2025 IMEKO TC26 International Conference on
Metrology for Archaeology and Cultural Heritage
Bergamo, Italy, October 15-17, 2025

Cultural Heritage visualization through Generative AI: a Unity-based plugin for real-time 3D object creation

Attilio Della Greca[1], Ilaria Amaro[1], Giovanni Nocerino[2], Paola Barra[3]

[1]*University of Salerno, adellagreca@unisa.it, iamaro@unisa.it*
[2]*University of Naples Federico II, PLINIVS-LUPT Study Centre, giovanni.nocerino@unina.it*
[2]*Parthenope University of Naples, paola.barra@uniparthenope.it*

***Abstract* – This work introduces a plugin developed for Unity based on generative artificial intelligence (GAI), designed for the real-time creation of 3D objects from textual descriptions, for the preservation and digitization of cultural heritage. The proposed solution aims to overcome the limitations of current extended reality (XR) techniques, reducing production times and costs, and offering an alternative accessible even to non-expert users. The system is divided into three main phases: object definition, generation via multimodal AI pipeline and integration into the virtual environment. Preliminary tests demonstrate the plugin's ability to generate coherent, visually detailed and ready-to-use resources in immersive contexts. This approach promotes new ways of participatory fruition of heritage, encouraging interaction, accessibility and customization of digital cultural experiences.**

I. INTRODUCTION

In recent years, Virtual Reality (VR) has been widely studied and applied in various contexts, including industry, healthcare, education, and the conservation of architectural and cultural heritage [15][10] [6]. In particular, in the field of Cultural Heritage (CH) conservation, VR has enabled immersive and interactive experiences, overcoming the limitations of traditional tools, such as 2D simulators and photo archives. The evolution of XR technologies, including VR, Augmented Reality (AR), and Mixed Reality (MR), has also introduced new opportunities for the digitization, preservation, and transmission of tangible, intangible, and natural heritage, thanks to the integration with mobile devices, wearable systems, and advanced visualization systems [11][13].

Despite the transformative potential of VR technologies in the cultural heritage sector, their application continues to present significant challenges. Among these are the high production and development costs, the technical complexities related to the integration of software and hardware, and the barriers to accessibility for different groups of users [16].

Recurring concerns also include the need for continuous updates, the risk of digital obsolescence, the challenge of designing universally accessible and emotionally engaging experiences for all users, especially non-expert users [18][8].

A valuable alternative to overcome the current difficulties related to the construction and use of virtual experiences is Generative Artificial Intelligence (GAI). This technology has enabled the creation of 3D images in the VR field using vocal and text prompts, thereby enhancing immersion and personalization in digital experiences while reducing costs and generation times [14].

It is in this context that the following work is presented, in which we propose a Generative Artificial Intelligence plugin for Unity, initially conceived for therapeutic environments in the healthcare sector (as described in our paper [7]) and subsequently adapted for the cultural and architectural heritage sector.

The plugin, based on a multimodal AI generation pipeline, enables the real-time creation of 3D objects from textual descriptions, providing an accessible tool for the visualization and digital reconstruction of deteriorated, partially lost, or not physically accessible artistic and architectural elements.

The integration of this technology within cultural paths or professional training courses can promote a deeper and more personalized interaction with the contents, favoring emotionally engaging and cognitively stimulating experiences. The proposed plugin is, therefore, configured as a tool with a dual functionality: (i) a potential educational support in the context of academic training and (ii) a valuable tool to enhance artistic experiences, such as exhibitions and displays.

The images generated by our plugin can improve the understanding and analysis of cultural and architectural elements that would otherwise be difficult to access due to historical or geographical reasons.

II. RELATED WORKS

In recent years, the application of XR technologies in the field of Cultural Heritage has gained increasing interest in the scientific community. These technologies have been widely used to reconstruct archaeological sites, create virtual tours in museums and offer immersive educational experiences that enrich traditional forms of cultural transmission [2][4][3]. Virtual environments provide the means to simulate historical contexts, allowing users to interact dynamically with artifacts and environments. Traditional XR applications in Cultural Heritage, however, often rely on pre-built 3D assets, whose production requires considerable time, skills and resources [1]. This static approach limits customization and restricts user autonomy, especially in participatory and educational contexts. To address these challenges, research has increasingly focused on generative technologies capable of dynamically creating 3D content from user input [17]. Generative AI, in particular models integrating Natural Language Processing (NLP) with 3D object generation pipelines, represents a promising alternative for creating scalable content in virtual environments designed for the dissemination and preservation of Cultural Heritage [5].

Several recent works have exploited AI-generated images and models to improve active participation and user engagement during virtual experiences. In particular, the combination of text-to-image models allows even non-expert users to create complex virtual assets through simple text prompts [19]. These tools have also proven effective in other fields, such as mental health therapy, where the co-creation of meaningful objects can strengthen emotional engagement and user trust [7].

Digital storytelling and gamification are further areas where generative tools intertwine with Cultural Heritage. However, such systems still largely rely on manually curated content and lack real-time adaptability [9].

Our Unity plugin addresses these limitations by integrating a modular AI pipeline that translates textual inputs into fully textured 3D objects. This innovation is in line with current research trends that promote more inclusive and interactive approaches to heritage preservation, where users are encouraged not only to explore but also to contribute to the digital representation of cultural memory.

By integrating generative AI with XR applications in Cultural Heritage, this interdisciplinary work contributes to the development of sustainable experiences accessible to a wide range of users. [12].

III. METHODOLOGY

The proposed methodology, as shown in figure 1, describes the use of a generative AI plugin developed for Unity, adapted to the creation of 3D assets for applications in the Cultural Heritage (CH) sector. The system allows the real-time generation of digital artifacts starting from natural language descriptions, which can then be integrated into immersive virtual environments. The process is divided into three main phases: (1) definition of the object to be generated, (2) execution of the generation pipeline, and (3) integration and interaction within the XR scene.

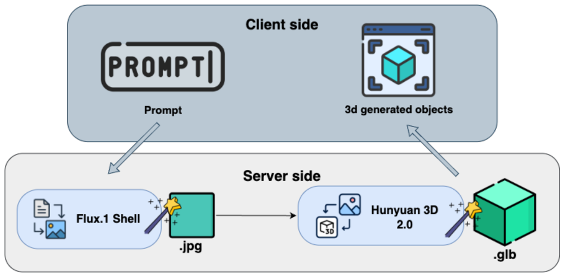

Fig. 1. Proposed creation pipeline performed by the Plugin

A. Phase 1: Definition of the object to be generated

In the first phase, domain experts or non-expert users define which object or element should be generated. This can include historical artifacts (e.g. ceramics, tools, furniture), symbolic objects (e.g. ritual objects, coats of arms), or scenographic elements (e.g. statues, architectural elements). The definition is expressed in natural language through a textual prompt that includes visual, historical, and stylistic details that will be integrated into the generated object.

Example prompt:
"Create a Doric column, with a fluted shaft without a base, a simple capital composed of an echinus and an abacus. The column must rest directly on the stylobate"

B. Phase 2: Generation via the AI plugin pipeline

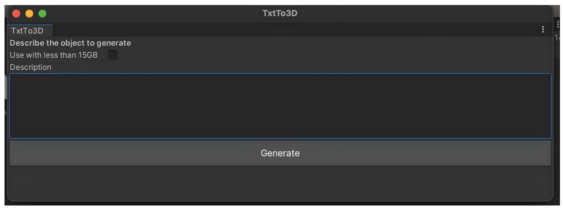

Fig. 2. User Interface of the plugin in Unity

The second phase concerns the actual generation of the 3D object via the Unity plugin, as shown in figure 2, based on a two-stage AI pipeline:

- Text-to-Image (T2I): First, the text prompt is processed by a diffusion-based model (FLUX.1-schnell), generating a high-resolution image that visually interprets the description. This image will serve as a starting point for the generation of the 3D object.
- Image-to-3D (I2M): In fact, the generated image is processed by a 3D reconstruction model (Hunyan3D 2.0), which produces a textured mesh. The resulting asset is compatible with real-time engines.

The plugin handles this process via a built-in API and isolated Conda environments. The overall generation process typically takes less than 30 seconds and produces ready-to-use 3D models, dramatically reducing development time. However, the generated models can be further improved and modified before being integrated into virtual environments.

C. Phase 3: Integration into Unity and use in XR context

Once generated, the object is automatically imported into the Unity scene and made available as a standard GameObject. In this context, it can be positioned, scaled and enriched with interactive elements via Unity's XR Interaction toolkit or other input systems.

This integration phase also includes runtime performance optimization (e.g. mesh simplification, occlusion elimination) and adaptation to AR/VR platforms (e.g. Meta Quest, Hololens).

IV. PRELIMINARY RESULTS AND DISCUSSION

To evaluate the applicability of the Unity plugin in Cultural Heritage contexts, we conducted generation tests based on historically and culturally significant object descriptions. Each test followed the three-phase methodology outlined in Section III: textual definition, AI-based generation, and Unity integration. The objective was to assess the plugin's ability to generate semantically coherent, visually detailed, and technically usable 3D objects suitable for XR applications.

In the work we presented two examples of 3D objects generated from the following prompts:

1. *"Create a Roman oil lamp in terracotta with a depiction of the god Mercury."*
2. *"Create a Corinthian capital. The capital should be decorated with two superimposed orders of stylised acanthus leaves, spiral corner volutes and a moulded, square top (abacus). The leaves should be carved with a high level of detail"*

Case 1: Roman Oil Lamp

The plugin successfully generated a terracotta-toned 3D model with accurate Roman styling and a relief representing Mercury. Surface textures included appropriate weathering and burn marks.

Case 2: Corinthian capital

The rendered figure shows a general similarity with the initial prompt. However, to obtain a more accurate representation, the proportions should be changed and the details of the entire 3D model should be improved, since the Corinthian capital was much more detailed and therefore more complex to generate. This example highlights the current limitations of the plugin that can be improved in future works achieving better accuracy. Despite the limitations highlighted, the plugin can be considered at present as a support tool for designers as it reduces the generation time of 3D models that can be subsequently edited and improved.

Fig. 3. 2D representation of the Roman Oil Lamp

Fig. 4. Final result of the Roman Oil Lamp with and without texture

V. CONCLUSION

This study presented a generative AI-based Unity plugin designed for real-time 3D content creation in cultural heritage applications. The proposed system enables users to create culturally relevant digital artifacts from simple text prompts, eliminating the need for advanced 3D modeling skills. The adopted methodology, divided into phases of definition, generation, and integration, offers a simplified

Fig. 5. 2D representation of the Corinthian capital

Fig. 6. Final result of the Corinthian capital with and without texture

and scalable operational framework aimed at creating immersive cultural and educational experiences. Preliminary results confirm the effectiveness of the plugin in producing detailed 3D objects optimized for XR environments. However, several potential improvements were observed, including the correspondence between prompts and generated content and the fidelity in representing complex objects.

Despite the limitations encountered, the model proposed in our work promotes a transition towards more interactive, inclusive, and emotionally engaging ways of experiencing cultural heritage while fostering active user participation and digital accessibility in the processes of cultural heritage conservation and valorization.

Table 1. Technical characteristics of the generated 3D meshes

Object	**Vertices**	**Gen-Time (s)**
Oil Lamp	24,179	21.3
Corinthian capital	32,562	28.6

REFERENCES

[1] Alonzo C Addison. Emerging trends in virtual heritage. *IEEE multimedia*, 7(2):22–25, 2002.

[2] Fabio Bruno, Stefano Bruno, Giovanna De Sensi, Maria-Laura Luchi, Stefania Mancuso, and Maurizio Muzzupappa. From 3d reconstruction to virtual reality: A complete methodology for digital archaeological exhibition. *Journal of Cultural Heritage*, 11(1):42–49, 2010.

[3] Mario Casillo, Francesco Colace, Brij B Gupta, Angelo Lorusso, Domenico Santaniello, and Carmine Valentino. The role of ai in improving interaction with cultural heritage: An overview. *Handbook of Research on AI and ML for Intelligent Machines and Systems*, pages 107–136, 2024.

[4] Mario Casillo, Francesco Colace, Angelo Lorusso, Domenico Santaniello, and Carmine Valentino. Integrating physical and virtual experiences in cultural tourism: an adaptive multimodal recommender system. *IEEE Access*, 2025.

[5] Kevin Chen, Christopher B Choy, Manolis Savva, Angel X Chang, Thomas Funkhouser, and Silvio Savarese. Text2shape: Generating shapes from natural language by learning joint embeddings. In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*, pages 100–116. Springer, 2019.

[6] Chris Christou. Virtual reality in education. In *Affective, interactive and cognitive methods for e-learning design: creating an optimal education experience*, pages 228–243. IGI Global Scientific Publishing, 2010.

[7] Attilio Della Greca, Ilaria Amaro, Paola Barra, Emanuele Rosapepe, and Genoveffa Tortora. Enhancing therapeutic engagement in mental health through virtual reality and generative ai: a co-creation approach to trust building. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 6805–6811. IEEE, 2024.

[8] Attilio Della Greca, Ilaria Amaro, Nicola Frugieri, Paola Barra, and Genoveffa Tortora. The impact of virtual scenarios on empathy: a user study on the role of empathic abilities and environmental context in emotional facial expression replication. In *2025 IEEE Conference on Virtual Reality and 3D User In-*

terfaces Abstracts and Workshops (VRW), pages 574–581. IEEE, 2025.

[9] Maria Economou. Heritage in the digital age. *A companion to heritage studies*, pages 215–228, 2015.

[10] Elliot Hu-Au and Joey J Lee. Virtual reality in education: a tool for learning in the experience age. *International Journal of Innovation in Education*, 4(4):215–226, 2017.

[11] Chiara Innocente, Luca Ulrich, Sandro Moos, and Enrico Vezzetti. A framework study on the use of immersive xr technologies in the cultural heritage domain. *Journal of Cultural Heritage*, 62:268–283, 2023.

[12] Thomas Kersten, Felix Tschirschwitz, and Simon Deggim. Development of a virtual museum including a 4d presentation of building history in virtual reality. In *TC II & CIPA 3D Virtual Reconstruction and Visualization of Complex Architectures, 1–3 March 2017, Nafplio, Greece*, pages 361–367. Copernicus, 2017.

[13] Chenming Lin, Guobin Xia, Farnaz Nickpour, and Yinshan Chen. A review of emotional design in extended reality for the preservation of culture heritage. *npj Heritage Science*, 13(1):86, 2025.

[14] Zhihan Lv. Generative artificial intelligence in the metaverse era. *Cognitive Robotics*, 3:208–217, 2023.

[15] Fabrizia Mantovani, Gianluca Castelnuovo, Andrea Gaggioli, and Giuseppe Riva. Virtual reality training for health-care professionals. *CyberPsychology & Behavior*, 6(4):389–395, 2003.

[16] Valeria Minucciani[1], Michela Benente[1], Francesco Strada, and Andrea Bottino. Virtual reality for cultural heritage: Emotional involvement and design for all. *Design for Inclusion*, page 12, 2024.

[17] Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. Point-e: A system for generating 3d point clouds from complex prompts. *arXiv preprint arXiv:2212.08751*, 2022.

[18] Giovanni Nocerino and Mattia Federico Leone. Workerbee: A 3d modelling tool for climate resilient urban development. In *INTERNATIONAL SYMPOSIUM: New Metropolitan Perspectives*, pages 16–26. Springer, 2024.

[19] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.