vision Large Language Models”. In: *Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology*. VRST ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 1–14. DOI: [10.1145/3641825.3687742](https://doi.org/10.1145/3641825.3687742).
- [135] Gonzalo Lorenzo et al. “Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders”. In: *Computers & Education* 98 (2016), pp. 192–205. DOI: [10.1016/j.compedu.2016.03.018](https://doi.org/10.1016/j.compedu.2016.03.018).
- [136] Louvre Museum. *Mona Lisa Beyond the Glass*. <https://www.louvre.fr/en/explore/life-at-the-museum/mona-lisa-beyond-the-glass-the-louvre-s-first-virtual-reality-experience>. Accessed: [5-November-2024]. 2019.
- [137] L.J. Lovell, R.J. Davies, and D.V.L. Hunt. “The Application of Historic Building Information Modelling (HBIM) to Cultural Heritage: A Review”. In: *Heritage* 6.10 (2023), pp. 6691–6717. DOI: [10.3390/heritage6100350](https://doi.org/10.3390/heritage6100350).
- [138] Haopeng Lu et al. “GazeTance Guidance: Gaze and Distance-Based Content Presentation for Virtual Museum”. In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2021, pp. 462–463. DOI: [10.1109/VRW52623.2021.00113](https://doi.org/10.1109/VRW52623.2021.00113).
- [139] E. Lucchi. “Digital twins for the automation of the heritage construction sector”. In: *Automation in Construction* 156 (2023), p. 105073. DOI: [10.1016/j.autcon.2023.105073](https://doi.org/10.1016/j.autcon.2023.105073).
- [140] Alexander Lussenhop et al. “Social participation of families with children with autism spectrum disorder in a science museum”. In: *Museums & Social Issues* 11.2 (2016), pp. 122–137. DOI: [10.1080/15596893.2016.1214806](https://doi.org/10.1080/15596893.2016.1214806).
- [141] Dechao Ma and Weiwei Wu. “Does artificial intelligence drive technology convergence? Evidence from Chinese manufacturing companies”. In: *Technology in Society* 79 (2024), p. 102715. DOI: [10.1016/j.techsoc.2024.102715](https://doi.org/10.1016/j.techsoc.2024.102715).

- [142] Xiaoyue Ma and Yudi Huo. “Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework”. In: *Technology in Society* 75 (2023), p. 102362. DOI: [10.1016/j.techsoc.2023.102362](https://doi.org/10.1016/j.techsoc.2023.102362).
- [143] Keren MacLennan et al. ““It Is a Big Spider Web of Things”: Sensory Experiences of Autistic Adults in Public Spaces”. In: *Autism in Adulthood: Challenges and Management* 5.4 (2023), pp. 411–422. DOI: [10.1089/aut.2022.0024](https://doi.org/10.1089/aut.2022.0024).
- [144] Elhassan Makled et al. “PathoGenius VR: VR medical training”. In: PerDis ’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–2. DOI: [10.1145/3321335.3329694](https://doi.org/10.1145/3321335.3329694).
- [145] Irene C. Mammarella, Ramona Cardillo, and Margaret Semrud-Clikeman. “Do comorbid symptoms discriminate between autism spectrum disorder, ADHD and nonverbal learning disability?” In: *Research in Developmental Disabilities* 126 (2022), p. 104242. DOI: [10.1016/j.ridd.2022.104242](https://doi.org/10.1016/j.ridd.2022.104242).
- [146] Kenji Mase, Rieko Kadobayashi, and Ryohei Nakatsu. “Meta-Museum: A Supportive Augmented-Reality Environment for Knowledge Sharing”. In: *VSM: International Conference on Virtual Systems and MultiMedia*. 1996.
- [147] Marianna Mauro et al. “Digital transformation in healthcare: Assessing the role of digital technologies for managerial support processes”. In: *Technological Forecasting and Social Change* 209 (2024), p. 123781. DOI: [10.1016/j.techfore.2024.123781](https://doi.org/10.1016/j.techfore.2024.123781).
- [148] Christofer Meinecke, Chris Hall, and Stefan Jänicke. “Towards Enhancing Virtual Museums by Contextualizing Art through Interactive Visualizations”. In: *Journal on Computing and Cultural Heritage* 15.4 (2022). DOI: [10.1145/3527619](https://doi.org/10.1145/3527619).
- [149] Ivo Alexander Meins et al. “VR-SOAP, a modular virtual reality treatment for improving social activities and participation of young people with psychosis: A study protocol for a single-blind multi-centre randomized controlled trial”. In: *Trials* 24.1 (2023), pp. 1–11. DOI: [10.1186/s13063-023-07241-z](https://doi.org/10.1186/s13063-023-07241-z).
- [150] Rafael Melendreras-Ruiz et al. “Analysis of the user experience (On site vs. Virtual Reality) through biological markers and cognitive tests in museums: the case of Museo Cristo de la Sangre (Murcia, Spain)”. In: *Virtual Reality* 28 (2024). DOI: [10.1007/s10055-023-00928-3](https://doi.org/10.1007/s10055-023-00928-3).
- [151] Nicolò Messina et al. “Museum for All: Wearable Immersive Virtual Tours in Museums for People with Neurodevelopmental Disorders”. In: *IOP Conf. Series: Materials Science and Engineering* 364.1 (2018), p. 012047. DOI: [10.1088/1757-899X/364/1/012047](https://doi.org/10.1088/1757-899X/364/1/012047).

- [152] Pat Mirenda. “Toward Functional Augmentative and Alternative Communication for Students With Autism”. In: *Language, Speech, and Hearing Services in Schools* 34.3 (2003), pp. 203–216. DOI: [10.1044/0161-1461\(2003/017\)](https://doi.org/10.1044/0161-1461(2003/017)).
- [153] Abhijit Mishra et al. *ReVision: A Dataset and Baseline VLM for Privacy-Preserving Task-Oriented Visual Instruction Rewriting*. 2025. DOI: [10.48550/arXiv.2502.14780](https://doi.org/10.48550/arXiv.2502.14780).
- [154] Nurul Fathihin Mohd Noor Shah and Masitah Ghazali. “A Systematic Review on Digital Technology for Enhancing User Experience in Museums”. In: *User Science and Engineering*. Ed. by Natrah Abdullah, Wan Adilah Wan Adnan, and Marcus Foth. Springer Singapore, 2018, pp. 35–46. DOI: [10.1007/978-981-13-1628-9_4](https://doi.org/10.1007/978-981-13-1628-9_4).
- [155] Nicolas Mollet, Stéphanie Gerbaud, and Bruno Arnaldi. “An Operational VR Platform for Building Virtual Training Environments”. en. In: *Technologies for E-Learning and Digital Entertainment*. Ed. by Kin-chuen Hui et al. Berlin, Heidelberg: Springer, 2007, pp. 140–151. DOI: [10.1007/978-3-540-73011-8_16](https://doi.org/10.1007/978-3-540-73011-8_16).
- [156] Jewoong Moon, Gi Woong Choi, and Joo Young Seo. “Revisiting multimedia learning design principles in virtual reality-based learning environments for autistic individuals”. In: *Virtual Reality* 27.4 (2023), pp. 3101–3113. DOI: [10.1007/s10055-023-00856-2](https://doi.org/10.1007/s10055-023-00856-2).
- [157] Martez Mott et al. “Accessible by Design: An Opportunity for Virtual Reality”. In: *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2019, pp. 451–454. DOI: [10.1109/ISMAR-Adjunct.2019.00122](https://doi.org/10.1109/ISMAR-Adjunct.2019.00122).
- [158] Linda Münch et al. “Accessibility Criteria for an Inclusive Museum for People with Learning Disabilities: A Review”. In: *Universal Access in Human-Computer Interaction. User and Context Diversity (Proc. HCII 2022)*. Ed. by Margherita Antona and Constantine Stephanidis. Cham: Springer Int. Publishing, 2022, pp. 371–385. DOI: [10.1007/978-3-031-05039-8_27](https://doi.org/10.1007/978-3-031-05039-8_27).
- [159] M. Murphy, E. McGovern, and S. Pavia. “Historic building information modelling (HBIM)”. In: *Structural Survey* 27.4 (2009), pp. 311–327. DOI: [10.1108/02630800910985108](https://doi.org/10.1108/02630800910985108).
- [160] W. Natephra et al. “Integrating building information modeling and virtual reality development engines for building indoor lighting design”. In: *Visualization in Engineering* 5.1 (2017), p. 19. DOI: [10.1186/s40327-017-0058-x](https://doi.org/10.1186/s40327-017-0058-x).
- [161] National BIM Standard Project Committee and others. *National building information modeling standard, version 1, Part 1: Overview, principles, and methodologies*. National Institute of Building Sciences. 2007.

- [162] G.A. van Nederveen and F.P. Tolman. “Modelling multiple views on buildings”. In: *Automation in Construction* 1.3 (1992), pp. 215–224. DOI: [10.1016/0926-5805\(92\)90014-B](https://doi.org/10.1016/0926-5805(92)90014-B).
- [163] Haemoon Oh, Ann Marie Fiore, and Miyoung Jeoung. “Measuring Experience Economy Concepts: Tourism Applications”. In: *J. of Travel Research* 46.2 (2007), pp. 119–132. DOI: [10.1177/0047287507304039](https://doi.org/10.1177/0047287507304039).
- [164] Edwin Juma Omol. “Organizational digital transformation: from evolution to future trends”. In: *Digital Transformation and Society* 3.3 (2023), pp. 240–256. DOI: [10.1108/DTS-08-2023-0061](https://doi.org/10.1108/DTS-08-2023-0061).
- [165] Laurens Jozef Nicolaas Oostwegel et al. “Digitalization of culturally significant buildings: Ensuring high-quality data exchanges in the heritage domain using OpenBIM”. In: *Heritage Science* 10.1 (2022), p. 10. DOI: [10.1186/s40494-021-00640-y](https://doi.org/10.1186/s40494-021-00640-y).
- [166] *OpenAI Prompt Engineering Guidelines*. <https://platform.openai.com/docs/guides/prompt-engineering>. Accessed: [23-April-2024].
- [167] K. Oumoumen, F. Aboubane, and E.-C. Younes. “Automation of Historical Buildings: Historical Building Information Modeling (HBIM) based Virtual Reality (VR)”. In: *Proc. of 10th International Meeting on Mediterranean Architectural Heritage (RIPAM10)*. Vol. 40. 2024, pp. 319–322. DOI: [10.21741/9781644903117-34](https://doi.org/10.21741/9781644903117-34).
- [168] Jan Packer and Roy Ballantyne. “Conceptualizing the Visitor Experience: A Review of Literature and Development of a Multifaceted Model”. In: *Visitor Studies* 19.2 (2016), pp. 128–143. DOI: [10.1080/10645578.2016.1144023](https://doi.org/10.1080/10645578.2016.1144023).
- [169] Alessandra Pagliano and Barbara Ansaldi. “Digital Visualization and Multimedia for Cultural Heritage Accessibility: Designing “For All” Video-Tours at the Archaeological National Museum of Naples (MANN)”. In: *SCIENTIFIC RESEARCH and Information Technology* 13 (2 2023), pp. 85–102. DOI: [10.2423/i22394303v13n2p85](https://doi.org/10.2423/i22394303v13n2p85).
- [170] Mengxu Pan et al. “ELLMA-T: an Embodied LLM-agent for Supporting English Language Learning in Social VR”. In: *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. DIS ’25. New York, NY, USA: Association for Computing Machinery, 2025, pp. 576–594. DOI: [10.1145/3715336.3735786](https://doi.org/10.1145/3715336.3735786).
- [171] Marina Paolanti et al. “Evaluating Learning Outcomes of Virtual Reality Applications in Education: A Proposal for Digital Cultural Heritage”. In: *Journal on Computing and Cultural Heritage* 16.2 (2023), pp. 1–25. DOI: [10.1145/3593432](https://doi.org/10.1145/3593432).
- [172] Joon Sung Park et al. *Generative Agents: Interactive Simulacra of Human Behavior*. 2023. DOI: [10.48550/arXiv.2304.03442](https://doi.org/10.48550/arXiv.2304.03442).

- [173] T. Penjor et al. “Heritage Building Information Modelling (HBIM) for Heritage Conservation: Framework of Challenges, Gaps, and Existing limitations of HBIM”. In: *Digital Applications in Archaeology and Cultural Heritage* (2024), e00366. DOI: [10.1016/j.daach.2024.e00366](https://doi.org/10.1016/j.daach.2024.e00366).
- [174] Margarida F. Pereira et al. “Application of AR and VR in hand rehabilitation: A systematic review”. In: *Journal of Biomedical Informatics* 111 (2020), p. 103584. DOI: [10.1016/j.jbi.2020.103584](https://doi.org/10.1016/j.jbi.2020.103584).
- [175] Krzysztof Pietroszek et al. “Museum in Your Living Room: Recreating the Peace Corps Experience in Mixed Reality”. In: *2019 IEEE Games, Entertainment, Media Conference (GEM)*. 2019, pp. 1–4. DOI: [10.1109/GEM.2019.8811547](https://doi.org/10.1109/GEM.2019.8811547).
- [176] B. Joseph Pine and James H. Gilmore. *The Experience Economy: Work is Theatre & Every Business a Stage*. Boston, MA: Harvard Business Press, 1999.
- [177] Preloaded. *Curious Alice: The VR experience*. <https://preloaded.com/work/curious-alice/>. Accessed: [5-November-2024]. 2021.
- [178] Anna Puig et al. “Lessons learned from supplementing archaeological museum exhibitions with virtual reality”. In: *Virtual Reality* 24.2 (2020), pp. 343–358. DOI: [10.1007/s10055-019-00391-z](https://doi.org/10.1007/s10055-019-00391-z).
- [179] C. Pybus et al. “NEW REALITIES FOR CANADA’S PARLIAMENT: A WORKFLOW FOR PREPARING HERITAGE BIM FOR GAME ENGINES AND VIRTUAL REALITY”. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XLII-2/W15. 2019, pp. 945–952. DOI: [10.5194/isprs-archives-XLII-2-W15-945-2019](https://doi.org/10.5194/isprs-archives-XLII-2-W15-945-2019).
- [180] Grega Radež and Ciril Bohak. “Integrating Environmental Awareness Into NPCs: Contextual Conversational Interaction in Games”. en. In: *Proceedings of the 9th Human-Computer Interaction Slovenia (HCI SI) Conference 2024*. Ed. by Ciril Bohak, Luka Čehovin Zajc, and Veljko Pejović. Vol. 3866. CEUR Workshop Proceedings. Ljubljana, Slovenia: CEUR, 2024, pp. 11–29. URL: <https://ceur-ws.org/Vol-3866/#paper2>.
- [181] Sudha Rao et al. *Collaborative Quest Completion with LLM-driven Non-Player Characters in Minecraft*. 2024. DOI: [10.48550/arXiv.2407.03460](https://doi.org/10.48550/arXiv.2407.03460).
- [182] Gabrielle Rappolt-Schlichtmann and Samantha G. Daley. “Providing Access to Engagement in Learning: The Potential of Universal Design for Learning in Museum Design”. In: *Curator: The Museum J.* 56.3 (2013), pp. 307–321. DOI: [10.1111/cura.12030](https://doi.org/10.1111/cura.12030).

- [183] João Reis and Nuno Melão. “Digital transformation: A meta-review and guidelines for future research”. In: *Heliyon* 9.1 (2023), e12834. DOI: [10.1016/j.heliyon.2023.e12834](https://doi.org/10.1016/j.heliyon.2023.e12834).
- [184] Simone Restivo et al. “Interacting with Ancient Egypt Remains in High-Fidelity Virtual Reality Experiences”. In: *Proc. of Eurographics Workshop on Graphics and Cultural Heritage*. Ed. by Alberto Bucciero et al. The Eurographics Association, 2023. DOI: [10.2312/gch.20231175](https://doi.org/10.2312/gch.20231175).
- [185] Thomas Rist. “Using a Large Language Model to turn Explorations of Virtual 3D-Worlds into Interactive Narrative Experiences”. In: *2024 IEEE Conference on Games (CoG)*. 2024, pp. 1–8. DOI: [10.1109/CoG60054.2024.10645583](https://doi.org/10.1109/CoG60054.2024.10645583).
- [186] Rebecca Roberts et al. “Are Extended Reality Interventions Effective in Helping Autistic Children to Enhance Their Social Skills? A Systematic Review”. In: *Review J. of Autism and Developmental Disorders* 10.4 (2023), pp. 729–748. DOI: [10.1007/s40489-022-00320-y](https://doi.org/10.1007/s40489-022-00320-y).
- [187] Richard Sandell. “Social Inclusion, the Museum and the Dynamics of Sectoral Change”. In: *Museum & Society* 1.1 (2015), pp. 45–62. DOI: [10.29311/mas.v1i1.13](https://doi.org/10.29311/mas.v1i1.13).
- [188] Martijn J. Schuemie et al. “Research on Presence in Virtual Reality: A Survey”. In: *CyberPsychology & Behavior* 4.2 (2001), pp. 183–201. DOI: [10.1089/109493101300117884](https://doi.org/10.1089/109493101300117884).
- [189] Elmedin Selmanović et al. “Improving Accessibility to Intangible Cultural Heritage Preservation Using Virtual Reality”. In: *Journal on Computing and Cultural Heritage* 13.2 (2020), pp. 1–19. DOI: [10.1145/3377143](https://doi.org/10.1145/3377143).
- [190] Jinseok Seo, Gerard Jounghyun Kim, and Kyo Chul Kang. “Levels of detail (LOD) engineering of VR objects”. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. VRST ’99. New York, NY, USA: Association for Computing Machinery, 1999, pp. 104–110. ISBN: 978-1-58113-141-3. DOI: [10.1145/323663.323680](https://doi.org/10.1145/323663.323680). (Visited on 04/15/2026).
- [191] Hamza Shahab et al. “Virtual Reality in Museums: Does It Promote Visitor Enjoyment and Learning?” In: *International Journal of Human-Computer Interaction* 39.18 (2023), pp. 3586–3603. DOI: [10.1080/10447318.2022.2099399](https://doi.org/10.1080/10447318.2022.2099399).
- [192] Maria Shehade and Theopisti Stylianou-Lambert. “Virtual Reality in Museums: Exploring the Experiences of Museum Professionals”. In: *Applied Sciences* 10.11 (2020). DOI: [10.3390/app10114031](https://doi.org/10.3390/app10114031).

- [193] Anderson Augusto Simiscuka, Tejas Moreshwar Markande, and Gabriel-Miro Muntean. “Real-Virtual World Device Synchronization in a Cloud-Enabled Social Virtual Reality IoT Network”. In: *IEEE Access* 7 (2019), pp. 106588–106599. DOI: [10.1109/ACCESS.2019.2933014](https://doi.org/10.1109/ACCESS.2019.2933014).
- [194] Jaibir Singh et al. “A Systematic Review of Blockchain, AI, and Cloud Integration for Secure Digital Ecosystems”. en. In: *International Journal of Networked and Distributed Computing* 13.2 (2025), p. 28. DOI: [10.1007/s44227-025-00072-1](https://doi.org/10.1007/s44227-025-00072-1).
- [195] Maria Skublewska-Paszkowska et al. “3D technologies for intangible cultural heritage preservation—literature review for selected databases”. In: *Heritage Science* 10.1 (2022), pp. 1–24. DOI: [10.1186/s40494-021-00633-x](https://doi.org/10.1186/s40494-021-00633-x).
- [196] Mel Slater and Sylvia Wilbur. “A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments”. en. In: *Presence: Teleoperators and Virtual Environments* 6.6 (1997), pp. 603–616. DOI: [10.1162/pres.1997.6.6.603](https://doi.org/10.1162/pres.1997.6.6.603).
- [197] Smithsonian American Art Museum. *Beyond the Walls*. <https://americanart.si.edu/beyond-the-walls>. Accessed: [5-November-2024]. 2018.
- [198] Hak Jun Song et al. “The Influence of Tourist Experience on Perceived Value and Satisfaction with Temple Stays: The Experience Economy Theory”. In: *J. of Travel & Tourism Marketing* 32.4 (2015), pp. 401–415. DOI: [10.1080/10548408.2014.898606](https://doi.org/10.1080/10548408.2014.898606).
- [199] Cesar A. de Souza, Cláudia Szafr-Goldstein, and Annabeth Aagaard. “IoT in the Context of Digital Transformation and Business Model Innovation: the case of a traditional Brazilian wholesaler”. In: *2020 Global Internet of Things Summit (GIoTSummit)*. 2020, pp. 1–6. DOI: [10.1109/GIoTSummit49054.2020.9119527](https://doi.org/10.1109/GIoTSummit49054.2020.9119527).
- [200] R. Spallone et al. “3D MODELLING AND VIRTUAL REALITY FOR MUSEUM HERITAGE PRESENTATION: CONTEXTUALISATION OF SCULPTURE FROM THE TANG ERA”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-2/W4-2024 (2024), pp. 413–420. DOI: [10.5194/isprs-archives-XLVIII-2-W4-2024-413-2024](https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-413-2024).
- [201] Roberta Spallone et al. “Immersive Experiences for the Re-contextualization of Statues of the Goddess Sekhmet”. In: *Advances in Representation: New AI- and XR-Driven Transdisciplinarity*. Ed. by Andrea Giordano, Michele Russo, and Roberta Spallone. Cham: Springer Nature Switzerland, 2024, pp. 141–159. DOI: [10.1007/978-3-031-62963-1_9](https://doi.org/10.1007/978-3-031-62963-1_9).

- [202] Molly Follette Story. “Maximizing Usability: The Principles of Universal Design”. In: *Assistive Technology* 10.1 (June 1998), pp. 4–12. ISSN: 1040-0435. DOI: [10.1080/10400435.1998.10131955](https://doi.org/10.1080/10400435.1998.10131955).
- [203] K. Suhari et al. *Exploring BIM-based queries for retrieving cultural heritage semantic data*. PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-3812967/v1>]. 2024.
- [204] Tirill Bjørkeli Svaler. “On making libraries and museums more accessible for autistic people”. In: *IFLA J.* 50.1 (2024), pp. 42–52. DOI: [10.1177/03400352231202516](https://doi.org/10.1177/03400352231202516).
- [205] S. Sylaiou et al. “From physical to virtual art exhibitions and beyond: Survey and some issues for consideration for the metaverse”. In: *Journal of Cultural Heritage* 66 (2024), pp. 86–98. DOI: [10.1016/j.culher.2023.11.002](https://doi.org/10.1016/j.culher.2023.11.002).
- [206] Yiliu Tang et al. “LLM Integration in Extended Reality: A Comprehensive Review of Current Trends, Challenges, and Future Perspectives”. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI ’25. Association for Computing Machinery, 2025. DOI: [10.1145/3706598.3714224](https://doi.org/10.1145/3706598.3714224).
- [207] The Kremer Collection. *The Kremer Museum*. <https://thekremercollection.com/>. Accessed: [5-November-2024]. 2017.
- [208] The Solomon R. Guggenheim Foundation. *Guggenheim for All*. <https://www.guggenheim.org/accessibility/guggenheim-for-all>. [Accessed: 14 Jan 2025]. 2008.
- [209] Sam Theriault and Beth Redmond Jones. “Constructing Knowledge Together: Collaborating with and Understanding Young Adults with Autism”. In: *J. of Museum Education* 43.4 (2018), pp. 365–374. DOI: [10.1080/10598650.2018.1525657](https://doi.org/10.1080/10598650.2018.1525657).
- [210] Hugo Touvron et al. *LLaMA: Open and Efficient Foundation Language Models*. 2023. DOI: [10.48550/arXiv.2302.13971](https://doi.org/10.48550/arXiv.2302.13971).
- [211] Georgios Trichopoulos et al. “Crafting a Museum Guide Using ChatGPT4”. In: *Big Data and Cognitive Computing* 7.3 (2023). DOI: [10.3390/bdcc7030148](https://doi.org/10.3390/bdcc7030148).
- [212] Unity Technologies. *Unity glTF Fast Documentation*. <https://docs.unity3d.com/Packages/com.unity.cloud.glTFast@5.2/manual/index.html>. Accessed: [02-August-2024]. 2023.
- [213] Savvas Varitimidis et al. “Graph-Based Conversational AI: Towards a Distributed and Collaborative Multi-Chatbot Approach for Museums”. In: *Applied Sciences* 11.19 (2021). DOI: [10.3390/app11199160](https://doi.org/10.3390/app11199160).

- [214] Gregory Vial. “Understanding digital transformation: A review and a research agenda”. In: *The Journal of Strategic Information Systems*. SI: Review issue 28.2 (2019), pp. 118–144. DOI: [10.1016/j.jsis.2019.01.003](https://doi.org/10.1016/j.jsis.2019.01.003).
- [215] Rebeka Vital et al. “Comparison of extended reality platforms and tools for viewing and exhibiting art”. In: *Digital Applications in Archaeology and Cultural Heritage* 31 (2023), e00298. DOI: [10.1016/j.daach.2023.e00298](https://doi.org/10.1016/j.daach.2023.e00298).
- [216] Vitomir Jevremovic, Alexander Fuerer. *VR-All-Art*. <https://vrallart.com/>. Accessed: [27-October-2024]. 2024.
- [217] B. Wagner, J. Kramer, and M. Wenzel. *LINQ to XML overview*. <https://learn.microsoft.com/en-us/dotnet/standard/linq/linq-xml-overview>. Accessed: [05-August-2024]. 2021.
- [218] Hongyu Wan et al. “Building LLM-based AI Agents in Social Virtual Reality”. In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 1–7. DOI: [10.1145/3613905.3651026](https://doi.org/10.1145/3613905.3651026).
- [219] Katherine Wang, Simon J. Julier, and Youngjun Cho. “Attention-Based Applications in Extended Reality to Support Autistic Users: A Systematic Review”. In: *IEEE Access* 10 (2022), pp. 15574–15593. DOI: [10.1109/ACCESS.2022.3147726](https://doi.org/10.1109/ACCESS.2022.3147726).
- [220] Michael Wilkinson, Sean Brantley, and Jing Feng. “A Mini Review of Presence and Immersion in Virtual Reality”. EN. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 65.1 (2021), pp. 1099–1103. DOI: [10.1177/1071181321651148](https://doi.org/10.1177/1071181321651148).
- [221] Isabell Wohlgenannt, Alexander Simons, and Stefan Stieglitz. “Virtual Reality”. en. In: *Business & Information Systems Engineering* 62.5 (2020), pp. 455–461. DOI: [10.1007/s12599-020-00658-9](https://doi.org/10.1007/s12599-020-00658-9).
- [222] J.-Y. Wong et al. “BIM-VR Framework for Building Information Modelling in Engineering Education”. In: *International Journal of Interactive Mobile Technologies (IJIM)* 14 (2020), p. 15. DOI: [10.3991/ijim.v14i06.1339](https://doi.org/10.3991/ijim.v14i06.1339).
- [223] Zhiheng Xi et al. *The Rise and Potential of Large Language Model Based Agents: A Survey*. 2023. DOI: [10.48550/arXiv.2309.07864](https://doi.org/10.48550/arXiv.2309.07864).
- [224] Fu-Chia Yang, Kevin Duque, and Christos Mousas. “The Effects of Depth of Knowledge of a Virtual Agent”. In: *IEEE Transactions on Visualization and Computer Graphics* 30.11 (2024), pp. 7140–7151. DOI: [10.1109/TVCG.2024.3456148](https://doi.org/10.1109/TVCG.2024.3456148).
- [225] Fu-Chia Yang et al. “Embodied Conversational Agents in Extended Reality: A Systematic Review”. In: *IEEE Access* 13 (2025), pp. 79805–79824. DOI: [10.1109/ACCESS.2025.3566698](https://doi.org/10.1109/ACCESS.2025.3566698).

- [226] Syed Fawad Mustafa Zaidi, Callan Moore, and Himanshu Khanna. “Towards integration of user-centered designed tutorials for better virtual reality immersion”. In: *Proceedings of the 2nd International Conference on Image and Graphics Processing*. ICIGP '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 140–144. DOI: [10.1145/3313950.3313977](https://doi.org/10.1145/3313950.3313977).
- [227] Tatiana Zemskova and Dmitry Yudin. *3DGraphLLM: Combining Semantic Graphs and Large Language Models for 3D Scene Understanding*. 2025. DOI: [10.48550/arXiv.2412.18450](https://doi.org/10.48550/arXiv.2412.18450).
- [228] Xiaolong Zhao, Jinju Lee, and Kwanseon Hong. “A Comparative Analysis of Museum Accessibility in High-Density Asian Cities: Case Studies from Seoul and Tokyo”. In: *Buildings* 13.8 (2023), p. 1886. DOI: [10.3390/buildings13081886](https://doi.org/10.3390/buildings13081886).
- [229] Jiarui Zhu et al. “Free-form Conversation with Human and Symbolic Avatars in Mixed Reality”. In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2023, pp. 751–760. DOI: [10.1109/ISMAR59233.2023.00090](https://doi.org/10.1109/ISMAR59233.2023.00090).
- [230] Emmanouil Zidianakis et al. “The Invisible Museum: A User-Centric Platform for Creating Virtual 3D Exhibitions with VR Support”. In: *Electronics* 10.3 (2021). DOI: [10.3390/electronics10030363](https://doi.org/10.3390/electronics10030363).

This Ph.D. thesis has been typeset by means of the T_EX-system facilities. The typesetting engine was pdfL^AT_EX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete T_EX-system installation.