

Is OpenAI the future for historical 3D reconstruction? Analysis of a suitable pipeline applied to the Iuvanum case study

Vincenzo Saverio Alfio^{1,*†}, Alessia Amelio^{1,†}, Massimiliano Pepe^{1,†}, Domenica Costantino^{2,†}

¹ University "G. d'Annunzio" of Chieti-Pescara, InGeo Department viale Pindaro 42, 65127 Pescara, Italy

² Polytechnic University of Bari, DICATECh Department, via E. Orabona, 70125, Bari, Italy

Abstract

The aim of this manuscript is to generate three-dimensional models of historical and archaeological heritage using Artificial Intelligence models specialised in text-to-3D. In particular, using ChatGPT as a semantic interpreter, the system extracts structured spatial, architectural and material information, which is then translated into input prompts or parameters for generative 3D modelling tools. Next, a process of textual and/or visual concept extraction is initiated, i.e. elements that can be described textually or represented graphically are recognised and, by combining natural language processing through Large Language Models, after a series of improvements through textual prompts, the data feeds AI platforms specialised in generating 3D models with AI-based geometric modelling systems. This approach allows the generation of detailed 3D reconstructions of archaeological sites where only partial remains or fragmentary structural traces are evident. By exploiting the interpretative potential of textual historical and archaeological sources, the method bridges the gap between incomplete material evidence and a coherent spatial representation, offering new opportunities for both scientific analysis and the enhancement and dissemination of Cultural Heritage.

Keywords

Artificial Intelligence, 3D Modelling, Cultural Heritage, Image-to-3D, ChatGPT, Meshy.ai

1. Introduction

In recent years, Artificial Intelligence (AI) has seen a sharp increase, transforming itself from a simple computational tool to a technology capable of learning, interpreting, and generating complex content. The fields of application are varied and apply in contexts such as natural language processing, Computer Vision (CV), medicine, engineering, etc., demonstrating a cross-cutting approach in many other sectors. Furthermore, an innovative aspect of this evolution is the introduction of the Chain of Thought (CoT), a technique that allows models to explicitly expose their reasoning steps during problem solving. This approach has proven particularly effective in increasing the ability of AI models to tackle tasks that require multi-step deductions, articulated explanations, and the generation of coherent and well-structured content, improving the performance of Large Language Models (LLMs) [1]. In this context, the application of AI in the field of Cultural Heritage (CH) promises great developments [2]; in fact, the ability to analyse historical texts, specialist documents, publications, etc., not only to extract information, but also to organise it according to transparent logical processes, allows for the construction of more accurate and verifiable descriptions [3]. Furthermore, AI models combined with Deep Learning (DL) and CV techniques can support professionals in discovery, description, classification, and conservation activities [4] as well as in the inspection and evaluation of historical architectures using image

^{*}VIPERC '25: 4th International Conference on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, 3-4 December 2025

^{1*} Corresponding author.

[†] These authors contributed equally.

✉ vincenzosaverio.alfio@poliba.it (V.S. Alfio); alessia.amelio@unich.it (A. Amelio); massimiliano.pepe@unich.it (M. Pepe); domenica.costantino@poliba.it (D. Costantino)

ORCID 0000-0003-4816-8913 (V.S. Alfio); 0000-0002-3568-636X (A. Amelio); 0000-0003-2508-5066 (M. Pepe) 0000-0002-1909-9261 (D. Costantino).



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

processing techniques and visual interpretation assistance [5]. In the context of CH, there are several applications of AI for the creation of 3D models. Stoean et al., 2024 analysed a DL approach in the field of CH restoration and conservation using the Neural Radiance Field (NeRF) model for 3D rendering, combined with the Stable Diffusion (SD) architecture to recreate part of a degraded ceramic artefact [6]. Jaramillo et al., 2024 explore the use of recent generative AI algorithms for the restoration of CH objects, exploiting a conditional diffusion model; this model is capable of reconstructing 3D point clouds and generating missing parts by predicting their geometry with high accuracy [7]. In the field of text-to-image, Arzomand et al., 2024 explore the use of AI and, in particular, text-to-image generation technologies for the reconstruction of damaged historical sites. Starting from a data collection, a series of images are generated by AI, which are then reviewed by experts and compared with historical sources to assess the accuracy of the reconstruction. These images represent a new approach and support in the field of historical heritage conservation [8].

In this context, the aim of this work is to develop a methodological approach and evaluate the potential, reliability and applicability of OpenAI and Meshy.ai for historical reconstruction and automatic generation of 3D models of CH. Establishing a collection of bibliographic, descriptive or iconographic data, the process exploits the generative capabilities of OpenAI to produce texts or images consistent with the cultural asset in question. These outputs are then imported into Meshy.ai, an automatic 3D model generation platform that enables the digital reconstruction of historical objects or environments. This methodology aims to facilitate the enhancement, preservation and digital enjoyment of CH, even in the absence of complete physical documentation.

2. Case study

In order to apply the methodological approach shown in this manuscript, some architectural remains at the Iuvanum site (see Figure 1) were analysed as part of the Lip3D project (Living forever the Past through a 3Digital world), funded by the Digital Europe programme. This case study was chosen due to the potential application of the digital methodologies tested within the project [9]. The origins of the archaeological site of Iuvanum date back to the Bronze Age, and the remains of several architectural elements can be recognised within it, such as the forum, the temple, the church and the theatre. The acropolis housed two Samnite temples built at different times and dating back to the 2nd century BC. Later, around 1000 AD, an abbey was built using the remains of the ancient city and temple.

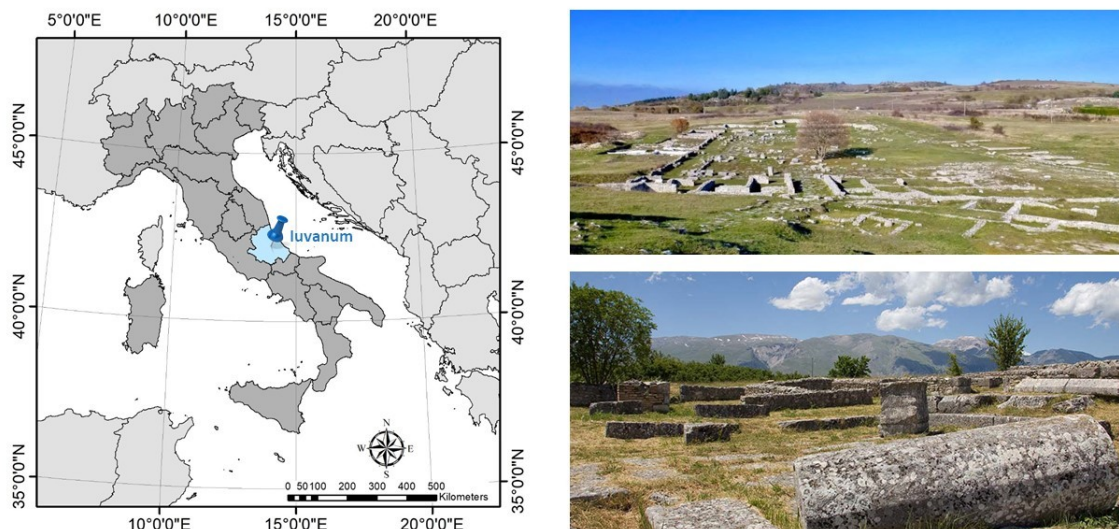


Figure 1: Location of the study area and view of the archaeological site of Iuvanum, Italy.

In this case study, the focus is on the two temples (named Temple A and Temple B), and various historical sources have been consulted and analysed in order to identify the proportions, architectural orders and styles characteristic of a specific historical period. The two temples were built during the early stages of the archaeological site's development, specifically during the original settlement of the Samnites, with traces dating back to the Bronze Age (between 2300 BC and 950-800 BC). In addition to the two temples side by side on the podium on the hill, a small theatre was built in the south-east of the hill, below the temple area, and a sacellum at the rear of temple B [10, 11]. Finally, regarding architectural order and styles, it has been suggested that the presence of columns with bases deviates from the canonical Tuscan style (traditionally without bases) and instead resembles an Italic Doric style with marked local influences. This indicates that the temple adopted a Samnite variant of the Doric order rather than adhering to a pure Tuscan model.

3. Material and Methods

The generation of a 3D model and its export to compatible formats (e.g. .obj) from a textual description requires a pipeline consisting of several tools. The approach to this task consists of four main steps: i) bibliographic collection and document optimisation, ii) text and/or image extraction, iii) 3D model generation, iv) file export.

In the first phase, documents, manuscripts, reports, and other documentation necessary to describe the object or artefact from a historical, architectural, and dimensional point of view are collected. This documentation is analysed by AI models. In particular, for the analysis of this case study, the GPT-4o version was used, which, compared to the previous one, introduces significant improvements in terms of speed, comprehension, and more accurate responses by better executing complex instructions. Subsequently, based on the interpretation of the uploaded documentation, a textual description of the object is extracted, appropriately calibrated according to the desired information, or a descriptive image can be generated. These outputs are then uploaded to AI platforms specialised in text-to-3D or image-to-3D for the generation of 3D models of the analysed architectural elements. In this work, the AI platform chosen is Meshy.ai, as it allows 3D models to be generated from text or image inputs, greatly simplifying workflows in the creation and download of three-dimensional content in standard formats. With this platform, it is also possible to download the created models in the appropriate export formats and via webapp or local pipeline. Meshy.ai, an AI platform that allows 3D models to be generated from text or image inputs, was chosen for the generation of 3D models, greatly simplifying workflows in the creation and download of three-dimensional content in standard formats. Furthermore, in order to test the functionality of the AI models, the methodological approach described has been divided into three separate frameworks.

In particular, in Framework 1, no bibliographic references are analysed by the AI model; only the generation of text or images from a text prompt is required.

In Framework 2, on the other hand, the AI model analyses scientific texts, papers, and other documentary information in order to analyse them semantically and extract concepts, cause-effect relationships, objectives, hypotheses and, where appropriate, technical aspects.

Finally, Framework 3 integrates the information from the previous framework with the data and information obtained from graphical, historical, and architectural analyses generated during the survey and three-dimensional modelling phase. In this way, the AI model will also be able to process coherent descriptions and perform proportional analyses between the different architectural components of the analysed assets, i.e., generate schematic or stylised representations that respect the proportions and relationships between the different elements.

In order to validate the process, the 3D models obtained from the different frameworks will be compared with a 3D model (3D model reconstructed manually and in a controlled manner based on historical information).

This model was obtained by integrating 3D surveys, proportions, and architectural rules as reference modules, ensuring a final 3D model that is historically consistent and visually faithful to the original context. The methodological approach described is schematised in the Figure 2.

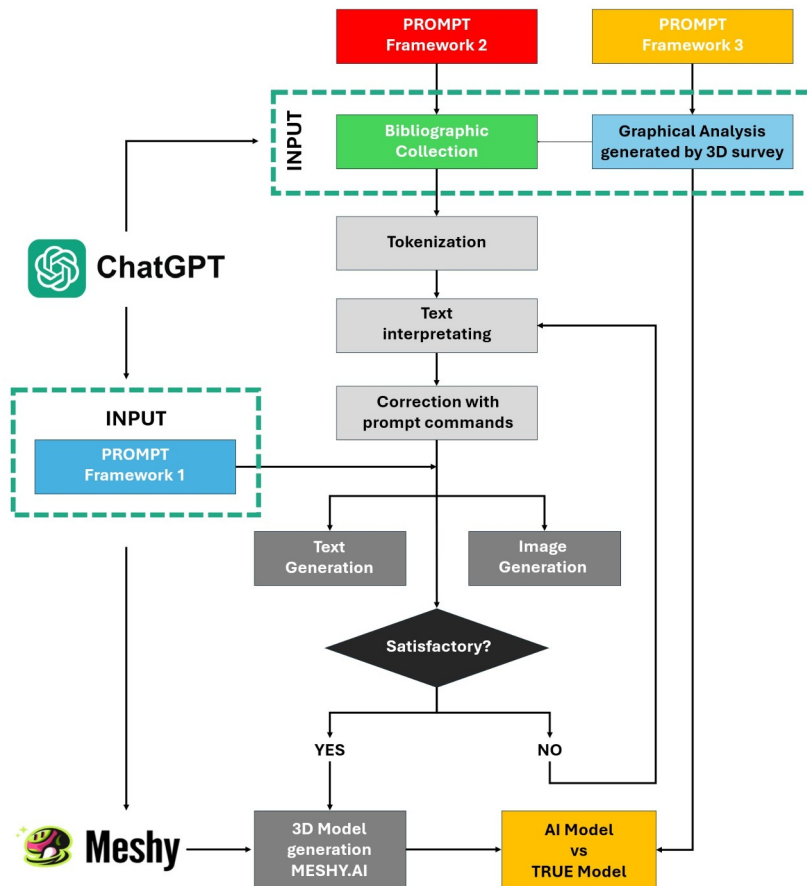


Figure 2: Schematic representation of the methodological approach used, which includes the phases of bibliographic collection, analysis, and modelling with AI platforms.

3.1. Bibliographic collection

The preliminary collection of documentation useful for generating an interpretative text is the fundamental preliminary phase of this approach, as it allows all the information necessary to build a solid knowledge base of the object under investigation to be found. This activity involves identifying relevant sources, carefully selecting the materials to be interpreted, and verifying the documents according to their type and intended use. An organised bibliographic collection guarantees excellent results in the subsequent stages, avoiding interpretation errors and ensuring that the final objectives of 3D modelling are met.

3.2. Reading and interpreting texts

At this stage, ChatGPT was used, a virtual assistant developed by OpenAI and based on GPT (Generative Pre-trained Transformer) architecture, an advanced linguistic model capable of understanding and generating natural language in a coherent and relevant manner. The system, pre-trained on a large corpus of public texts and subsequently optimised using supervised and reinforcement learning techniques, produces responses through four main stages: input acquisition, processing and interpretation, response generation, and textual output [12]. Without direct access to the internet, consciousness, or emotions, the model can nevertheless generate errors and has knowledge limited to a specific update date. Furthermore, ChatGPT's architecture is based on a Transformer neural network, designed to analyse semantic relationships between words and predict the next word in a sequence. Training utilises large amounts of text and the Reinforcement Learning from Human Feedback (RLHF) approach, in which human evaluations guide the qualitative improvement of the model [13, 14]. The system processes language through tokens, elementary linguistic units converted into numerical values [15], managed within a context window of variable size, which in more advanced models, such as GPT-4, allows for the processing of extended texts. Thanks to self-attention and positional encoding mechanisms [16], the model identifies relationships and dependencies between words, generating coherent texts based on statistical associations rather than semantic understanding. Finally, the ability to process multimodal inputs (text and images) increases its flexibility and naturalness in dialogue and language generation tasks [17].

3.3. Text and image generation

After reading a document, the text is divided into tokens, i.e. linguistic units that allow natural language to be translated into numerical data that can be processed by the model. ChatGPT then applies its linguistic capabilities to analyse the semantic context, identify logical structures and conceptual relationships, and provide relevant answers. The context window limits the amount of text that can be handled in a single processing, although advanced models such as GPT-4 can handle large documents. In the presence of descriptive texts, the model can also initiate a process of visual transposition: after semantic analysis and visual concept extraction, ChatGPT processes a visual prompt for an image generator, such as DALL-E [18, 19], which uses convolutional neural networks or diffusion models to create coherent representations, respecting the spatial relationships and semantic style of the source text.

3.4. 3D modelling with Meshy.ai

3D modelling on the Meshy.ai platform, based on text or images as input, is also divided into several stages. In particular, about text input, the first stage involves semantic analysis of the

prompt i.e. Natural Language Processing (NLP), where the system understands the required elements (objects, materials, shapes, proportions) and creates a structured representation of the object. The text is then converted into a vector embedding, i.e. a series of numbers representing the visual meaning of the prompt. This embedding is projected into a latent 3D space, i.e. a mathematical space in which each point represents a potential 3D shape, using models such as Shap-E (OpenAI), DreamFusion (Google) and GET3D (NVIDIA). In this step, the system imagines the shape of the object based solely on the description, leveraging knowledge learned from millions of text examples, and generates a rough volumetric shape, usually using techniques such as Neural Radiance Fields (NeRF), Voxel Grids (three-dimensional matrices) or Signed Distance Functions (SDF) [20, 21, 22]. The result is a 3D “mass cloud” that has not yet been converted into a mesh. At this point, it is possible to convert the volume into a real 3D mesh, formed by vertices (points in 3D space), faces (triangles or polygons that join the vertices) and normals (to indicate the directions of the surfaces) and proceed with the texturing and, if necessary, final export of the model.

If the input is an image, it is processed with depth estimation models to calculate depth, followed by segmentation to isolate the main object. In this way, the object is reconstructed as a partial 3D map. Next, the three-dimensional shape is generated using a 3D generative model based on deep neural networks with typical tools such as NeRF for volumetric representation of the scene, Point cloud generation for point clouds or Mesh generation for conversion into geometry. The model generates a latent three-dimensional representation, i.e. a 3D point cloud that is transformed into concrete geometry (mesh). The algorithm also imagines hidden parts and uses statistical data and large generative 3D models to complete the object (3D completion or hallucination). Once a geometric representation has been obtained, a triangular mesh is constructed, and the result is transformed into polygonal meshes; normals are added for lighting and rendering, and textures are applied.

4. Results

4.1. Framework No.1

For Framework No. 1, we asked ChatGPT to generate an image representing a historical reconstruction of the two temples at the archaeological site of Iuvanum. As a result of the process, the AI model produced two images, the first of which (Figure 3a) presented several discrepancies with respect to the analysed structure (number of columns, architectural styles, etc.). After a few minutes of reasoning and consulting some online bibliographic sources, the AI model generated a second image (Figure 3b) that was more realistic and considered the characteristics schematised in Table 1.

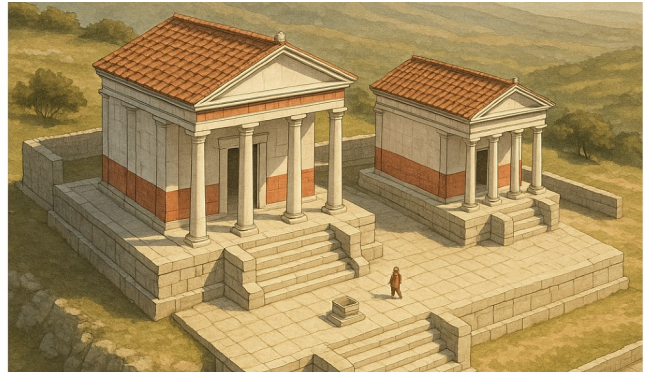
Table 1
Information extracted from the ChatGPT platform for image generation for Framework No 1.

Elements	Image Generation
Temple A	Rebuilt as a Doric tetrastyle temple in antis/prostyle on a high podium in square form covered with cladding, with a floor plan measuring approximately 21.3 × 12.6 m (reconstructed measurements), with central front entrance; traces of stone cladding (travertine) have been found.
Temple B	Built shortly afterwards (also in the 2nd century BC) about 4 m from the main temple; today, only the podium remains, so the architectural design (reduced front colonnade, lower elevation) is hypoal and indicative.
Note	The temple complex was part of the sacred area of the acropolis, which

included at least two religious buildings and a sacellum behind them, in visual contact with the forum and the theatre below in the Romanised municipal town after the Social War. The colours of the plaster (white with a red base), tile roofing and decorative details are plausible comparative reconstructions based on similar Italic-Roman sanctuaries from the same period; there is no definitive evidence for their exact colours at Iuvanum.



(a)

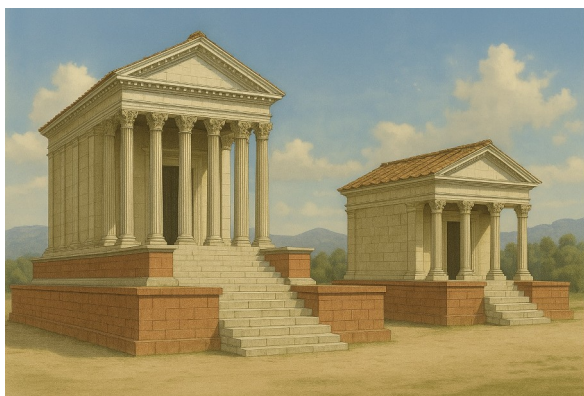


(b)

Figure 3: Framework No. 1: image generation after the request prompt (a) and final image with the requested corrections applied (b).

4.2. Framework No.2

For framework No.2, OpenAI was asked to generate an image by analysing some scientific texts describing the archaeological site of Iuvanum. In this case, compared to the previous one, the image generation was more accurate. Despite this, in the first image generated, the number, style and position of the columns in Temple A are inconsistent (Figure 4a). For this reason, manual prompts were used to request that these errors be modified and corrected. After the image was corrected, the result of this second approach is shown in Figure 4b.



(a)



(b)

Figure 4: Framework No. 2: generation of the first image with obvious semantic and dimensional errors (a) and final image generated after the requested corrections (b).

4.3. Framework No.3

After compiling a bibliography of scientific texts, supplemented by an architectural, stylistic and dimensional description of the archaeological site of Iuvanum, with particular focus on Temple A and Temple B, the following text prompt was entered into the ChatGPT platform: “*Analyse the documents and extract an image of Temple A and Temple B*”. Although an image of excellent graphic and interpretative quality was generated, some architectural elements, such as the capital, do not accurately reflect what is described in the input text. For this reason, again using text prompts, we asked the AI to analyse the text provided in detail, outline the main information for the reconstruction of the temples (see Table 2) and improve the image generation phase.

Table 2

Schematisation of information extracted from the ChatGPT platform for the generation of images.

Temple	Elements to be reconstructed	Suggested Style
Temple A	2.5 m high podium, prostyle colonnade with 4 columns (5.95 m × 0.85 m), balustrade, architrave, frieze and pediment with decorations, double torus at the bases, steps.	Combine wireframe and painting to highlight details and materials.
Temple B	2.05 m podium with steps, platform, light reconstruction of parapets and hypothetical cell measuring approximately 9 × 5 m.	Simple reconstruction style, in line with the “jungle” or clear pictorial image.

Furthermore, as can be seen in Table 2, during the image generation process, the AI model was able to analyse the data provided by interpreting not only the semantic content but also the dimensional relationships between the elements described. This aspect is particularly important because it was possible to generate coherent visual representations in which the proportions between objects are faithfully respected. In fact, the image produced is not only a visualisation of the objects but also a structured expression of the information contained in the input data, suitable for use in subsequent application contexts. The result is shown in Figure 5a, where it is possible to see the correspondence with the architectural styles described in the input text and the representation of both time periods. In this further generation, however, the mutual orientation of the time periods was not well interpreted, despite this information being described. For this reason, through a further session of textual prompts, additional information on the layout of the structures and their distance was provided to the AI model.

Through the manual prompt, it was requested, for example, to add columns in an exact manner consistent with the analysed text or to modify the orientation of the temple. For example, the prompt used in the latter case was: “*The main elevations of Temple A (the largest) and Temple B (the smallest) are aligned and placed on the same plane as viewed by the viewer. Regenerate image*”, and the result is shown in Figure 5b.



Figure 5: Result of the image generation phase: initial reconstruction with temple B rotated (a) and final congruent reconstruction (b).

After importing the image onto the AI Meshy platform, the image obtained from the OpenAI platform was uploaded to follow the Image-to-3D process. The modelling parameters were entered in the settings, namely the number of triangles that make up the mesh model and texture generation. After choosing the best model representing the case study, the final 3D model was generated. As can be seen from Figure 6a, the model is fairly faithful to the processed image. Despite this, there are several topological and structural errors in the mesh, which therefore require a further retopology process to improve the final result.

Furthermore, especially in Temple B, the reconstructions of the interior of the structure have not been faithfully reproduced; in fact, it can be seen that the AI algorithms were unable to generate closed environments. For this reason, it was decided to test the functionality of the AI modelling platform by focusing on a single structure at a time (Figure 6b). In this case, although some errors in the modelling of certain architectural details are still evident, the result appears to be better than the previous one in terms of both topology and mesh structuring (Figure 6c).

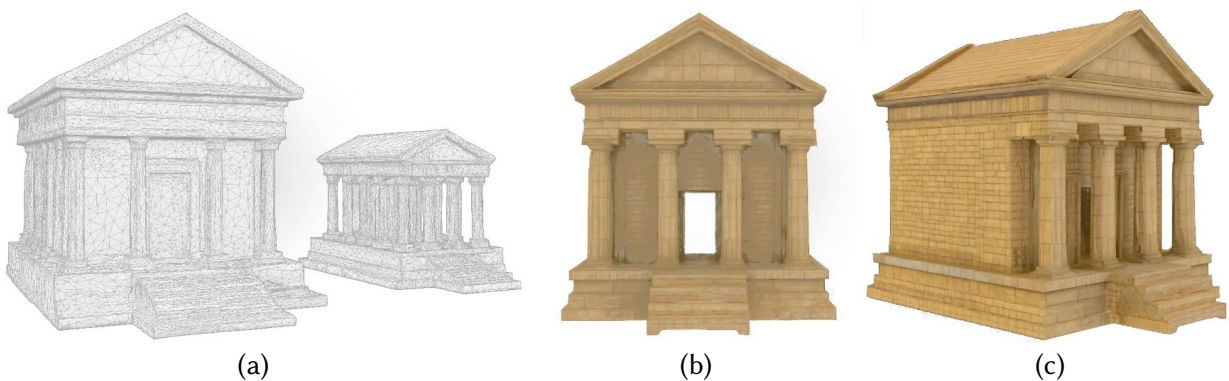


Figure 6: Image-to-3D process in Meshy.ai: generation of the 3D model of temples A and B (a), front view (b) and axonometric view (c) of temple A only.

4.4. Accuracy control and quantitative validation

One limitation encountered at this stage concerns the metric aspect of the automatically generated 3D models: although the proportions between architectural elements are generally consistent with the descriptions contained in the scientific texts analysed with AI, the absolute dimensions do not correspond to the actual metric data. This misalignment is attributable to the nature of the generative algorithm, which produces models from visual or descriptive inputs without a standardised metric reference.

With this methodological approach, in order to ensure the geometric accuracy of the models with respect to the actual dimensions obtained from 3D surveys, it is therefore necessary to implement an additional phase after the generation of the 3D models. This data post-processing phase must include a plane rotation and translation operation for the correct orientation of the model in three-dimensional space, as well as the definition of a uniform scale factor calculated on the basis of measurements resulting from the survey. In this way, it is possible to realign the virtual model to the actual dimensions while preserving the original proportions. Accurate algorithms for comparing datasets, such as Cloud to Mesh (C2M) and Multiscale Model to Model Cloud Comparison (M3C2), can validate the final model from a metric point of view.

After this post-processing phase, together with retopology applications and final data validation, it will then be possible to improve the final quality of the 3D mesh, and the resulting model can be used not only in reconstruction or educational contexts but can also be integrated into H-BIM (Heritage Building Information Modelling) platforms or GIS (Geographic Information System) systems in the field of monitoring, enhancement and promotion of CH.

5. Discussion

The introduction of basic historical, architectural, and metric information has proved to be an essential element in the methodological approach proposed for the automatic generation of 3D models for archaeological contexts. The quality and accuracy of the final result depend largely on the density and reliability of the documentary sources used in the preliminary phase, which play a key role in computational processing. An adequate historical and architectural framework of the context is essential not only to ensure the morphological consistency of the model, but also to correctly support the interpretation of the reconstruction data.

In this sense, the proposed methodology, based on AI and automatic generation from textual data or images, represents a significant advance over traditional three-dimensional modelling processes, which are often manual and highly time-consuming and resource intensive. However, it cannot yet fully replace human intervention, which remains essential in several stages: from the critical selection of sources to interpretative verification and the finishing of the model. In particular, the intervention of a specialist is necessary to avoid automatic reading errors or simplifications that could lead to interpretative errors.

This study aims to test an approach based on historical information through an iterative method and the development of applications consolidated in literature. Each element of the structure is constructed, through the controlled use of AI [23], based on historical and archaeological texts, as well as defined by architectural and geometric rules established by the user, which help to define dimensions and proportions consistent with the period in which the artefact was created. Furthermore, in the case study analysed, as the two temples are not physically present (only some remains of the columns and perimeter walls are present), the final validation of the model is therefore carried out through the objective verification of the correct proportions and the correct use of architectural styles [24].

From a computational point of view, image processing times in the various frameworks were encouraging. In fact, the average waiting time for image generation varied from a minimum of about 20 seconds to a maximum of a few minutes. This is a variable factor that depends on the complexity of the requested image and the interaction with the inference server that executes the diffusion model. In general, the response time in image generation increased significantly during the image editing phase following the insertion of a text prompt to modify or correct certain elements. Nevertheless, in order to assess its applicability in the real world, it is necessary to analyse the consistency of processing times and their scalability in operational contexts. Although

the variability observed is acceptable in the experimental phase, use in production scenarios requires greater stability and predictability of performance. This implies the optimisation of computational resources and inference models to ensure efficient integration into real workflows.

Furthermore, regarding the generation of 3D models, the methodology still has some significant technical limitations. These include the need for retopology operations to optimise the model meshes for subsequent integration into other platforms. Furthermore, as already discussed, it is essential to scale the model based on 3D survey data, without which the models cannot be used for design purposes. Finally, another important aspect is related to the georeferencing of the model, as this not only allows the model to be integrated into multidisciplinary and interoperable environments such as BIM platforms or GIS systems, but also into virtual reconstruction or augmented reality projects, with potential applications in the field of public use and heritage enhancement.

6. Conclusion

The methodological approach tested has demonstrated how the integration of generative AI techniques and automatic modelling tools can provide valuable support in the digitisation and reconstruction of CH structures. The pipeline based on textual analysis using NLP models (ChatGPT), the generation of architectural images and their subsequent conversion into three-dimensional models (via Meshy.ai) allows complex scientific content to be translated into synthetic and navigable digital representations. From an operational point of view, the use of iterative prompts has made it possible to progressively improve the consistency between descriptive data and generated images, with particular attention to proportions, spatial orientations and stylistic details. However, the need to refine the textual instructions has emerged in order to compensate for the current interpretative limitations of the models, especially in cases where complex spatial relationships or the generation of interior environments are required.

The subsequent 3D conversion phase showed promising results, but also critical issues related to the topological quality of the generated meshes. In particular, the absence of volumetric closure and the presence of non-optimised geometries require post-processing steps (retopology, refinement) to obtain models suitable for advanced use (e.g. virtual reality, 3D printing, simulations). The analysis also highlighted greater effectiveness in the process when the focus is on individual objects or structures. Overall, the methodology proves to be a solid basis for a semi-automatic process of digital reconstruction of CH, although it still requires skilled human intervention for the validation and optimisation of results. The approach is scalable, repeatable, and easily adaptable to different case studies, representing an interesting solution for projects involving the documentation, enhancement and communication of historical and architectural heritage.

Furthermore, based on the results obtained, 3D models generated using generative AI techniques can also be a strategic resource for integration into BIM/GIS platforms. In the BIM field, these models can be used as a basis for the semantic and informative reconstruction of historic buildings, supporting conservation design, digital management of built heritage and intervention planning. In GIS, integration with spatial data allows three-dimensional models to be enriched with georeferenced information, opening scenarios for territorial analysis, environmental simulations, and enhanced enjoyment of cultural landscapes.

In addition, the approach shown differs from traditional methods of cultural heritage reconstruction based on artificial intelligence due to its semantic and multimodal nature. Instead of relying exclusively on visual data or three-dimensional surveys, the methodology uses large language models (such as ChatGPT) to analyse historical and technical texts, extracting architectural and proportional information useful for reconstruction. In this way, the process

combines linguistic analysis, visual generation, and geometric modelling, enabling the digital reconstruction of archaeological sites that integrates documentary knowledge and spatial representation, offering a more informed and interdisciplinary reconstruction.

In conclusion, the development of multimodal models capable of interpreting complex textual descriptions or historical images to directly generate structured 3D models will open new frontiers in heritage digitisation. The combined adoption of NLP, computer vision and 3D modelling will enable the production of coherent and interoperable three-dimensional representations, allowing for faster reconstruction of historical and archaeological heritage compared to traditional approaches.

Acknowledgements

The study presented in this article was funded by the Project Lip3D: Living forever the Past through a 3Digital world. Digital Europe Programme (DIGITAL), Cloud Data - AI 05 Cultheritage. Data Space for Cultural Heritage.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 for research purposes only. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] Feng, G., Zhang, B., Gu, Y., Ye, H., He, D., Wang, L. (2023). Towards revealing the mystery behind chain of thought: a theoretical perspective. *Advances in Neural Information Processing Systems*, 36, 70757-70798.
- [2] Tiribelli, S., Pansoni, S., Frontoni, E., Giovanola, B. (2024). Ethics of artificial intelligence for cultural heritage: opportunities and challenges. *IEEE Transactions on Technology and Society*.
- [3] Spennemann, D. H. (2024). Will artificial intelligence affect how cultural heritage will be managed in the future? Responses generated by Four genAI Models. *Heritage*, 7(3), 1453-1471.
- [4] Gîrbacia, F. (2024). An Analysis of Research Trends for Using Artificial Intelligence in Cultural Heritage. *Electronics*, 13(18), 3738.
- [5] Mishra, M., Lourenço, P. B. (2024). Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review. *Journal of Cultural Heritage*, 66, 536-550.
- [6] Stoean, R., Bacanin, N., Stoean, C., Ionescu, L. (2024). Bridging the past and present: AI-driven 3D restoration of degraded artefacts for museum digital display. *Journal of Cultural Heritage*, 69, 18-26.
- [7] Jaramillo, P., Sipiran, I. (2024, September). Cultural heritage 3d reconstruction with diffusion networks. In *European Conference on Computer Vision* (pp. 104-117). Cham: Springer Nature Switzerland.
- [8] Arzomand, K., Rustell, M., Kalganova, T. (2024). From ruins to reconstruction: Harnessing text-to-image AI for restoring historical architectures.
- [9] Pepe, M., Menozzi, O., Costantino, D., Alfio, V.S., Maravelakis, E., Mailis, A., Herban, S., Narloch, K., Abate, D. (2025, June). LIP3D Project: Living forever the Past through a 3Digital

World. In 2025 6th International Conference in Electronic Engineering & Information Technology (EEITE) (pp. 1-7). IEEE.

- [10] Lapenna, S. (2006). Iuvanum: L'area archeologica. Synapsi Edizioni.
- [11] Potestà, G. (2019). Il complesso teatro-tempio nell'area archeologica del santuario italico di Pietrabbondante: documentazione, analisi e ricostruzione digitale UID per il disegno: 2019, 935-944.
- [12] Shafik, W. (2024). Introduction to ChatGPT. In Advanced applications of generative AI and natural language processing models (pp. 1-25). IGI Global.
- [13] Haque, M. A., Li, S. (2024). Exploring ChatGPT and its impact on society. AI and Ethics, 1-13.
- [14] Zhou, J., Ke, P., Qiu, X., Huang, M., Zhang, J. (2023). ChatGPT: potential, prospects, and limitations. Frontiers of Information Technology & Electronic Engineering, 1-6.
- [15] Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P., Sun, L. (2024). A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT. International Journal of Machine Learning and Cybernetics, 1-65.
- [16] Tian, F., Lu, Y., Liu, F., Ma, G., Zong, N., Wang, X., Liu, C., Wei, N., Cao, K. (2024). Supervised abnormal event detection based on ChatGPT attention mechanism. Multimedia Tools and Applications, 83(41), 89501-89519.
- [17] Orrù, G., Piarulli, A., Conversano, C., Gemignani, A. (2023). Human-like problem-solving abilities in large language models using ChatGPT. Frontiers in artificial intelligence, 6, 1199350.
- [18] Fitria, T. N. (2023, March). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. In ELT Forum: Journal of English Language Teaching (Vol. 12, No. 1, pp. 44-58).
- [19] Liu, V., Vermeulen, J., Fitzmaurice, G., Matejka, J. (2023, July). 3DALL-E: Integrating text-to-image AI in 3D design workflows. In Proceedings of the 2023 ACM designing interactive systems conference (pp. 1955-1977).
- [20] Pepe, M., Alfio, V. S., Costantino, D. (2023). Assessment of 3d model for photogrammetric purposes using ai tools based on nerf algorithm. Heritage, 6(8), 5719-5731.
- [21] Heeramaglore, M., Kolbe, T. H. (2022). Semantically enriched voxels as a common representation for comparison and evaluation of 3D building models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10, 89-96.
- [22] Park, J. J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S. (2019). Deepsdf: Learning continuous signed distance functions for shape representation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 165-174).
- [23] Mitric, J., Radulovic, I., Popovic, T., Scekcic, Z., & Tinaj, S. (2024, February). AI and computer vision in cultural heritage preservation. In 2024 28th International Conference on Information Technology (IT) (pp. 1-4).

- [24] Foka, A., & Griffin, G. (2024). AI, cultural heritage, and bias: Some key queries that rise from the use of genAI. *Heritage* 7 (11), 6125–6136.