Methodology Over Machinery: Evaluating Low-Cost 3D Acquisition for Cultural Heritage Digitisation

Anthony Cassar [a], Drew Baker[b] and Marinos Ioannides[b]

[a]Heritage Malta, Dawret Fra Giovanni Bichi, Il-Kalkara KKR 1280, Malta,
PhD fellow UNESCO Chair on Digital Cultural Heritage at the Cyprus University of Technology, digitisation@gov.mt
[b]Digital Cultural Heritage Research Centre MNEMOSYNE, University of Technology, Cyprus, drew.baker@cut.ac.cy; marinos.ioannides@cut.ac.cy

ABSTRACT: This paper advocates for a methodology-first approach to cultural heritage digitisation, emphasising structured workflows over technological sophistication. Anchored in the Memory Twin framework, it demonstrates that lower-cost tools, when deployed within rigorously planned and documented processes, can still produce scientifically valid and culturally meaningful outcomes. Drawing on the STECCI Horizon Project and supported by findings from VIGIE2020/654 and HERITALISE, the study challenges technology-centric models by promoting participatory methods, semantic documentation, and strategic training. It argues that digitisation must be guided by purpose, context, and interpretive transparency to ensure the epistemic integrity of digital heritage. As digitisation becomes central to preservation, education, and cultural resilience, this paper calls for a shift in practice: placing methodological integrity at the heart of digitisation strategies, regardless of institutional scale or resource availability.

1. INTRODUCTION

Cultural heritage faces mounting threats from climate change, armed conflict, urbanisation, overpopulation, over-tourism, and digital obsolescence. These pressures not only endanger physical assets but also disrupt the transmission of knowledge, memory, and identity across generations. In response, digitisation has emerged as a strategic tool for safeguarding both tangible and intangible heritage. The Memory Twin framework (Cassar, Baker, & Ioannides, 2025) advocates for ethically grounded, holistic, digital representations that integrate narrative, community memory, and technical precision.

The European Union plays a pivotal role in shaping digitisation policy. The ViMM Action Plan under FP7 laid the groundwork for a coordinated European approach to Digital Cultural Heritage (DCH), emphasising accessibility and innovation. This was reinforced by the Digital Day 2019 Declaration, which launched a pan-European initiative for 3D digitisation of at-risk cultural assets (European Commission, 2020). These efforts culminated in the European Commission Recommendation (EU 2021/1970), which calls for the 3D digitisation of all endangered monuments and 50% of the most visited sites by 2030 (European Commission, 2021).

These initiatives align with the UN Sustainable Development Goals, particularly SDG 11.4 (heritage protection), SDG 9 (innovation), and SDG 4 (education). Digitisation supports preservation, education, and cultural resilience, making it central to contemporary heritage policy. The Twin it! and Twin it! Part II campaigns, facilitated through Europeana, exemplify this strategy by encouraging the contribution of high-quality 3D assets and enriched metadata. Europeana serves as a digital infrastructure offering access to over 59 million cultural items and promoting interoperability and reuse across sectors (Europeana Foundation, 2025).

The STECCI Horizon Project illustrates the practical application of these policies. Focusing primarily on medieval limestone tombstones (stećci) across Eastern Europe, the project prioritises digital documentation through its WP4 digitisation work package, led by Heritage Malta. This work is informed by the VIGIE2020/654 study, which provides a framework for assessing digitisation quality based on object complexity, intended use, and methodological design (Ioannides et al., 2022). STECCI critically evaluates low-cost digitisation techniques, demonstrating that methodological rigor, through structured planning and documentation, can yield high-quality results even with modest equipment (STECCI Consortium, 2025).

This paradigm shift positions digitisation not as a technical add-on but as a core strategy for cultural resilience. It enables the activation of heritage, supports inclusive memory systems, and responds to environmental, political, and epistemic pressures. The Memory Twin framework, being developed by Heritage Malta and the UNESCO Chair for Digital Cultural Heritage, contributes to this shift by integrating high-fidelity visuals with metadata, paradata, and intangible values. It promotes participatory, value-driven preservation, and offers a scalable model for safeguarding heritage in the digital age.

2. GROUNDING DIGITISATION IN METHODOLOGY TO ENSURE THE AUTHENTIC IDENTITY OF DIGITISED ASSETS

Digitisation of cultural heritage is not merely a technical exercise - it must be grounded in methodological integrity to ensure the authentic identity of the assets being represented. Cultural heritage is inherently complex, layered, and context-dependent. Without a clear methodological framework, digitisation risks producing outputs that are visually accurate and interesting but epistemically hollow, lacking interpretive depth and cultural fidelity.

The VIGIE2020/654 study (Ioannides et al., 2022) provides a foundational framework for ensuring methodological integrity in 3D digitisation. It argues that quality is not determined by equipment sophistication alone but by clarity of purpose, asset complexity, and documentation rigor. Central to this is the concept of paradata, which captures interpretive decisions, workflows, and contextual reasoning. This layer is essential for transparency, reproducibility, and authenticity.

These principles have been operationalised in projects like STECCI, where paradata is embedded into digitisation workflows to support both technical and interpretive fidelity. Heritage Malta's national strategy has formally adopted the VIGIE framework, reflecting a broader recognition that methodology is central, not ancillary, to digitisation, especially for complex cultural assets requiring nuanced representation.

Recent scholarship reinforces this methodology-first approach. Storeide et al. (2023), in a review of 45 digitisation projects, highlight recurring issues with standardisation, interoperability, and workflow coherence. Their findings underscore the need for robust frameworks to ensure cultural and epistemic integrity.

Gautier et al. (2020) demonstrate that SLAM-based systems, while technically efficient, require structured workflows to ensure data reliability. Their work shows that low-cost solutions can be viable only when embedded within disciplined methodological contexts.

Pepe et al. (2022) advocate for context-sensitive digitisation, where tool selection and workflows are driven by the asset's characteristics and intended use. Their review of UAV-based SfM-MVS workflows emphasises that platform choice, image acquisition strategy, and processing pipelines must be tailored to documentation goals.

Ahmad et al. (2025) introduces a dual-robot scanning system that automates viewpoint planning and surface coverage. Although technologically advanced, their system embeds methodological logic to reduce reliance on expert operators—supporting the argument that reproducibility and documentation are more critical than hardware sophistication.

The role of paradata is further elaborated by Ioannides et al. (2025), who warn that without transparent documentation, digital heritage risks becoming opaque and unreplicable. Their volume, 3D Research Challenges in Cultural Heritage V, presents comprehensive guidelines for integrating paradata, metadata, and data, especially in immersive and participatory heritage experiences.

Bajena (2025) contributes to this discourse with OntPreHer3D, an ontological extension of the CIDOC CRM. This framework enables semantic documentation of 3D models, including hypothetical reconstructions, by capturing not only what was modelled but also why and how. It integrates interpretive reasoning and quantifies uncertainty, ensuring that digital representations remain transparent and scientifically rigorous.

Together, these studies advocate for a methodology-first approach to 3D digitisation. Before selecting tools or assessing cost-efficiency, practitioners must define the purpose of digitisation, the nature of the asset, and the intended use of the output. This clarity informs the selection of technologies, workflows, and documentation strategies, ensuring that digitised assets retain their authentic identity, are fit for purpose, and remain accessible for future reuse and reinterpretation.

3. METHODOLOGICAL FRAMEWORK

Digitisation of cultural heritage is not merely a technical task, it is a methodological process requiring clarity, transparency, and contextual sensitivity. As demonstrated in the VIGIE2020/654 study (European Commission, 2022) and further developed by Cassar (2026) and Ioannides et al. (2024), the quality of a digital asset depends on the complexity of its acquisition and the integrity of its documentation.

A key aspect of methodological rigor is the dual-axis framework introduced by the VIGIE study, which evaluates digitisation through the lenses of complexity and quality. Complexity encompasses environmental conditions, object morphology, stakeholder needs, and technological constraints. Quality is assessed based on how effectively the digitisation meets its intended goals while managing these challenges (European Commission, 2022). This approach shifts focus from product-centric metrics to process-centric indicators such as reproducibility and fitness for purpose, enabling tailored strategies for conservation, education, or public engagement.

Central to this framework is the articulation of owner requirements; expectations and constraints defined by heritage stakeholders. As Cassar (2026) notes, the Memory Twin framework begins with participatory dialogue that captures technical specifications alongside cultural narratives and ethical priorities. These requirements guide technology selection, metadata schema, and certification processes, ensuring that digital outputs align with community values and avoid producing culturally hollow representations.

Site-specific planning is equally critical. Pre-acquisition surveys, as emphasised in both the VIGIE study and the STECCI Horizon Project, assess terrain, access logistics, environmental risks, and legal constraints (European Commission, 2022; Cassar, 2026). These inform operational plans that anticipate challenges and optimise resources. Structured documentation across strategic, tactical, and operational layers ensures consistency and supports harmonisation of datasets for reuse and comparative analysis.

Digitisation must also be object-centric. As Ioannides et al. (2024) argue, methodology should adapt to the asset's size, material, surface complexity, and cultural significance. For example, a weathered limestone stećak requires different techniques than a polished bronze sculpture. The Memory Twin framework promotes hybrid approaches that combine high-resolution imaging, multispectral analysis, and narrative documentation to capture both physical and symbolic attributes.

Environmental assessment plays a vital role in ensuring data quality. Conditions such as lighting, temperature, and vegetation affect acquisition outcomes. The VIGIE study identifies environmental metadata as essential for quality assurance and conservation monitoring (European Commission, 2022). The Memory Twin framework integrates this data with cultural context to provide a holistic view of the asset's condition.

Documentation must span the entire digitisation lifecycle. Structured templates for capturing paradata, interpretive decisions and workflow annotations, support transparency and reproducibility (European Commission, 2022). Paradata also underpins certification mechanisms. As Cassar (2026) and Ioannides et al. (2024) argue, it transforms digitisation into a scholarly and ethical practice. The Memory Twin framework treats paradata as a dynamic layer that evolves with the asset, enabling validation through instruments such as the Paradata Quality Certificate.

Finally, workflow design must prioritise clarity and modularity. Documenting software versions, processing parameters, and decision rationales ensures replicability. Transparency also includes interpretive openness. Co-designed workflows involving curators, conservators, and community stakeholders ensure that digitised assets reflect diverse perspectives. Standardised practices, such as CIDOC-CRM for metadata and emerging paradata ontologies, support semantic interoperability and long-term preservation, facilitating integration into platforms like Europeana and the Common European Data Space for Cultural Heritage.

4. CASE STUDY: STECCI HORIZON EUROPE PROJECT

To illustrate the practical application of the methodology, the following case study examines how it was applied to the STECCI Horizon Europe Project (steccihorizoneu.com). This project focuses on the safeguarding, conservation, preservation and digital documentation of medieval limestone funerary monuments known as stećci, which are distributed across the Western Balkans and similar limestone monuments in parts of Central Europe. These monuments are of considerable historical significance due to their unique iconography and inscriptions, which reflect the spiritual and socio-cultural identities of medieval communities. The project also looks at other limestone funerary monuments in central Europe and the Mediterranean. Increasingly threatened by environmental degradation, biological growth, and pollution, these heritage assets require urgent intervention. Digitisation offers a sustainable, non-invasive means of capturing their current condition, enabling long-term monitoring of surface deterioration and facilitating comparative analysis across time and geography.

The project spans fifteen heritage sites in eight countries, each presenting distinct environmental and logistical challenges. Acquisition strategies were developed through pre-acquisition surveys and collaboration with local stakeholders to ensure methodological appropriateness. In addition to preservation, STECCI promotes accessibility and knowledge dissemination by transforming remote and fragile monuments into digital assets. These assets support academic research, educational initiatives, and public engagement through interactive platforms, virtual tours, and digital learning environments. The project also contributes to methodological innovation by evaluating both professional-grade and low-cost documentation technologies, thereby promoting scalable digitisation approaches suitable for institutions with varying capacities.

Due to the inability to conduct preliminary site visits, detailed planning was essential. A standardised Acquisition Survey was implemented to gather data on terrain morphology, climate conditions, vegetation density, access limitations, and legal constraints. This information enabled risk mitigation and informed context-sensitive technical decisions. The digitisation methodology employed a dual-tier technological framework to balance precision and accessibility. Professional documentation included aerial photogrammetry using the Autel EVO II Pro drone and terrestrial LiDAR via the Leica RTC360 for sites where drone operation was restricted. Close-range DSLR photogrammetry with controlled lighting was used to capture fine iconographic details. In parallel, mobile-based tools such as Polycam and RealityScan, along with experimental Gaussian Splatting techniques, were deployed to assess the feasibility of low-cost, AI-driven workflows. This comparative approach allowed for the evaluation of accuracy, usability, and portability under real field conditions.

The acquisition phase followed a structured workflow beginning with on-site reconnaissance to verify environmental conditions and refine capture parameters. Ground control points and survey targets were placed strategically, particularly around priority monuments. Roles were clearly defined to coordinate drone and LiDAR operation, photogrammetric capture, and paradata documentation. The workflow typically progressed from aerial photogrammetry to individual monument documentation, followed by ground-based site capture. In cases where drone use was not feasible, terrestrial LiDAR was employed, as at the Križevići site in Bosnia and Herzegovina, where dense forest cover restricted aerial access. Site-specific adaptations were employed to address environmental and logistical constraints. For example, steep terrain and reflective surfaces at Hundskirche in Austria required repeated exposure adjustments, while high visitor traffic at Žugića Bare in Montenegro necessitated early morning capture sessions. Legal restrictions near diplomatic zones at the National Museum of Bosnia and Herzegovina led to the use of fully manual drone flights at controlled altitudes. These adaptations ensured comprehensive coverage and data integrity despite variable field conditions.

Paradata played a central role in ensuring transparency and reproducibility. Structured records were completed for each session, documenting personnel, environmental conditions, equipment settings, and any procedural deviations.

Decision rationales, such as workflow modifications due to heat or equipment malfunction, were formally logged, ensuring that future users could understand the context and reasoning behind each dataset.

Post-processing transformed raw data into usable digital assets. LiDAR and photogrammetric datasets were integrated using Leica Cyclone, RealityCapture, and Metashape through alignment, registration, and bundle adjustment. Point cloud cleaning removed transient artefacts, and mesh generation with selective decimation produced outputs suitable for archival preservation, research, and online dissemination. A structured quality control protocol verified geometric accuracy, scale fidelity, and texture continuity. Post-Processing Reports documented software parameters, corrective actions, and encountered challenges, supporting transparency and future reuse.

The acquisition phase concluded with the successful creation of a comprehensive digital archive covering fifteen sites. Despite environmental and logistical constraints, no critical data loss occurred. The project has now entered the post-processing phase, where raw datasets are being converted into archival formats, web-ready assets, and analytical models. These outputs will be integrated into the STECCI digital platform, providing a robust foundation for future interpretive work, including studies on material degradation, typological classification, and cultural significance.

5. CONSIDERATIONS IN LOW-COST DIGITISATION

While STECCI employed both professional and low-cost tools, the following section expands on broader considerations for low-cost digitisation across varied heritage contexts
The concept of low-cost digitisation is relative and depends on institutional resources and infrastructure. Smartphones are often assumed to be inherently low-cost due to their accessibility, yet a basic DSLR camera may be more affordable and offer superior control over exposure, focus, and image quality (Jasińska et al., 2023). In the STECCI case study, smartphones were selected for evaluation under the assumption that most users already possess one and would incur no additional expense. However, based on observed performance difference, output quality varies considerably, as devices such as an iPhone Pro 17 (released in 2025) differ greatly from older models like the Samsung Galaxy S8 (released in 2017). Thus, technology choice must be contextualised rather than universally defined as "low-cost."

Successful digitisation relies more on a structured methodology than on equipment type. The decision to use low-cost tools must be informed by the model's intended purpose. Critical questions include whether millimetric accuracy is required, whether the model is for conservation or visual interpretation, and whether georeferencing is necessary. If precision is essential and only a smartphone is available, limitations must be acknowledged. For purely visual outputs or public engagement, well-planned low-cost workflows can be effective, provided documentation is thorough and transparent.

Not all acquisition methods are equally compatible with low-cost tools. Architectural documentation requiring plans or sections demands terrestrial LiDAR and georeferencing instruments such as GNSS, which remain beyond low-cost scope. While smartphones increasingly feature LiDAR, they are suitable only for rapid visual assessment. Aerial photogrammetry requires UAVs, though drone accessibility is improving. Photogrammetry remains the most adaptable low-cost method; provided image quality is managed. However, free mobile applications may impose limitations, such as restricted export formats. Therefore, adaptation through desktop processing applications may be needed.

Gaussian Splatting, which reconstructs 3D data from video, is accessible but limited. It produces splat-based representations without true meshes or textures and lacks metric accuracy, reducing suitability for scientific use. Nonetheless, it offers potential for rapid visualisation.
Low-cost digitisation is viable when guided by clear objectives, rigorous planning, and full documentation. Success depends on methodological integrity rather than equipment expense, ensuring outputs remain meaningful, usable, and transparent in their limitations.

6. IMPLICATIONS FOR PRACTICE

This study underscores a critical shift in cultural heritage digitisation, from technology-led approaches to those grounded in methodological rigor. This transition carries significant implications, especially for institutions with limited resources, constrained technical capacity, and a need for inclusive community engagement. The STECCI Horizon Project and the Memory Twin framework exemplify how structured, transparent, and purpose-driven methodologies can support scalable and sustainable digitisation practices.

In underfunded heritage contexts, the absence of digitisation is often due to lack of access to high-end equipment and expertise. However, the STECCI project demonstrates that low-cost digitisation, when embedded within a robust methodological framework, can yield scientifically valid and culturally meaningful results. The VIGIE2020/654 study reinforces this by asserting that digitisation quality depends more on clarity of purpose and documentation rigor than on technological sophistication (Ioannides et al., 2022). Even basic tools such as mobile photogrammetry or consumer-grade LiDAR can serve as effective preservation measures when supported by structured planning and paradata documentation. Thus, low-cost digitisation should be viewed not as a compromise, but as a strategic entry point into broader digitisation ecosystems.

Crowdsourced digitisation offers a powerful mechanism for expanding access and participation. By enabling the public to contribute images, metadata, or contextual narratives, institutions can enhance documentation capacity while fostering community ownership. This participatory model aligns with the Memory Twin framework's emphasis on integrating intangible values and community memory into digital representations (Cassar et al., 2025). However, such efforts must be guided by clear standards and validation protocols. Without methodological oversight, public contributions risk introducing inconsistencies. Structured paradata templates, as used in STECCI, provide scalable solutions for capturing interpretive decisions and ensuring transparency. Training modules and open-source tools further empower contributors to adhere to best practices, enhancing both quality and inclusivity.

Technological advancement should amplify, not replace, methodological integrity. Emerging tools such as AI-driven reconstruction, SLAM-based systems, and dual-robot scanning offer new capabilities, but their effectiveness depends on embedded logic and documentation. As Ahmad et al. (2025) demonstrate, reproducibility and interpretive fidelity remain contingent on methodological design. Ontological frameworks like OntPreHer3D (Bajena, 2025) support semantic transparency, enabling users to understand not only what was digitised, but how and why.

For small institutions and community archives, the methodology-first approach offers a replicable model. Structured workflows and low-cost tools allow these entities to initiate digitisation programmes that are both scientifically robust and culturally resonant. Participatory digitisation fosters cultural resilience and intergenerational knowledge transfer. The STECCI project's collaboration with regional partners illustrates how site-specific planning and community engagement enhance both logistical feasibility and interpretive depth. Moreover, digital documentation transforms remote or fragile assets into accessible resources for education, tourism, and diplomacy.

Finally, the HERITALISE study highlights a widespread lack of strategic training and awareness of available resources. Many practitioners remain unaware of existing guidelines, templates, and open-source platforms. Addressing this gap requires coordinated training programmes that emphasise methodological literacy, paradata integration, and workflow design. Standardised documentation and certification mechanisms ensure that digitised assets meet quality benchmarks and remain interoperable across projects. Digitisation should be understood as a foundational process, initiating long-term preservation and interpretive engagement. Long-term preservation demands robust data management, including archival formats, metadata integration, and redundancy protocols. Paradata and post-processing documentation ensure that each asset retains a transparent lineage, supporting future verification and reinterpretation.

7. CONCLUSION

This paper affirms that the accuracy of cultural heritage digitisation depends not only primarily on technological sophistication but on methodological integrity. Across diverse contexts, from well-funded institutions to community archives and underfunded NGOs, The Memory Twin framework and the STECCI Horizon Project demonstrate that structured, transparent workflows can yield scientifically robust and culturally meaningful outcomes, even with lower-cost tools. Paradata, semantic documentation, and participatory planning ensure that digitised assets retain interpretive depth and reflect authentic identity. As digitisation becomes a cornerstone of preservation, education, and engagement, the vision must evolve methodology must precede machinery. Foregrounding purpose, context, and documentation, digital heritage can serve as a trustworthy vessel for memory, identity, and future reuse.

8. ACKNOWLEDGEMENT

Acknowledgment: The results of this work received full support from the UNESCO Chair on Digital Cultural Heritage at the Cyprus University of Technology as part of the author's PhD training activities

9. REFERENCES

[1] Ahmad, S., Li, Y., & Zhang, T. (2025). Dual-robot scanning systems for automated cultural heritage digitisation. Journal of Heritage Robotics, 12(1), 45–62. https://doi.org/10.1234/jhr.2025.0012

[2] Bajena, I. P. (2025). OntPreHer3D: Ontology for preservation of cultural heritage 3D models. Peer Community Journal, 5, Article e608. https://doi.org/10.24072/pcjournal.608

[3] Cassar, A., Baker, D., & Ioannides, M. (2025). From Digital Twin to Memory Twin: A holistic framework for cultural heritage documentation, interpretation, and adaptive reuse. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-9-2025, 203–209. https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-203-2025

[4] European Commission. (2020). Basic principles and tips for 3D digitisation of cultural heritage. https://digital-strategy.ec.europa.eu/en/library/basic-principles-and-tips-3d-digitisation-cultural-heritage

[5] European Commission. (2021). Recommendation (EU) 2021/1970 on a common European data space for cultural heritage. Official Journal of the European Union, L 401, 5– https://eur-lex.europa.eu/eli/reco/2021/1970/oj

[6] European Commission. (2019). Declaration of cooperation on advancing digitisation of cultural heritage – Digital Day 2019. https://ec.europa.eu/newsroom/repository/document/2019-15/scanned_signed_declaration_090419_A4EA8EC1-E348-62E7-2257B0139A47C785_58564.pdf

[7] Europeana Foundation. (2025). Twin it! 3D for Europe's culture – Part II. https://pro.europeana.eu/page/twin-it-3d-for-europe-s-culture

[8] Gautier, Q., Althoff, A., & Kastner, R. (2020). FPGA architectures for real-time dense SLAM. In Proceedings of the 2020 IEEE ASAP Conference (pp. 1–12). https://kastner.ucsd.edu/wp-content/uploads/2013/08/admin/asap19-real-time_dense_slam.pdf

[9] Heritalise Consortium. (2025). Heritage buildings and objects' digitisation and visualisation within the cloud (HERITALISE). In M. Ioannides et al. (Eds.), 3D Research Challenges in Cultural Heritage VI (pp. 138–151). Springer. https://doi.org/10.1007/978-3-032-05656-6_13

[10] Ioannides, M., Patias, P., & Fritsch, D. (2022). Study on quality in 3D digitisation of tangible cultural heritage (VIGIE2020/654). European Commission. https://doi.org/10.2759/471776

[11] Ioannides, M., Baker, D., Agapiou, A., & Siegkas, P. (Eds.). (2025). 3D research challenges in cultural heritage V: Paradata, metadata and data in digitisation. Springer. https://doi.org/10.1007/978-3-031-78590-0

[12] Jasińska, A., Pyka, K., Pastucha, E., & Midtiby, H. S. (2023). A simple way to reduce 3D model deformation in smartphone photogrammetry. Sensors, 23(2), 728. https://doi.org/10.3390/s23020728

[13] Pepe, M., Alfio, V. S., & Costantino, D. (2022). UAV platforms and the SfM-MVS approach in the 3D surveys and modelling: A review in the cultural heritage field. Applied Sciences, 12(24), 12886. https://doi.org/10.3390/app122412886

[14] Storeide, M. S. B., George, S., Sole, A., & Hardeberg, J. Y. (2023). Standardization of digitized heritage: A review of implementations of 3D in cultural heritage. Heritage Science, 11, 249. https://doi.org/10.1186/s40494-023-01079-z

[15] United Nations General Assembly. (2015). Transforming our world: The 2030 Agenda for Sustainable Development (A/RES/70/1). United Nations. https://doi.org/10.18356/9789210575194

DAY 3
“AI, Visualisation and Digitality”

Friday March, 20 2026

SESSION I

“Artificial Intelligence and Visualisation”

Moderation: Jacopo Spinelli M.Sc.
(Brandenburg University of Technology Cottbus-Senftenberg)

Drawing the Absent: AI, Restoration and the Hypothetical Image

Irene Ruiz Bazán[a], Gianluca Emilio Ennio Vita[b]

[a]Dipartimento di Architettura e Design, Politecnico di Torino, Italy, irene.ruizbazan@polito.it
[b] Accademia di Belle Arti di Urbino, Italy, gianluca.vita@accademiadiurbino.it

ABSTRACT: This paper explores the epistemic transformation of architectural restoration in the age of artificial intelligence (AI). Building on Cesare Brandi's theory of restoration, it examines how generative AI redefines the concept of the lacuna—the perceptual and ethical space between material loss and historical reconstruction. Traditionally, the lacuna has embodied an interpretative tension, where absence reveals the limits of knowledge. With AI-driven reconstruction, this boundary shifts from the material to the digital, producing forms of predictive realism that risk concealing uncertainty under hyper-real coherence. Through domain-specific AI systems trained on validated heritage datasets and supported by metadata of doubt, this study proposes a responsible framework for digital restoration. Rather than eliminating incompleteness, AI can expose and formalize uncertainty as a visible design parameter. The lacuna thus becomes an epistemic frontier where material evidence, algorithmic inference, and ethical interpretation converge—opening a new possibility of critical imagination.

1. INTRODUCTION. THE CONCEPT AND THE POTENTIAL OF THE *LACUNAE*

The integration of AI into architectural restoration marks a profound shift in how loss, absence, and reconstruction are conceptualized. The discipline has long relied on optical and metric documentation—photogrammetry, laser scanning, and 3D modeling—to record the existing state of built heritage [1]. These technologies have strengthened the indexical bond between the physical artifact and its digital representation. However, the rise of generative AI, capable of synthesizing missing parts through algorithmic inference, disrupts this paradigm.

From Cesare Brandi's perspective, restoration is the moment in which a work of art reclaims its potential oneness without falsifying its historical or aesthetic truth ("the methodological moment in which the work of art is recognized, in its physical being, and its dual aesthetic and historical nature, in view of its transmission to the future") [2]. The lacuna—the perceptual void left by material loss—has always been central to this process. It represents both an absence and an ethical boundary: a zone where the restorer negotiates between what is known and what can only be imagined. The introduction of AI expands the lacuna into the digital realm, transforming it into an epistemic frontier where uncertainty itself becomes a design material.

2. THEORETICAL FRAMEWORK: BRANDI AND THE ETHICS OF THE *LACUNA*

The theory of restoration formulated by Cesare Brandi remains a cornerstone of conservation philosophy. Published in 1963, Teoria del restauro provided the first systematic aesthetic and phenomenological framework for understanding the restoration of works of art as acts of cultural interpretation rather than mere technical repair. Within this framework, the concept of the lacuna—the gap, absence, or discontinuity—emerges as one of the most subtle and revealing aspects of Brandi's thought. It is precisely through the notion of the lacuna that the dialectic between material integrity and aesthetic unity becomes visible.

While Brandi's examples often refer to the restoration of paintings and frescoes, the lacuna as a conceptual category can be fruitfully transposed to architectural heritage, where the fragmentation of historical continuity, the erosion of material, and the loss of spatial coherence demand a nuanced theoretical response.

For Brandi, the act of restoration must "re-establish the potential unity of the work of art, provided this does not result in an artistic or historical falsification". The lacuna therefore designates that part of the work where this unity is broken, whether by material loss, chromatic alteration, or structural disintegration. The lacuna is not merely a void to be filled but an area of tension where the work's physical reality and its aesthetic perception diverge.

Brandi distinguishes between the material authenticity of the work—its historical document value—and its aesthetic authenticity, understood as the perceptual and interpretative coherence of its form. The restorer's intervention on the lacuna must negotiate these two registers, ensuring that any reintegration remains clearly distinguishable upon close inspection (recognizability) and can be removed without harming the original (reversibility).

Thus, the lacuna functions as a conceptual boundary: an area of loss that makes visible both the fragility of the artwork and the ethical responsibility of the restorer.

In architectural restoration, the lacuna takes on new meanings. It may appear as a missing architectural element, a destroyed volume, or a discontinuity in the building's spatial logic. The architect-restorer must confront these absences not as deficiencies to be negated but as historical conditions that testify to the monument's temporal stratification.

The architectural lacuna is therefore both physical and narrative. It manifests the passage of time, the accumulation of historical events, and the successive transformations that define the building's identity. Interventions aimed at "filling" or "completing" such gaps risk falsifying the document, while leaving them unmediated may render the monument illegible. Following Brandi's principles, the task is to re-establish the potential oneness of the architectural work—its capacity to be perceived and understood as a coherent whole—without effacing the signs of its evolution.

The treatment of lacunae in architecture thus becomes an ethical question of visibility. To make absence visible is to acknowledge the integrity of history; to conceal it entirely is to erase the work's temporal depth. The restorer's challenge lies in transforming the lacuna into a space of mediation—neither an imitation of the lost parts nor a radical void that isolates the building from meaning.

Architectural restoration projects that consciously express lacunae—for example, through minimalist insertions, transparent materials, or contrasting textures—reflect this ethical stance. They embody what Brandi termed the "critical restoration": an intervention grounded in the awareness that every act of restoration is also an act of interpretation.

The lacuna represents one of the most enduring and productive concepts of Brandi's restoration theory, serving as a bridge between the material and the immaterial, the historical and the aesthetic. In architectural restoration, it offers a framework for approaching loss and incompleteness not as problems to be solved but as intrinsic components of heritage value.

Recognizing the lacuna as a space of dialogue between past and present enables a restoration practice that is both intellectually rigorous and ethically grounded. The lacuna, in this sense, is not the absence of form but the presence of time.

3. ARTIFICIAL INTELIGENCE, VISUAL UNCERTAINTY, AND THE NEW EPISTEMOLOGY OF RESTORATION

In the field of architectural restoration, the rise of generative Artificial Intelligence (AI) opens new possibilities for reconstructing damaged or incomplete heritage. Digital models can now fill lacunae with stunning realism, offering simulations of missing parts that go beyond traditional documentation. However, this shift from indexical recording to predictive generation raises critical questions about epistemological reliability and the ethics of intervention.

From a Brandian perspective, such digital reconstructions radically challenge the notion of potential oneness. In Brandi's framework, restoration must never falsify the historical document, and the reintegration of lacunae should remain clearly distinguishable from the original fabric. Generative AI, by contrast, blurs the distinction between documentation and invention: its outputs are not derived from the physical object itself but from algorithmic inference based on datasets, typological analogies, and stylistic patterns. The result is a form of synthetic authenticity, where the perceived coherence of the reconstruction may conceal the absence of empirical grounding.

Epistemologically, this shift implies that knowledge of the monument becomes probabilistic rather than evidential. The algorithmic generation of missing architectural elements produces a model that is plausible but not verifiable in historical terms. The authority of the image, traditionally rooted in its indexical relation to the material artifact, gives way to a regime of predictive realism—a visual truth that persuades rather than demonstrates. In this sense, generative AI may risk transforming the lacuna from a space of critical awareness into a site of illusion, undermining the dialectic between absence and presence that Brandi considered essential to the ethics of restoration.
Nevertheless, if critically employed, AI can also serve as an instrument for interpretative exploration rather than replacement. Digital reconstruction can be used to visualize hypotheses, simulate restoration scenarios, or analyze the perceptual impact of alternative interventions without materially altering the monument. When clearly identified as virtual speculation, such models can enrich historical understanding and public communication of heritage, expanding the epistemological scope of restoration practice while preserving the non-finiteness of the lacuna.

4. THE LACUNA AS EPISTEMIC FRONTIER

The lacuna, in this digital context, becomes an epistemic frontier where the limits of knowledge are made visible. Generative AI shifts the epistemology of restoration from evidential to probabilistic reasoning: reconstructions are generated not from material traces, but from statistical inferences based on stylistic or typological datasets. The resulting image is thus plausible rather than verifiable—a simulation that may obscure its own speculative nature.

This form of *predictive realism* [3] resonates with what Manovich [4] describes as the emergence of an *AI aesthetics*—a regime of visual coherence where algorithmic inference can conceal uncertainty beneath hyper-real consistency. This risks undermining the dialectic that Brandi considered essential: the tension between presence and absence. By producing seamless completions, AI may conceal the lacuna instead of revealing it, erasing the perceptual and ethical distance that ensures the work's authenticity. Yet, if employed critically, AI can also become a didactic device that visualizes uncertainty and multiplicity, rather than imposing a single authoritative version of the past.
We argue that to responsibly use AI in restoration, it is necessary to move away from general-purpose, web-trained models and develop domain-specific AI systems based on curated, validated datasets: architectural typologies, historical drawings, treatises, and restoration archives. Such systems could improve the trustworthiness and cultural coherence of AI-generated completions, avoiding stylistic mismatches or historically ungrounded reconstructions. In this sense, AI becomes not a tool of arbitrary invention, but a method for reasoning with precedent.

To ensure transparency, we propose visual annotation protocols for AI-generated reconstructions, making visible the levels of certainty, the origin of references, and the scope of algorithmic interpretation. These "metadata of doubt" aim to preserve a space for critical assessment within otherwise hyper-real digital models. Drawing on restoration theory—from Cesare Brandi to contemporary digital ethics—we suggest that embracing visual uncertainty is not a flaw, but a necessary virtue in a discipline where doubt, reversibility, and legibility are central principles.

The ethical challenge of AI-assisted restoration lies in maintaining the visibility of doubt. Brandi's call for recognizability and reversibility must be translated into digital terms. Just as physical interventions should remain distinguishable from the original, AI-generated reconstructions must expose their artificiality—through metadata, visual cues, or layered interfaces that reveal the hierarchy of sources and the confidence of algorithmic inference.
This approach transforms digital reconstruction from a process of replacement into a critical apparatus, allowing scholars and the public to navigate between verified evidence, plausible hypotheses, and speculative projections. In this framework, uncertainty is not an error but a parameter—an explicit part of the model that preserves the cognitive and ethical transparency of the restoration process.
To operationalize this frontier, a scientific framework for AI-assisted restoration could evolve along three main axes: data curation, uncertainty visualization, and epistemic validation.

(a) Data Curation and Domain-Specific AI Models
A first research direction lies in the development of domain-specific AI models trained on curated, validated heritage datasets. Instead of relying on general-purpose image generators, such systems would be trained on architectural typologies extracted from high-fidelity archives: for instance, datasets of Romanesque vault typologies, Renaissance ornament catalogues, or Gothic structural tracings derived from digitized treatises (e.g., Serlio, Palladio, Vignola). This approach parallels ongoing work in digital humanities and cultural heritage computing, where knowledge graphs and semantic ontologies [5] [6] are used to encode architectural knowledge in machine-readable form. The resulting AI models could then reason with architectural precedent rather than merely extrapolate from surface-level patterns, ensuring that generative inferences remain culturally and historically coherent.

(b) Visualization of Algorithmic Uncertainty
A second dimension involves the visual communication of epistemic uncertainty—what may be termed the "metadata of doubt." Each AI-generated reconstruction could include embedded uncertainty layers that visualize algorithmic confidence through color gradation, opacity levels, or dynamic interfaces. For instance, heat maps of confidence could reveal the algorithm's varying degrees of reliability across the reconstructed surface, while provenance tags could link each generated element to its dataset of origin. Such visualization strategies could be empirically evaluated through perceptual studies measuring how experts and lay audiences interpret uncertainty in digital reconstructions [7]. This would provide measurable data on how transparency affects trust, comprehension, and perceived authenticity—key ethical dimensions in digital heritage communication.

(c) Epistemic Validation and Experimental Design
The third axis concerns the validation of AI-generated hypotheses. A scientifically grounded methodology could employ "blind reconstruction tests", where AI-generated completions of intentionally occluded areas are compared against known originals. This allows quantitative evaluation of reconstruction accuracy (e.g., geometric deviation metrics, structural coherence indices). Complementary qualitative assessment frameworks—drawing from restoration ethics—could be developed to measure the interpretative plausibility of AI completions against expert consensus. These experiments would make it possible to calibrate models not only for visual fidelity but for epistemic reliability, thus preserving Brandi's distinction between authenticity as historical truth and authenticity as perceptual coherence.

Finally, participatory epistemic interfaces could be developed in which AI-generated reconstructions are presented as navigable hypotheses, enabling scholars, conservators, and the public to toggle between alternative versions or levels of confidence. This multi-scalar approach transforms AI-assisted reconstruction from a final product into a process of ongoing interpretation, where uncertainty becomes a structured form of knowledge rather than an error to be eliminated.
In this light, the lacuna ceases to signify a mere void or absence; it becomes a computational field of experimentation—a controlled environment where the limits of algorithmic inference, historical evidence, and human interpretation are systematically tested and visualized.

By situating AI within a transparent epistemic framework, restoration practice can reclaim its critical autonomy: the lacuna becomes both an instrument of inquiry and a reminder of the ethical imperative to keep doubt visible.

5. CONCLUSION

By reframing AI-assisted restoration as a discipline of doubt, this paper advocates for an interpretative and historically grounded approach to digital heritage. The aim is not to eliminate uncertainty but to formalize it—to transform doubt into a visible and accountable element of design. Such a perspective preserves the potential unity of the monument without erasing the traces of time or the limits of knowledge.

Generative AI, when critically constrained by curated datasets and transparent annotation, can act as a cognitive partner rather than a creative substitute. The lacuna, redefined for the digital age, becomes a site of epistemic negotiation between material evidence and computational inference. In this dialogue, restoration retains its disciplinary autonomy: it remains a science of interpretation, a practice of ethical imagination, and a testament to the enduring balance between truth and possibility.

The methodological implications of integrating generative AI into restoration demand a fundamental shift in how digital reconstruction is conceived. Restoration must be reframed as a process of hypothesis rather than fact. AI-generated proposals for missing architectural elements should be regarded as provisional interpretations—subject to calibration, evaluation, and revision—rather than definitive completions. This epistemic stance aligns with Brandi's conception of the lacuna as a space of interpretative tension, where knowledge is partial, situated, and revisable.

Digital restoration workflows should therefore adopt layered modelling systems that explicitly encode uncertainty. Embedding metadata such as confidence scores, provenance tags, and version histories allows each inferred element to remain visibly distinct from verified fabric. These annotations transform uncertainty into an operational variable, preserving the interpretative transparency of the model. A human-in-the-loop approach remains essential: AI may assist in generating hypotheses, but professional expertise—architectural, historical, and conservation-based—must guide all curatorial and design decisions. Human judgment ensures that interventions remain recognizable, reversible, and ethically accountable.

To guarantee epistemic robustness, restoration research should adopt interdisciplinary validation frameworks. Experimental protocols such as blind testing, comparative metric analyses, and expert assessment can evaluate the reliability and biases of AI-generated reconstructions. These methods transform restoration into a reflexive science of testing and verification. Equally, communication practices must evolve: AI-generated visualizations should include explanatory layers that clarify what is original, reconstructed, or speculative, and indicate the confidence levels and data sources underlying each element. Such transparency fosters public trust and critical literacy in the interpretation of digital heritage.

The ethical dimension of AI-assisted restoration extends beyond methodological precision to encompass authenticity, transparency, and inclusivity. In accordance with Brandi's theory, any digital reintegration must safeguard the historical document value of the monument. The lacuna must remain perceptible—either visually or through metadata—so that temporal stratification and the traces of loss continue to testify to the building's history. Erasing absence in pursuit of aesthetic completeness risks falsifying the document itself. Equally critical is the governance of datasets and algorithms: domain-specific AI models must rely on curated and validated sources, with full documentation of dataset provenance, algorithmic parameters, and bias audits. Transparency in data governance is integral to the ethical legitimacy of digital restoration.

The classical principles of legibility, recognizability, and reversibility must be reinterpreted for the digital domain. Even within virtual models, reconstructed elements should remain clearly identifiable and removable without compromising the integrity of original data. These principles ensure that the digital act of restoration remains faithful to its ethical lineage. Moreover, AI systems must promote cultural diversity and inclusivity: generative reconstructions should not reproduce a single stylistic canon or dominant historiographical narrative. The lacuna signifies the plurality of possible readings; digital reconstructions should therefore embrace multiplicity and reflect the diversity of cultural memory.

Public trust in AI-assisted heritage interpretation depends on fostering critical literacy. Institutions should accompany digital exhibitions and reconstructions with explanatory content about the speculative and interpretative nature of AI outputs. By making uncertainty visible, restoration reaffirms its role as a discipline of critical awareness rather than technological illusion.

The integration of AI into architectural restoration also opens significant opportunities for methodological refinement and empirical research. A key priority is the development of benchmark datasets specifically designed for heritage restoration tasks. Such datasets, incorporating ground-truth cases with intentionally masked zones, would enable the calibration of AI models and facilitate standardized comparisons of predictive performance. This would establish a rigorous foundation for assessing the reliability and limitations of AI-generated reconstructions.

Equally important is the need to explore user perception studies. Visual cues for uncertainty—such as color coding, opacity gradients, or provenance annotations—may strongly influence how both experts and non-experts interpret reconstructed heritage. Systematic studies on perceptual and cognitive responses can inform best practices for interface design, ensuring that uncertainty remains perceptible and meaningful. AI workflows should also support multi-scenario reconstructions: instead of a single "restored" version, systems could generate multiple plausible alternatives, transforming the lacuna from a void into a generative field that invites interpretative exploration.

Longitudinal research is needed to understand how digital reconstructions evolve over time. Data decay, model updates, and new historical findings may alter the accuracy and relevance of previously generated reconstructions. Heritage institutions must therefore develop strategies for versioning, obsolescence management, and systematic updates of AI-derived assets to maintain authenticity and transparency. Finally, policy and standardization frameworks should align conservation principles—such as those articulated in the Nara Document on Authenticity—with emerging norms in AI ethics, data governance, and digital preservation. These guidelines will bridge technical innovation and disciplinary ethics, ensuring responsible and culturally sensitive practices in AI-assisted restoration.

In conclusion, generative AI—when constrained by curated datasets, transparent metadata, and human-centered workflows—can serve not as a creative substitute but as a cognitive partner in restoration. The lacuna, redefined for the digital age, becomes a site of epistemic negotiation between material evidence and computational inference. Restoration thus retains its disciplinary autonomy: a science of interpretation, an ethics of imagination, and a balance between truth and possibility. The ultimate goal is not perfect visual completeness but intelligible incompleteness—a heritage that preserves its temporal depth, its provisionality, and the visible tension between what we know and what we infer. In this sense, the lacuna, whether material or digital, remains a fundamental dimension of architectural heritage—an enduring reminder of loss, time, and the limits of knowledge.

6. REFERENCES

[1] Remondino, Fabio, and Rizzi, Andrea: Reality-based 3D documentation of natural and cultural heritage sites—Techniques, problems, and examples. *Applied Geomatics*, vol. 2, no. 3, pp. 85–100, 2010.

[2] Brandi, Cesare: *Theory of Restoration*, Nardini Editore, Florence, 2005 (original work published 1963).

[3] Carpo, Mario: *Beyond Digital: Design and Automation at the End of Modernity*, MIT Press, Cambridge (MA), 2021.

[4] Manovich, Lev, Arielli Emanuele (2024): *Artificial Aesthetics: Generative AI, Art and Visual Media*. Available from: https://manovich.net/index.php/projects/artificial-aesthetics (retrieved 25/10/2025)

[5] Doerr, Martin: *The CIDOC CRM – An Ontological Approach to Semantic Interoperability of Metadata.*, *AI Magazine*, vol. 24, 2003.

[6] Niccolucci, Franco, Achille Felicetti, and Sorin Hermon: "Populating the Data Space for Cultural Heritage with Heritage Digital Twins" *Data* 7, no. 8: 105, 2022.

[7] Correll, Michael, et al. "Value-driven evaluation of uncertainty visualizations." *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 2018.

Architectural Representation Conditioning Stack (Arcs): Generative Process Multimodal Control

Enrico Pupi[a], Piergiuseppe Rechichi[b]

[a] Department of Architecture and Design, Politecnico di Torino, Italy, enrico.pupi@polito.it;
[b] Department of Energy, Systems, Territory and Buildings Engineering,
Università di Pisa, Italy, piergiuseppe.rechichi@phd.unipi.it

ABSTRACT: This research addresses the limitations of the Text-to-Image (TTI) paradigm in the domain of Generative Artificial Intelligence (GenAI) architectural representation, which relegates designers to a passive human-on-the-loop (HOTL) role. We propose the Architectural Representation Conditioning Stack (ARCS), a hybrid, multimodal framework promoting a human-in-the-loop (HITL) workflow. ARCS integrates three conditioning layers to inference Diffusion Models (DMs) with precision: geometric (L1) via 3D models and ControlNet for spatial coherence; semantic (L2) via Large Language Models (LLMs) for prompt engineering; and stylistic (L3) via Low-Rank Adaptation (LoRA) fine-tuning to encode a specific visual lexicon. Incremental benchmarking demonstrates the framework's success in aligning generated outputs with user intent and expectations. The research further investigates an iterative feedback loop through contextual editing and explores frontiers, such as Image-to-3D/Video, to overcome the limitations of static representation, pointing towards a future of interactive spatial simulation. ARCS offers a draft methodology to transform GenAI from a black box into a creative partner, thereby fostering the algorithmic literacy necessary for contemporary architectural practice.

1. BEYOND THE TEXTUAL PARADIGM: THE WEAKNESS OF THE PROMPT

The integration of Generative AI (GenAI) into architectural representation is driving a profound transformation. Recent literature has systematically documented its potential across all design phases [1], establishing a new upskilling imperative for the Architecture, Engineering, and Construction (AEC) industry and raising fundamental pedagogical questions [2]. In the architectural domain, Diffusion Models (DMs) have opened new possibilities for conceptual exploration and visualization [3]. However, current applications are predominantly confined to the Text-to-Image (TTI) paradigm, especially during initial creative stages [4]. Research has focused on this approach for ideation, praising its capacity for serendipitous discovery [5] while also noting the risks associated with design fixation [6]. This line of inquiry has extended from specific form-finding processes [7] to generating master-quality drawings directly from text prompts [8].

This approach, while democratizing access to powerful visual tools, presents an epistemological disconnect between design intent and its representation. The fundamental research problem lies in the inadequacy of verbal language as the sole means of conveying architectural intent. Natural language, being inherently polysemic [9], excels at describing image semantics but fails to define the spatial and relational syntax central to architecture rigorously. This limitation undermines scientific representation and raises issues of shared authorship [10].
Consequently, the designer is relegated to a passive, human-on-the-loop (HOTL) role, iterating through prompt-crafting with little precise control. This dynamic fuels concerns that an uncritical reliance on AI may lead to a de-skilling of human thought and a loss of personal character in design [11].
The central thesis of this research is that significant progress lies not in refining this flawed approach, but in superseding it. We posit the necessity of a transition towards hybrid, multimodal workflows that shift the designer from a passive HOTL role to an active Human-in-the-

Loop (HITL) one [12]. In this model, human intervention is an integrated, strategic component that orchestrates the generative process, fostering a synergy between design thinking and GenAI [13] to achieve an augmented intuition where intent is encoded directly into the computational process [14].
Therefore, this research defines, implements, and critically analyzes an operational framework named the Architectural Representation Conditioning Stack (ARCS). ARCS enables designers to exert layered control over the output by superimposing geometric (from 3D models), semantic (from text prompts), and stylistic (from fine-tuned visual lexicons) conditioning channels. The primary objective is to demonstrate how this multi-layered conditioning aligns the computational power of GenAI with human intentionality, transforming the generative model from a black-box into a precise and directable computational partner.
This investigation is contextualized within the wider educational program at the Department of Architecture and Design of the Politecnico di Torino. This setting is not incidental; it builds upon a pedagogical trajectory established in teaching parametric modeling at the same institution [15], positioning the ARCS framework as a necessary pedagogical evolution.
The contemporary educational challenge is not merely to teach new software, but to design new processes [16]. This requires training professionals who can balance technology adoption and adaptation [17] and possess a sophisticated algorithmic literacy—the critical competence to deconstruct, assemble, and direct complex systems, for instance, by using semantic AI models to guide ideation [18]. Through the development and analysis of the ARCS framework, this research aims to provide an applied contribution, outlining an approach that promotes a conscious and critical use of emerging generative technologies.

2. DEFINING THE STACK: THE THEORETICAL MODEL OF LAYERED CONDITIONING

The transition from the TTI paradigm to multimodal workflows is no longer a theoretical postulation but an established trajectory, driven by the convergence of academic research and enterprise-level solutions. Before detailing its technical structure, the ARCS framework must be situated within the broader context of architecture's digital turn.
Mario Carpo [19] described the first digital turn as a shift from the alphabet to the algorithm, enabling non-standard seriality. More recently, he theorized the second digital turn as design beyond intelligence [20], where computational power generates solutions so complex they appear alien to human logic, compelling the designer to accept outputs that are no longer intuitively understood. This prospect of an alien intelligence echoes critical concerns about the loss of personal authorship in design [11].
The ARCS framework stands as a direct rebuttal to this thesis. It aims not to operate beyond intelligence but to establish a model for Human-Machine Co-Intelligence. It is a deliberate attempt to re-inject human intentionality, semantics, and stylistic authorship—via its three conditioning layers—into a process that would otherwise become alien.
The state-of-the-art is actively moving beyond an exclusive reliance on language by integrating geometric and spatial inputs as primary constraints for compositional control. Current research is mapping this territory, identifying Transformer Models as having the most significant potential for early-stage design [21] and framing this evolution as a shift from generative algorithms to architectural intelligence within Computer-Aided Architectural Design (CAAD) [22].
Emblematic of this trend is the recent NVIDIA AI Blueprint for 3D-guided generative AI [23]. This pre-configured workflow utilizes a 3D scene as a control scaffold to generate conditioning maps, fundamentally shifting the 3D model's role from a final artifact to a generative input. Industry leaders are developing it, confirming that explicit geometric control is the recognized solution to the compositional uncertainty inherent in TTI approaches.
However, while solving spatial control, such workflows only partially address stylistic and tectonic authorship. Reliance on a pretrained model's intrinsic knowledge or generic prompts often yields results that are aesthetically plausible yet lack authorial nuance. This is the gap ARCS is designed to fill. It does not aim to invent multimodal control but to systematize and extend it. The ARCS framework, therefore, formalizes and stratifies these emerging trends, defining an approach that integrates explicit geometric conditioning with equally explicit complementary controls for the representation style and semantics.

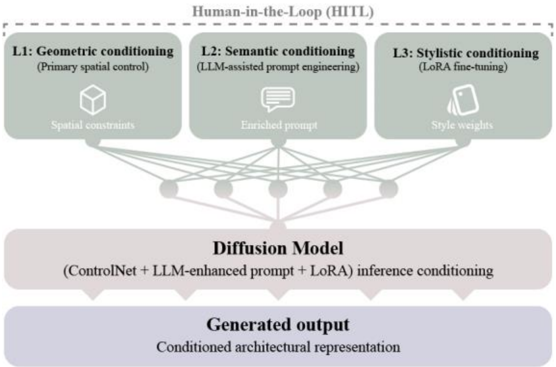

Figure 1: *The ARCS conceptual diagram illustrates how the three conditioning layers are linked to guide the diffusion model's inference process.*

The proposed stack consists of three hierarchical layers that progressively increase the degree of conditioning (Fig. 1):

Primary geometric conditioning (L1): This foundational layer implements the principle demonstrated by solutions like the NVIDIA Blueprint. It imposes spatial structure and ensures architectural coherence via control maps extracted from a 3D model. This approach parallels research on fusing parametric models for architectural design [24].

Assisted semantic conditioning (L2): At this level, the prompt's function is redefined and enhanced through LLM-assisted engineering. The resulting prompt, a product of a hybrid human-LLM dialogue, defines the scene's residual content and atmosphere, establishing a robust semantic base for the final layer.

Authorial stylistic conditioning (L3): This is the core contribution ARCS systematizes, superimposing a precise stylistic intent over the geometric control. This is achieved via Low-Rank Adaptation (LoRA) fine-tuning on a curated visual corpus. The method is validated through parallel research, which defines it as a framework for learning a curated architectural lexicon [25]. Here, the curator's role becomes a fundamental meta-representational act: designing how the project is represented. The critical selection of references is encoded into a computational artifact that infuses the generative process with a specific materiality, atmosphere, and visual language. This enables augmenting large-scale language-image models with a synthetic architecture that is absent from the original training datasets [26]. This research trajectory is rooted in earlier experiments using Generative Adversarial Networks (GANs) [27] for architectural layout generation [28].

Adopting such a layered approach thus requires advanced algorithmic literacy. The architect's required competency shifts from that of a tool user to a process designer, capable of deconstructing, understanding, and consciously orchestrating these complex workflows. This implies a critical understanding beyond functional labels, recognizing that internal mechanisms like self-attention in transformers operate not as human-like focus but as perceptual grouping processes based on feature similarity [29]. In this sense, ARCS is not merely an operational methodology but a framework for exercising critical and intentional control over a process whose computational nature, though complex, can be discretized and directed.

3. PROCESS ARCHITECTURE: TOOLS AND LOGIC OF THE ARCS MULTIMODAL WORKFLOW

The experimental protocol implemented to validate the ARCS framework constitutes the research's technical core, defining the multimodal workflow architecture, tool selection, and operational logic. The protocol was developed entirely within the ComfyUI environment, chosen for its modular and open-source nature. This Visual Programming Language (VPL) environment provides detailed control over every generative parameter, enabling the construction of complex, customized pipelines.

The methodology is founded on the exclusive use of open-weight DMs, specifically FLUX.1-dev [30] by Black Forest Labs. This choice deliberately eschews API-based solutions, as direct access to network weights is a prerequisite for reproducible and extensible academic research. Indeed, only open-weight models permit local fine-tuning—the core operation for authorial stylistic conditioning, as defined by the ARCS framework.

Within ComfyUI, the input is designed as a flexible cluster capable of integrating diverse information sources for a potentially real-time generation process. Geometric intent from 3D modeling software can be fed into the system via two modes: asynchronous loading of pre-rendered views or, more dynamically, using Mixlab's Screen Share node. The latter is particularly significant for experimentation, as it enables a real-time workflow by directly capturing the modeling software's viewport. While constrained by computational power and DM inference times—which could be optimized with lighter models requiring fewer steps (e.g., FLUX.1-schnell)—this approach enables a continuous production cycle and an interactive dialogue between modeling and generation. Notably, this same node can capture input from an

external camera, enabling a significant alternative workflow that involves physical scale models and bridges generative digital representation with traditional plastic modeling practices.
The inference process is managed by the X-Labs sampler, which is selected for its efficiency and compatibility with FLUX.1-dev. The final output is a 2D image that synthesizes the information from the various input channels, processed through the three conditioning layers of the ARCS framework (Fig. 2).

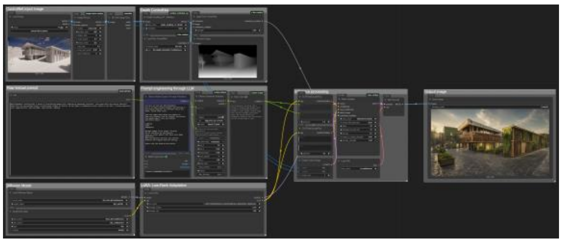

Figure 2: *The nodal interface of the ARCS workflow implemented in ComfyUI. This visual programming environment enables the construction and control of multimodal generative pipelines.*

3.1 GEOMETRIC CONDITIONING: THE 3D MODEL AS A SPATIAL MATRIX

The foundational layer of the ARCS framework is geometric conditioning (L1), which anchors the stochastic generation of the DM to an explicit and controlled spatial and compositional intent. Its function is to translate the three-dimensional architecture into a set of two-dimensional information that acts as the primary constraint for inference, guaranteeing the coherence of form, perspective, and inter-element relationships.
The selected case study is the redevelopment project of Plaza Ponce de León in Seville, originally a 2019 undergraduate thesis by Enrico Pupi. This project was strategically selected for two reasons. Firstly, its intermediate scale—a public space intervention integrating a new building—allows for testing the workflow on a significant yet manageable level of architectural and urban complexity. Secondly, the dense, layered context of Seville provides a robust testbed for evaluating the system's ability to manage the relationship between new interventions and pre-existing heritage, a central challenge in contemporary architectural practice.
The process begins by defining key views within the modeling software. From each view, multiple ControlNet preprocessors can be employed to generate distinct control maps [31]. For this experiment, three types of ControlNet preprocessors in ComfyUI were tested:

Depth Map: This grayscale map encodes the distance of each pixel from the camera. It is the most powerful control for defining the overall volumetrics, scene depth, and spatial arrangement of elements. Even a simple 3D massing model is sufficient to generate an effective depth map that guides the AI in correctly placing objects.
Canny Edge Map: This processor detects sharp contours and discontinuities, making it ideal for imposing the main compositional lines, defining building silhouettes, and precisely constraining well-defined architectural elements.
SoftEdge Map: Preprocessors such as HED (Holistically-Nested Edge Detection) yield softer, more pictorial edge maps resembling freehand sketches. These maps are helpful in the early exploratory phases, providing a less rigid compositional guide that suggests the scene's lines of force without imposing overly restrictive geometric detail.
The selection of these maps within the ComfyUI workflow is the primary method for resolving the compositional randomness inherent in the TTI paradigm, ensuring that every generated image, regardless of its style or content, adheres to the established spatial and formal intent.

3.2 SEMANTIC CONDITIONING: PROMPT ENGINEERING VIA LLM

The second layer of ARCS (L2) is assisted semantic conditioning, managed through LLM-driven text prompt engineering. Within this framework, the prompt is not the sole driver of generation but acts as a control layer, guiding aspects not defined by other conditioning levels. Its importance, however, remains, as prompt quality significantly influences the final image output.
To systematically explore the descriptive potential of language, this protocol integrates an assisted prompt-crafting approach. This aligns with emerging research aimed at enhancing prompt comprehension in DMs through the integration of LLMs [32]. The objective is not to delegate creativity but to leverage an LLM as a tool to expand a core idea into a richer, more varied prompt that is stylistically consistent with the linguistic conventions learned by DMs. For automated prompt generation, the Llama 3 model was employed [33], specifically its 3-billion-parameter (3B) variant, selected for its optimal balance between linguistic capability and computational efficiency. The 3B model proved sufficient for lexical enrichment, while ensuring

entirely local execution and rapid inference times that are compatible with the overall workflow.
Practical implementation was managed via the Ollama framework, which simplifies local LLM deployment. Integration into the ComfyUI pipeline was achieved using custom nodes that act as a bridge, allowing the ComfyUI interface to send basic descriptive inputs to the locally running Ollama service and receive a standardized, enriched prompt in return.
To automate and standardize this engineering process, a system prompt was defined to instruct the Llama 3 model to act as a prompt design expert for DMs. This system prompt enforces precise rules: the model must enrich a user-provided description by adding details on lighting, style, and atmosphere without introducing unsolicited new elements. Formal constraints, such as a maximum length, ensure a clean, directly usable output. Furthermore, the system prompt instructs the LLM to always include the specific trigger word for the intended LoRA, ensuring the generated text is pre-configured to activate the correct stylistic filter in the next layer of the stack. This transforms prompt writing into a hybrid dialogue, where initial human intent is augmented and structured by an LLM before being integrated with the other conditioning layers in the workflow.

3.3 STYLISTIC CONDITIONING: REFERENCES AS A META-REPRESENTATIVE ACT

The third and final layer of ARCS (L3) implements authorial stylistic conditioning. This layer aims to overcome the aesthetic genericity of DMs by injecting a controlled stylistic intent into the process. This is achieved via Parameter-Efficient Fine-Tuning (PEFT), a set of methods designed to adapt large-scale models to specific tasks while minimizing computational costs [34].
Among the available PEFT techniques, including Dreambooth [35], Hypernetworks [36], and Textual Inversion [37], Low-Rank Adaptation (LoRA) was selected due to its superior efficiency [38]. LoRA transforms a reference image set from a moodboard into a tool for transferring specific aesthetic qualities.
The process begins with dataset curation, which is framed here as a meta-representative act. This aligns with emerging CAAD research on teaching designers a curated architectural lexicon [25]. The critical selection of images is not a random assembly but a declaration of intent—an operation that distills the visual essence to be encoded.
To validate the framework’s versatility, three distinct datasets of 40 images each were curated, each representing a well-defined stylistic domain:
ARCS-Biomimetic: Focuses on the symbiotic relationship between architecture and nature, featuring green facades and ambiguous indoor-outdoor boundaries.
ARCS-Hightech: Centers on the celebration of technology and construction detail, highlighting tectonic precision through steel nodes, exposed joints, and technological facades.
ARCS-Brutalist: Explores mass and monumentality, emphasizing the plasticity of exposed reinforced concrete, powerful chiaroscuro effects, and raw textures from formwork.
Once assembled, the technical process was executed within the FluxGym infrastructure. The first step is image captioning. Instead of a manual process, the multimodal model Florence-2 [39], integrated into FluxGym, was used to automatically generate detailed, descriptive captions. These captions were then supervised and refined, with a unique trigger word added to each to activate the specific LoRA during the inference phase.
Using the FluxGym interface, each dataset was used to fine-tune the base FLUX.1-dev model, creating three distinct LoRA files. Training parameters were standardized for comparability. All dataset images were pre-processed to a 512x512 pixel resolution. Training was configured with 10 repeats per image and a maximum of 16 training epochs. This resulted in 6400 training steps for ARCS-Biomimetic (avg. loss 0.33, time ~2h 6m) and ARCS-Brutalist (avg. loss 0.279, time ~2h), and 6240 steps for ARCS-Hightech (avg. loss 0.29, time ~2h 4m). Each fine-tuning process utilized 20 GB of VRAM. Due to LoRA’s efficiency, training required relatively modest computational resources and did not alter the original model’s weights. The result is three lightweight, portable authorial filters—computational artifacts ready for integration into the ComfyUI workflow, completing the ARCS stack.

4. RESULTS, ANALYSIS, AND ARCS FRAMEWORK ASSESSMENT

The critical analysis of the results from ARCS focuses on a systematic and comparative examination of the generative process itself. The aim is to empirically demonstrate how each layer of the framework progressively aligns the output with the design intent, evaluating the degree of control and adherence that each additional stage provides.

To this end, an incremental benchmarking approach was adopted: a series of controlled generative tests where conditioning layers are activated sequentially, allowing the specific impact of each stack component to be isolated and evaluated. All computational operations, from DM inference to LoRA fine-tuning and LLM execution, were performed on a workstation equipped with an Intel(R) Core(TM) i9-14900K processor, 128 GB of RAM, and an NVIDIA GeForce RTX 4090 GPU with 24 GB of GDDR6 VRAM.

The following analysis is structured in three experimental phases:

Phase 1 tests the efficacy of geometric conditioning alone, establishing a baseline to demonstrate how L1 ensures spatial coherence while retaining the base model's generic style.

Phase 2 investigates LLM-assisted semantic conditioning, evaluating how the integration of L2 influences the richness and atmosphere of the image.

Phase 3 presents the complete ARCS results, integrating authorial stylistic conditioning via LoRA. This final stage critically analyzes the synergy between L1, L2, L3, and discusses the balance of their relative weights.

4.1 THE EFFECTIVENESS OF SPATIAL CONTROL VIA CONTROLNET

The first benchmarking phase isolates the efficacy of L1 to establish an operational baseline. In this stage, the workflow exclusively uses Depth Maps as the primary ControlNet input, paired with simple, user-formulated text prompts that indicate a basic stylistic intent. Depth Maps were selected to maximize spatial and volumetric fidelity, as preliminary tests demonstrated their superior capacity to convey the scene's three-dimensional layout to the DM. The results from this phase are summarized in a visual matrix (Fig. 3). The grid is organized for a direct comparison of inputs and outputs. The top row displays the four source views extracted from the 3D model, rendered as clay models to emphasize their pure geometric form. The subsequent rows display the conditioned generation results for each view, utilizing a distinct stylistic intent via its corresponding text prompt.

Figure 3: *The top row displays the source 3D views. Subsequent rows display outputs generated using only ControlNet conditioning (L1), exhibiting high spatial coherence but yielding generic stylistic results based on simple text prompts.*

Analysis of the results reveals a twofold outcome. On the one hand, the success of geometric conditioning is unequivocal. As a vertical comparison across the columns demonstrates, the spatial and compositional structure of each view is imposed with high fidelity across all stylistic variations. The building's volumetrics, its urban context, and furnishing layouts are accurately reproduced. This confirms that L1 achieves its primary objective: ensuring stable geometric control and producing an architecture that is spatially coherent with the design intent defined in the 3D model. On the other hand, the results are only partially stylistically convincing. A horizontal analysis of the rows reveals the DM's limitations when guided solely by simple prompts. In the biomimetic variations, the model correctly applies textures and vegetation but fails to capture the style's deep material sensibility. Similarly, for the high-tech style, it translates the input into metallic and glazed surfaces. However, it fails to render the tectonic details that are its essence, resulting in largely generic architecture. The brutalist style yields a more plausible result; nevertheless, it often lacks the intended monumentality, mass, and dramatic chiaroscuro, at times appearing more unfinished than intentional.

In conclusion, L1 proves to be a necessary and sufficient condition for guaranteeing spatial coherence. However, it also demonstrates that text prompts alone are an insufficient tool for achieving in-depth stylistic control. This finding empirically confirms literature observations that GenAI outputs often require significant manual interpretation and post-processing [2].

4.2 THE ADDITION OF SEMANTIC CONDITIONING VIA LLM

The second benchmarking phase introduces the second ARCS layer, assisted semantic conditioning (L2), while maintaining the constant L1 geometric control. This test aims to evaluate the degree to which text prompt enrichment, driven by the Llama 3 model, can overcome the aesthetic genericity observed in the baseline. The sole variable introduced is the formulation of the prompt, shifting from a simple, direct human input to one engineered by the LLM.

Figure 4*: Results from integrating LLM-engineered prompts (L1+L2). Compared to the baseline, the outputs exhibit greater semantic richness and more sophisticated material and lighting qualities.*

A comparative analysis (Fig. 4) reveals a significant qualitative improvement over the baseline. While maintaining formal fidelity, the generated images exhibit superior chromatic richness and a more sophisticated material rendering. The engineered prompt, being more specific in describing light conditions, atmospheric effects, and surface qualities, acts as a semantic amplifier, guiding the DM toward more complex and targeted interpretations of the requested style. This second phase demonstrates that L2 is an effective tool for enhancing the quality of the output and stylistic coherence, serving as a refined layer of semantic control. However, it is essential to note that this is still an interpretation—albeit a more sophisticated one—of the base model's intrinsic stylistic knowledge. The result is an improved version, but not one necessarily aligned with a specific reference lexicon.

4.3 COMPLETE CONDITIONING VIA LORA

The third and final experimental phase represents the full implementation of the ARCS framework, integrating L1, L2, and L3. Building upon the preceding layers, this test introduces authorial stylistic conditioning (L3) by activating specific LoRA models. The objective is to evaluate the capacity of ARCS to produce images that faithfully adhere to a predefined and designer-curated visual lexicon.

Figure 5: *Results from the complete ARCS workflow (L1+L2+L3), integrating stylistic conditioning via LoRA. The outputs show high adherence to the specific visual lexicon of each curated dataset.*

The analysis of the results (Fig. 5) demonstrates the success of the ARCS framework. Images produced in this phase achieve a level of detail and stylistic adherence markedly superior to that of the previous stages. The synergy among the three conditioning layers yields an output where geometric coherence is preserved, semantic richness is maintained, and stylistic adherence is specific and intentional. Comparing these results with previous ones, where the DM merely interpreted a style, it now applies a precise visual pattern learned from the reference corpus.

This result directly addresses critical concerns from the literature regarding the loss of personal character [11] and the issue of shared authorship [10]. The complete ARCS workflow proves to be a practical methodology for preserving and embedding authorial intent. The results embody the synthetic architectures for which theoretical research argues—the creation of stylistic and formal syntheses that do not exist in the original training datasets [26].

However, this operational success introduces a further dimension of control: the balancing of weights. Generation is not a static but a dynamic act of orchestration. The architecture of DMs is designed for such control, primarily through mechanisms like Classifier-Free Guidance scale (CFG), which allows manipulating the model's adherence to provided conditioning [40]. Within ComfyUI, the designer can leverage this principle to adjust the influence of each component. For instance, the ControlNet weight can be increased for higher geometric fidelity—at the risk of overpowering the LoRA's influence—or slightly decreased to allow the DM to propose style-coherent micro-variations. Similarly, the LoRA weight can be modulated for stricter or looser stylistic adherence, even enabling the exploration of stylistic hybrids.

5. BEYOND THE STATIC IMAGE: REVIEWING OUTPUT AND THE FRONTIERS OF REPRESENTATION

While the ARCS framework demonstrates the potential for multimodal control over the generative process, the workflow described thus far culminates in an intrinsically static artifact: the 2D image. Architectural practice, by contrast, is not a linear process ending in a single representation, but an iterative cycle of proposal, analysis, and revision. Therefore, the generated image, however faithful to the initial intent, is not an end-product but a visual hypothesis that must be subjected to critique, discussion, and refinement.

We therefore propose two complementary research directions (Fig. 6). The first aims to investigate technologies that enable direct interaction with and refinement of the generated output, eliminating the need to restart the entire process.

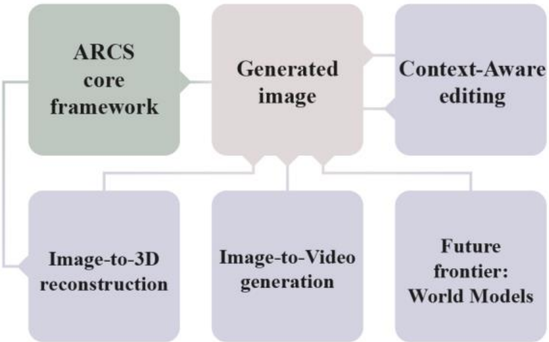

Figure 6: *This diagram illustrates the research trajectories beyond the static 2D image, including an iterative refinement loop (context-aware editing) and explorations into dynamic and three-dimensional outputs.*

The second direction explores emerging technologies designed to overcome the fundamental limitations of 2D representation. The production of conditioned yet static images constitutes a cognitive and operational bottleneck for a discipline that is inherently spatial and temporal in its nature. Consequently, advanced research is shifting its focus from static outputs to experiential simulations and the generation of three-dimensional content.

5.1 ITERATIVE REFINEMENT: CONTEXT AWARE EDITING

The high-quality image produced by the ARCS framework is not the endpoint of the design process, but rather the starting point for a critical phase of revision and refinement. Effectively closing this feedback loop requires tools that enable a dialogic interaction with the output, allowing for targeted modifications without regenerating the entire scene from scratch. This section explores this possibility through the use of FLUX.1-kontext [41], a next-generation contextual editing DM. Unlike simpler techniques such as inpainting, FLUX.1-kontext exhibits an advanced contextual understanding of the entire image, enabling complex and coherent modifications via text-based instructions. The selection of FLUX.1-kontext is, once again, driven by the open-weight principle. Although proprietary alternatives like Gemini 2.5 Flash Image (Nano Banana) exist, access to the model's weights is essential in a research context, as it enables future fine-tuning of the editing model itself for specialization in the architectural representation domain.

Inference was handled locally via a dedicated ComfyUI workflow (Fig. 7), using the 12 final ARCS-generated images as input. Interaction with the model was conducted through text prompts describing the desired modification. Tested interventions spanned various scales of complexity, including altering lighting conditions, modifying the materiality of specific elements, and reconfiguring the interior space by adding or removing components.

Figure 7: *The ComfyUI pipeline for iterative refinement using FLUX.1-kontext, which takes an ARCS-generated image as input and allows for targeted modifications through text-based instructions.*

Analysis of the results (Fig. 8) demonstrates the high capability of FLUX.1-kontext to interpret instructions correctly and apply them coherently, preserving the overall style and visual quality previously established by the ARCS framework. The modifications integrate seamlessly into the image, confirming the efficacy of this approach for an efficient iterative refinement phase.

Figure 8: *Results of contextual editing. This matrix displays examples of targeted modifications, demonstrating the model's ability to refine elements while preserving overall stylistic coherence.*

However, the technology's current limitations must be acknowledged. While excellent for in-scene edits, spatial coherence may degrade in response to requests for radical viewpoint shifts. Consequently, future research should explore the integration of complementary models, such as SeeDream 4.0 [42], which excel specifically in maintaining spatial consistency across multiple viewpoints, thereby prefiguring more advanced refinement workflows.

5.2 TOWARDS DYNAMIC REPRESENTATION: IMAGE-TO-VIDEO AND IMAGE-TO-3D EXPLORATIONS

While the iterative loop based on 2D image editing has proven effective for compositional refinement, it remains confined by the intrinsic limitations of static representation. Architecture is inherently an experience that unfolds in space and time, and project evaluation cannot be detached from these dimensions. The final phase of this research thus extends beyond the static image, exploring two emerging technological frontiers that aim to reintroduce temporality and three-dimensionality into the generative process.
The first frontier is Image-to-Video generation. These tools animate a static image by simulating limited camera motion or dynamic environmental effects, enabling a pre-visualization of the *promenade architecturale* to convey a sense of spatial experience. The experiment involves testing various models to map the state-of-the-art, including LTX-Video [43], Hunyuan Image-to-Video [44], and Tongyi Wanxiang 2.2. While proprietary alternatives, such as VEO3 [45] or SORA 2 [46], demonstrate remarkable capabilities, preliminary tests show that even more accessible solutions can provide short clips useful for evaluating spatial dynamics.
The second, and arguably most promising, frontier is Image-to-3D reconstruction [47, 48]. These models reverse the rendering process by inferring three-dimensional geometry from a single 2D image. This capability is critical for closing the design loop: an image generated and refined via ARCS can be converted back into a preliminary 3D model for re-import into modeling software, creating a bidirectional bridge between 2D generation and 3D modeling. The experiment involves testing state-of-the-art solutions, such as Hunyuan3D-2 [49], TripoSR [50], PartPacker [51], and Stable Point Aware 3D [52]. Their ability to reconstruct an approximate point cloud or polygon mesh from the ARCS workflow output prefigures a future where 2D stylistic and material exploration can directly and rapidly inform the subsequent digital modeling phase. While this mapping includes closed-source solutions, the long-term research trajectory must pivot towards the adoption and development of open-source or at least open-weight tools. Only an open approach ensures the transparency, reproducibility, and, most importantly, the ability to specialize and refine these models for the specific needs of the architectural domain, guaranteeing they remain user-controllable and manipulable tools.

6. THE PARADIGM SHIFT: CONCLUSIONS AND FUTURE OUTLOOK

This research has systematized generative conditioning in GenAI-assisted architectural representation. Incremental benchmarking demonstrates that the ARCS framework, proposed here as a draft operational model, is a practical methodology for overcoming the limitations of the TTI paradigm.
A critical analysis of the results reveals a fundamental paradigm shift: architectural representation is transformed into a design process that is hybrid, iterative, and multimodal. The ARCS framework, in particular, offers a direct technical and methodological response to urgent critiques in recent literature. This technical demonstration shows that the process designer can avoid depriving individuals of their thoughts and personal character [11]. The adoption of such multimodal frameworks appears to be a primary pathway toward a sustainable balance between technology adoption and adaptation [17] for both the profession and academia.
The implications for the discipline of representation are profound. Future architects must be trained as professionals possessing a critical algorithmic literacy, capable of deconstructing,

assembling, and directing these complex systems. The ARCS framework actualizes the synergies between design thinking and AI [13]. It provides an operational protocol for building augmented intuition [14], which must be the goal of our interaction with such systems.

It must be emphasized that ARCS is a work-in-progress framework, a model that is undergoing consolidation, whose evolution is intrinsically linked to the accelerated development of technology. Looking beyond the current horizon, the most compelling trajectory lies not in refining the generation of images, videos, or 3D models, but in the emergence of World Models [53]. State-of-the-art technologies, such as DeepMind's Genie3 or Word Labs' Marble, prefigure a future where AI simulates interactive experiences rather than merely representing space. The revolutionary characteristic of these models is their ability to generate experiential and freely navigable 3D environments without underlying 3D geometry, with each frame generated in real-time. Although open-source alternatives like Matrix-Game by SkyworkAI or Hunyuan-GameCraft by Tencent are still in their infancy, their potential for architecture is notable. The focus will no longer be on generating static artifacts, but on dynamic, interactive environments for testing the perception, movement, and experience of a space at a preliminary conceptual stage. This frontier shifts the focus from the representation of form to the simulation of experience, holding the potential to become one of the most effective tools in architecture in the GenAI era.

7. ACKNOWLEDGMENT

This paper is the result of scientific research carried out by the authors. E. Pupi wrote paragraphs 3, 3.1, 3.2, 3.3, 4, 4.1, 4.2, and 4.3. P. Rechichi wrote paragraphs 2, 5, 5.1, and 5.2. The authors wrote paragraphs 1 and 6 together.

8. REFERENCES

[1] Bölek, Buse; Tutal, Osman; Özbaşaran, Hakan: A systematic review on artificial intelligence applications in architecture. *Journal of Design for Resilience in Architecture and Planning*, vol. 4, no. 1, pp. 91–104, 2023.

[2] Onatayo, Damilola; Onososen, Adetayo; Oyediran, Abiola Oluwasogo; Oyediran, Hafiz; Arowoiya, Victor; Onatayo, Eniola: Generative AI Applications in Architecture, Engineering, and Construction: Trends, Implications for Practice, Education & Imperatives for Upskilling—A Review. *Architecture*, vol. 4, no. 4, pp. 877–902, 2024.

[3] Li, Chengyuan; Zhang, Tianyu; Du, Xusheng; Zhang, Ye; Xie, Haoran: Generative AI Models for Different Steps in Architectural Design: A Literature Review. *arXiv:2404.01335*, 2024.

[4] Bank Stigsen, Mathias; Moisi, Alexandra; Rasoulzadeh, Shervin; Schinegger, Kristina; Rutzinger, Stefan: AI Diffusion as Design Vocabulary - Investigating the use of AI image generation in early architectural design and education. In: Dokonal, W; Hirschberg, U; Wurzer, G (eds), *Digital Design Reconsidered - Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2023) - Volume 2*, Graz, 2023, pp. 587–596.

[5] Paananen, Ville; Oppenlaender, Jonas; Visuri, Aku: Using text-to-image generation for architectural design ideation. *International Journal of Architectural Computing*, vol. 22, no. 3, pp. 458–474, 2024.

[6] Hanafy, Nervana Osama: Artificial intelligence's effects on design process creativity: "A study on used A.I. Text-to-Image in architecture". *Journal of Building Engineering*, vol. 80, 2023.

[7] Jaruga-Rozdolska, Anna: Artificial intelligence as part of future practices in the architect's work: MidJourney generative tool as part of a process of creating an architectural form. *Architectus*, vol. 3, no. 71, pp. 95–104, 2022.

[8] Chen, Junming; Wang, Duolin; Shao, Zichun; Zhang, Xu; Ruan, Mengchao; Li, Huiting; Li, Jiaqi: Using Artificial Intelligence to Generate Master-Quality Architectural Designs from Text Descriptions. *Buildings*, vol. 13, no. 9, pp. 2285, 2023.

[9] Bolojan, Daniel; Vermisso, Emmanouil; Yousif, Shermeen: Is Language All We Need? A Query Into Architectural Semantics Using a Multimodal Generative Workflow. In: van Ameijde, Jeroen; Gardner, Nicole; Hyun, Kyung Hoon; Luo, Dan; Sheth, Urvi (eds), *POST-CARBON - Proceedings of the 27th CAADRIA Conference*, Sydney, 2022, pp. 353–362.

[10] Mancini, Matteo Flavio; Menconero, Sofia: AI-aided Design? Text-to-image Processes for Architectural Design. *diségno*, no. 13, pp. 57–70, 2023.

[11] Kwon, Choulwoong; Ahn, Yonghan: Critical views on AI (Artificial Intelligence) in building design. *International Journal of Sustainable Building Technology and Urban Development*, vol. 15, no. 2, pp. 240–246, 2024.

[12] Shneiderman, Ben: Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction*, vol. 36, no. 6, pp. 495-504, 2020.

[13] Sreenivasan, Aswathy; Suresh, M.: Design thinking and artificial intelligence: A systematic literature review exploring synergies. *International Journal of Innovation Studies*, vol. 8, no. 3, pp. 297–312, 2024.

[14] Bernhard, Matthias; Smigielska, Marta; Dillenburger, Benjamin: Augmented Intuition. Encoding Ideas, Matter and Why It Matters. In: As, Imdat; Basu, Prithwish (eds), *The Routledge Companion to Artificial Intelligence in Architecture*, Routledge, London, 2021, pp. 405–419.

[15] Lo Turco, Massimiliano; Tomalini, Andrea; Bono, Jacopo: The teaching of parametric modeling in the School of Architecture at the Politecnico di Torino. 20 Architecture Experience - Proceedings of International Conference of Architecture and Design, Bucarest, 30-31/10/2023, Accademia Adrianea Edizioni, Bucarest, pp. 72–81.

[16] Atalay, Fatmanur: Artificial Intelligence in Architectural Education: New Learning and Design Approaches. *Proceedings of the international conference of contemporary affairs in architecture and urbanism-ICCAUA*, vol. 8, no. 1, pp. 343–355, Jul./2025.

[17] Mance, Damir: Architectural practice in the Digital Age: Balancing Adoption and Adaptation. In: Barosio, Michela; Vigliocco, Elena; Gomes, Santiago (eds), *School of Architecture(s) - New Frontiers of Architectural Education*, Springer Nature Switzerland, Cham, 2025, pp. 349–361.

[18] Vermisso, Emmanouil: Semantic AI models for guiding ideation in architectural design courses. 13th International Conference on Computational Creativity, Bozen-Bolzano, Italy, 27/06/2022-01/07/2022, Association for Computational Creativity.

[19] Carpo, Mario: *The Alphabet and the Algorithm, Writing Architecture*, MIT Press, Cambridge, MA, 2011.

[20] Carpo, Mario: *The second digital turn: Design beyond intelligence*, The MIT Press, Cambridge, MA, 2017.

[21] Vissers-Similon, Elien; Dounas, Theodoros; De Walsche, Johan: Classification of artificial intelligence techniques for early architectural design stages. *International Journal of Architectural Computing*, vol. 23, no. 2, pp. 387–404, 2025.

[22] Stojanovski, Todor; Zhang, Hui; Frid, Emma; Chhatre, Kiran; Peters, Christopher; Samuels, Ivor; Sanders, Paul; Partanen, Jenni; Lefosse, Deborah: Rethinking Computer-Aided Architectural Design (CAAD) -- From Generative Algorithms and Architectural Intelligence to Environmental Design and Ambient Intelligence. In: Gerber, David; Pantazis, Evangelos; Bogosian, Biayna; Nahmad, Alicia; Miltiadis, Constantinos (eds), *Computer-Aided Architectural Design. Design Imperatives: The Future is Now*, Springer Singapore, Singapore, 2022, pp. 62–83.

[23] NVIDIA: Generative AI 3D Guided [online], Available from: https://build.nvidia.com/nvidia/genai-3d-guided (retrieved on 22 October 2025).

[24] Sebestyen, Adam; Özdenizci, Ozan; Legenstein, Robert; Hirschberg, Urs: AI-Infused Design: Merging parametric models for architectural design. In: Kontovourkis, O; Phocas, MC; Wurzer, G (eds), *Data-Driven Intelligence - Proceedings of the 42nd Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2024) - Volume 1*, Nicosia, 2024, pp. 489–498.

[25] Kim, Frederick Chando; Johanes, Mikhael; Huang, Jeffrey: Text2Form Diffusion: Framework for learning curated architectural vocabulary. In: Dokonal, W; Hirschberg, U; Wurzer, G (eds), *Digital Design Reconsidered -*

Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2023) - Volume 1, Graz, 2023, pp. 79–88.

[26] Koehler, Daniel: More than anything: Advocating for synthetic architectures within large-scale language-image models. *International Journal of Architectural Computing*, vol. 21, no. 2, pp. 242-255, 2023.

[27] Goodfellow, Ian J.; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua: Generative adversarial nets. 28th International Conference on Neural Information Processing Systems (NIPS'14), Montreal, Canada, 08-13/12/2014, MIT Press, Cambridge, MA, pp. 2672–2680.

[28] Chaillou, Stanislas: AI and architecture. An experimental perspective. In: As, Imdat; Basu, Prithwish (eds), *The Routledge Companion to Artificial Intelligence in Architecture*, Routledge, London, 2021, pp. 27–48.

[29] Mehrani, Paria; Tsotsos, John K.: Self-attention in vision transformers performs perceptual grouping, not attention. *Frontiers in Computer Science*, vol. 5, 2023.

[30] Greenberg, Or: Demystifying Flux Architecture. *arXiv:2507.09595*, 2025.

[31] Zhang, Lvmin; Rao, Anyi; Agrawala, Maneesh: Adding Conditional Control to Text-to-Image Diffusion Models. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 02-06/10/2023, IEEE, Piscataway, NJ, pp. 3813–3824.

[32] Lian, Long; Li, Boyi; Yala, Adam; Darrell, Trevor: LLM-grounded Diffusion: Enhancing Prompt Understanding of Text-to-Image Diffusion Models with Large Language Models. *arXiv:2305.13655*, 2023.

[33] Grattafiori, Aaron et al.: The Llama 3 Herd of Models. *arXiv:2407.21783*, 2024.

[34] Han, Zeyu; Gao, Chao; Liu, Jinyang; Zhang, Jeff; Zhang, Sai Qian: Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey. *arXiv:2403.14608*, 2024.

[35] Ruiz, Nataniel; Li, Yuanzhen; Jampani, Varun; Pritch, Yael; Rubinstein, Michael; Aberman, Kfir: DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 18-22/06/2023, IEEE, Piscataway, NJ, pp. 22500–22510.

[36] Ha, David; Dai, Andrew; Le, Quoc V.: HyperNetworks. *arXiv:1609.09106*, 2016.

[37] Gal, Rinon; Alaluf, Yuval; Atzmon, Yuval; Patashnik, Or; Bermano, Amit H.; Chechik, Gal; Cohen-Or, Daniel: An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. *arXiv:2208.01618*, 2022.

[38] Hu, Edward J.; Shen, Yelong; Wallis, Phillip; Allen-Zhu, Zeyuan; Li, Yuanzhi; Wang, Shean; Wang, Lu; Chen, Weizhu: LoRA: Low-Rank Adaptation of Large Language Models. *arXiv:2106.09685*, 2021.

[39] Xiao, Bin; Wu, Haiping; Xu, Weijian; Dai, Xiyang; Hu, Houdong; Lu, Yumao; Zeng, Michael; Liu, Ce; Yuan, Lu: Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks. *arXiv:2311.06242*, 2023.

[40] Ho, Jonathan; Salimans, Tim: Classifier-Free Diffusion Guidance. *arXiv:2207.12598*, 2022.

[41] Batifol, Stephen et al.: FLUX.1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space. *arXiv:2506.15742*, 2025.

[42] Chen, Yunpeng et al.: Seedream 4.0: Toward Next-generation Multimodal Image Generation. *arXiv:2509.20427*, 2025.

[43] HaCohen, Yoav et al.: LTX-Video: Realtime Video Latent Diffusion. *arXiv:2501.00103*, 2024.

[44] Kong, Weijie et al.: HunyuanVideo: A Systematic Framework For Large Video Generative Models. *arXiv:2412.03603*, 2025.

[45] Google (2024): Veo 3 Technical Report [online], Available from: https://storage.googleapis.com/deepmind-media/veo/Veo-3-Tech-Report.pdf (retrieved on 22 October 2025).

[46] OpenAI (2024): Video generation models as world simulators [online], Available from: https://openai.com/index/video-generation-models-as-world-simulators/ (retrieved on 22 October 2025).

[47] Poole, Ben; Jain, Ajay; Barron, Jonathan T.; Mildenhall, Ben: DreamFusion: Text-to-3D using 2D Diffusion. *arXiv:2209.14988*, 2022.

[48] Mildenhall, Ben; Srinivasan, Pratul P.; Tancik, Matthew; Barron, Jonathan T.; Ramamoorthi, Ravi; Ng, Ren: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In: Vedaldi, Andrea; Bischof, Horst; Brox, Thomas; Frahm, Jan-Michael (eds), *Computer Vision -- ECCV 2020*, Springer International Publishing, Cham, 2020, pp. 405–421.

[49] Zhao, Zibo et al.: Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation. *arXiv:2501.12202*, 2025.

[50] Tochilkin, Dmitry; Pankratz, David; Liu, Zexiang; Huang, Zixuan; Letts, Adam; Li, Yangguang; Liang, Ding; Laforte, Christian; Jampani, Varun; Cao, Yan-Pei: TripoSR: Fast 3D Object Reconstruction from a Single Image. *arXiv:2403.02151*, 2024.

[51] Tang, Jiaxiang; Lu, Ruijie; Li, Zhaoshuo; Hao, Zekun; Li, Xuan; Wei, Fangyin; Song, Shuran; Zeng, Gang; Liu, Ming-Yu; Lin, Tsung-Yi: Efficient Part-level 3D Object Generation via Dual Volume Packing. *arXiv:2506.09980*, 2025.

[52] Huang, Zixuan; Boss, Mark; Vasishta, Aaryaman; Rehg, James M.; Jampani, Varun: SPAR3D: Stable Point-Aware Reconstruction of 3D Objects from Single Images. *arXiv:2501.04689*, 2025.

[53] Ha, David; Schmidhuber, Jürgen: World Models. *arXiv:1803.10122*, 2018.

From Early Design Geometry to Architectural Vision: Open-Source Generative Workflows for Speculative Design

Giulio Lucio Sergio Sacco and Carlo Battini

DICCA, Dept. of Civil, Chemical and Environmental Engineering, University of Genoa, Italy,
giulioluciosergio.sacco@edu.unige.it, carlo.battini@unige.it

ABSTRACT: Recent advances in generative AI have introduced powerful tools for creating architectural imagery, yet most workflows remain opaque, uniquely prompt-driven, and stylistically biased. This paper uses an open workflow for guiding text-to-image generation using Blender and ControlNet-based conditioning, integrating parametric modelling with textual input through depth maps. The result is a controlled, reproducible, and locally executed rendering pipeline that can combines design logic with speculative visual output in an autonomous, non-proprietary environment that could be used in early design phases for the exploration visual possibilities.

Beyond illustrating this workflow, this research proposes a reflection on the current limitation of generative AI in architectural field. It compares a gallery of outputs generated referencing distinct artistic styles, like the ones of Piranesi or Bosch, and more general stylistic categories to preliminary assess recurrent compositional biases, visual stereotypes and limited stylistic fidelity that could constrain architectural imagination and may homogenise visual culture, urging designers to engage critically with its aesthetic boundaries.

1. INTRODUCTION

Over the past few years, generative artificial intelligence has profoundly transformed the way visual content is conceived, produced, and disseminated. Diffusion and transformer-based image models now allow users to produce complex imagery from natural-language prompts alone, reshaping creative workflows across arts, design, and architecture. With the announcement of Sora 2, OpenAI also released a self-contained social platform complete with feed exclusively dedicated to AI content creation and fruition. Yet, despite their accessibility, these systems remain opaque in both their training and operation, offering limited transparency about how specific styles, compositions, or iconographies are learned and reproduced. This issue has already been highlighted in social and media studies, both regarding text-to-image models and LLMs [1, 2]. In the same way, in visual representation, they risk of narrowing rather than expanding the range of visual expression available to their users.

The aim of this paper is to start an investigation on how AI models may present and, therefore, reproduce stylistic biases impairing their rendering capabilities, and consequently influence the aesthetic palette available for architectural visualisation and speculative design. The research question arises during the development of a hybrid 3D/text-to-image workflow aimed at generating prompt guided architectural rendering. Here we describe the workflow and propose some rendering examples to problematize how such stylistic biases could gradually shape our shared visual culture by privileging certain canons, atmospheres, or compositional logics over others.

To highlight this issue, two artists with very distinctive styles were selected: Giovanni Battista Piranesi and Hieronymus Bosch, that could be exemplified by their famous *Carceri d'Invenzione* [3] and *The Garden of Earthly Delights* [4] respectively. Both are highly recognisable, historically unique, and likely underrepresented in training datasets, making them ideal references for prompt generation with the goal of evaluating the capacity of generative systems to emulate stylistic intent without prior overexposure. Subsequently, more prompts were used to generate images with definite but less specific styles.

A comparative test was conducted using a set of several open-source diffusion and transformer models executed locally through ComfyUI, (e.g. Stable Diffusion Family, Flux, Hunyuan) and a proprietary one (DALL·E 3). Each model was prompted using identical textual inputs, both in prompt-only configurations and in geometry-conditioned mode, where depth maps generated from parametric 3D geometries in Blender were employed through ControlNet conditioning. This dual setup allows qualitative assessment of the compositional capabilities of the different models and how 3D derived depth conditioning could be used in early design phases analysing spatial composition and stylistic adherence

This research, therefore, positions itself as both a technical and critical reflection: it outlines an open, reproducible workflow for geometry-guided image generation, while questioning the possible cultural consequences of its use for design purposes due to lacks in stylistic representation.

2. MAIN ASPECTS

The proposed mixed workflow aims to directly bridge Blender parametric modelling, or, in principles, any 3D software capable to render required the required images, with ComfyUI. This integration leverage ComfyUI's flexible node-based interface to orchestrate the generation process, enabling users to structure complex AI pipelines and access ControlNet-based conditioning [5] for most diffusion models. It runs locally on the user's machine and hosts a lightweight web server, enabling seamless communication with Blender.

The architectural geometry used as conditioning input was modelled parametrically in Blender Geometry Nodes. It represents an imaginary complex structure loosely inspired by Piranesi's Carceri, defined only in broad compositional terms. The depth map (Figure 1.a) of the scene was rendered from the camera view using Cycles engine, then the compositor Z-distance output was normalised and saved as a PNG. Other possible conditioning inputs, such as normal maps (Figure 1.b), Canny edge maps and semantic segmentation maps were tested but gave unstable or less coherent results compared to the depth maps; therefore, only the latter were eventually used.

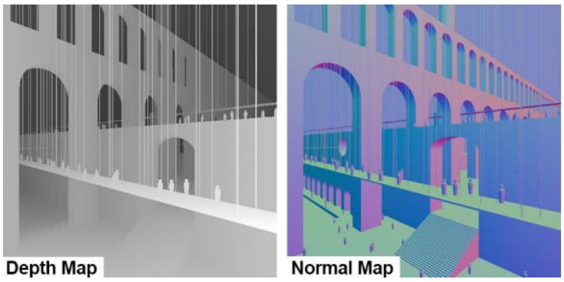

Figure 1: *Example of Depth and Normal maps representing a preliminary volumetric model.*

The models listed in Table 1 were chosen to represent both the historical evolution of diffusion models (stable diffusion family) and the current state of the art. The main Stable Diffusion models were included from version 1.5, released in August 2022 to 3.5, released in October 2024, while Flux.1 Dev and Hunyuan 1.2 DiT were added as examples of most recent open-source systems. These models were used locally through the latest version of ComfyUI. Additionally, to represent the most accessible prompt-only systems available to the average user, the same prompts were provided to DALL·E 3 accessed through ChatGPT browser interface. Microsoft Copilot and Grok were also tested but running a derivative of DALL·E 3 and Stable Diffusion 2.1 their results were considered redundant and omitted here. In the latter three cases, ControlNet was no available; therefore, only the pure text-to-image configuration was used for testing.

Model	Architecture	Size	Training dataset	Release date	Owner licence
SD 1.5	LDM	1B	LAION-2B	Aug '22	Stability AI open source
SD XL	LDM	2B	Nd	Jul '23	Stability AI open source
SD 3.5	LDM	nd	LAION + licensed	Oct '24	Stability AI open source*
Hunyuan 1.2 DiT	DiT	nd	Nd	Jul '24	Tencent open source
Flux.1 Dev	DiT	12B	LAION + licensed	Aug '24	Black Forest open source°
DALL·E 3 ChatGPT	Transformer	nd	Nd	Oct '23	OpenAI closed source

Table 1: *Summary of models tested. * community licence; ° Dev version only; nd = no data disclosed.*

The offline image generation were carried out on a workstation equipped with an NVIDIA RTX 4080 using CUDA acceleration. Each image was produced with 25 sampling steps and a random seed. ControlNet-Depth was used with a strength of 0.66, without any post-processing or compositing applied afterwards. Reducing the strength of the depth conditioning was needed to allow the model to apply variations to the input flat surfaces. All input maps and resulting images were square and kept at 1024 × 1024 pixels for standardization and consistency, except for Stable Diffusion 1.5, which was limited to 512 × 512 pixels due to its native architecture.

Each model was prompted with the same textual prompt for the two main stylistic targets (i.e. Piranesi and Bosch). Three additional prompts were added to generate "renaissance" and modern architecture, and cyberpunk scenery for a comparison baseline. For each case, 32 images were generated both using the text prompt and depth map, and with textual prompt only as a reference. When reported in the paper, out of the 32 images generated only the one considered most representative of the group for style, architectural and rendering quality was arbitrary selected and reported. This choice was made through direct visual judgement rather than quantitative scoring, in line with the qualitative and exploratory nature of the research.

2.1 ARTIST-SPECIFIC PROMPTING

The pictures in Figures 2 and 3 were generated with the following prompt, explicitly reporting the desired style by the author's name: *"A vast, imaginary prison interior rendered in the style of Piranesi's Carceri d'Invenzione. Etched textures, dramatic chiaroscuro, and intricate hatching define staircases, arches, and suspended bridges fading into shadow. The architecture expands infinitely in every direction, a visionary depiction of confinement, scale, and sublime architectural imagination".* The first set of images is reported in Figure 2 and shows a certain adherence to the desired style of Piranesi's Carceri, however, excluding DALL·E which provide a coherence both in terms of composition and rendering style, the other models show a significant bias towards realistic rendering, especially in lights and shadows. A significant bias for central perspective, not representative of Piranesi's works, was observed in the composition across all models: 25, 21 and 29 central perspective compositions out of the 32 images generated for each model were observed in Stable Diffusion 1.5, XL and 3.5 respectively, and 25 and 18 over 32 for Flux and Hunyuan. Out of the remaining, two-point perspective is the most represented, but three-point perspectives were present as well. This significant deviation from Piranesi's style [6] could be motivated by a lack of references in the training sets.

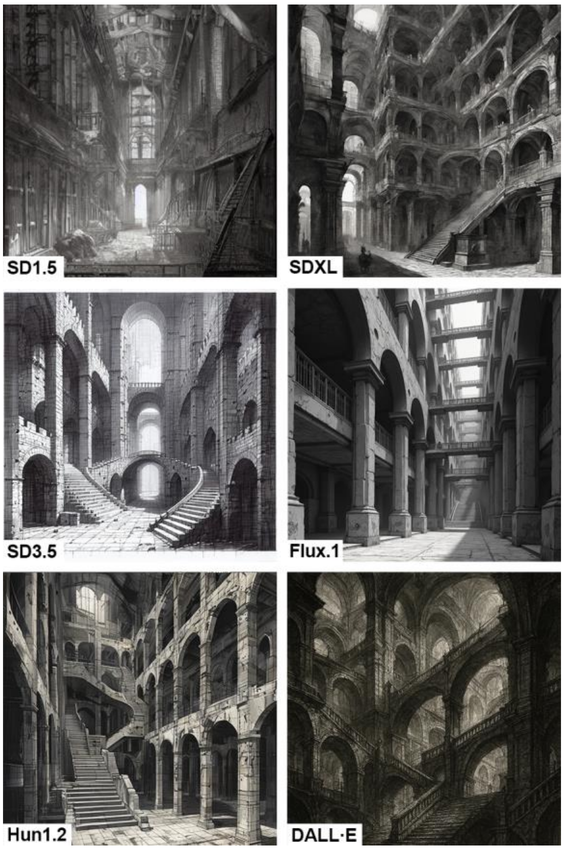

Figure 2: *Images generated with "Piranesi" prompt and different models.*

The images in Figure 3 were generated using depth conditioning. The use of a depth ControlNet adds a complexity layer that makes the direct comparison between the models less direct as the control networks interact differently with each one. For this reason, a fixed strength value was adopted only for consistency, but better results would be obtained adjusting this parameter for each case. However, the previous observation regarding the overall style adhesion holds also for this image set: Stable Diffusion models seem more capable of recreating a drawing rendering; among the three Stable diffusion 3.5 is the best in reproducing an hatch-like texture as it tends to produce finer and more intricate details and texture in general; and Flux and Hunyuan seems more biased toward photorealism, although preserving a stylized drawing rendering. Regarding composition, satisfactory results were obtained in all case; however the adherence to Piranesi style was further reduced likely due to conditioning competition.

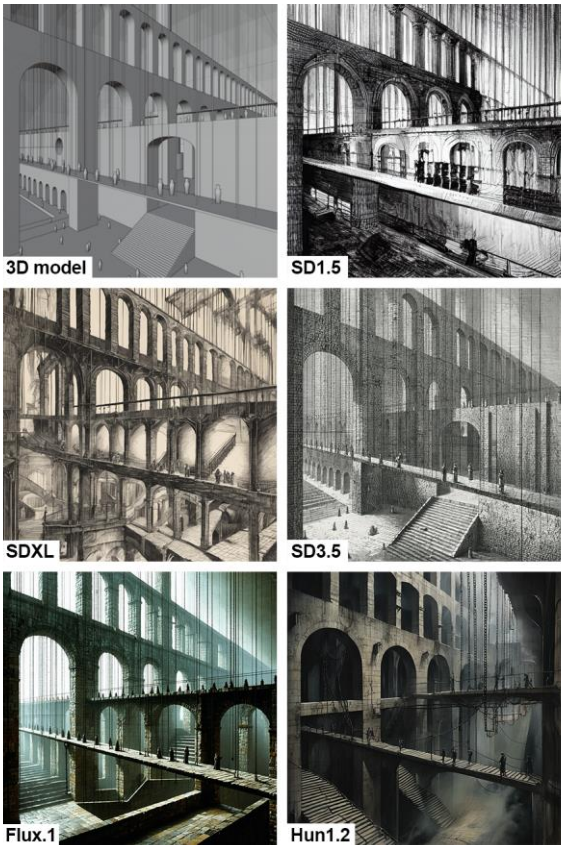

Figure 3: *Images generated with "Piranesi" prompt and depth ControlNet. The first picture reports the reference 3D model.*

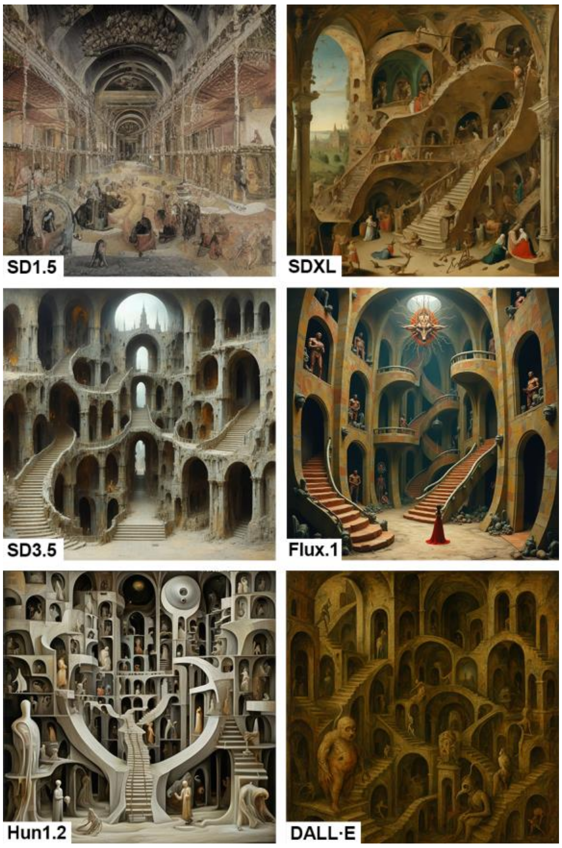

Figure 4: *Images generated with "Bosch" prompt and different models.*

The same approach was adopted to generate the second set of images based on the style of Hieronymus Bosch using the following prompt: *"A labyrinthic interior drawn in the style of Hieronymus Bosch, filled with surreal staircases, arches, and bridges twisting through dreamlike chambers. Hybrid figures and mechanical relics inhabit the architecture, fusing the organic and the artificial. The atmosphere is both sacred and grotesque, evoking a visionary world of allegory and chaos where architecture becomes a living, moral landscape".* The results are reported in Figure 4 and 5. Overall, no model was able to replicate the required style nor compositionally nor chromatically and the addition of depth conditioning further worsened the issue. This is probably due to the fact that the prompt is explicitly requiring a subject and a composition which is not represented in Bosch corpus. It is noteworthy that using the prompt *"Hieronymus Bosch painting and compositional style"* only the Stable Diffusion models generated outputs approximating Bosch's style, while only SDXL consistently recreated its characteristic colour palette.

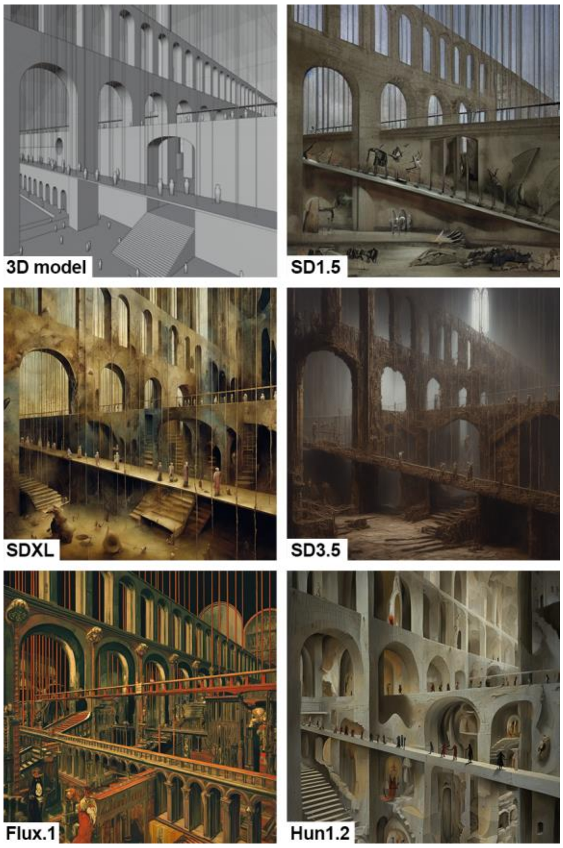

Figure 5: *Images generated with "Bosch" prompt and depth ControlNet. The first picture reports the reference 3D model.*

2.2 GENERALISED STYLE PROMPTS

To test less restrictive and broader stylistic conditioning a set of three prompts were generated using ChatGPT-5 providing the previous two prompts as a reference and asking for a renaissance, a contemporary with curtain walls and vegetation, and a cyberpunk variation. As a further restriction it was asked not to include names of architects or illustrators in the prompts. The resulting prompts and the corresponding images are reported below and in Figure 6 respectively:

A grand vaulted hall with staircases winding between columns and arches, connected by bridges beneath a coffered ceiling. Sunlight reveals proportion and harmony across the stone geometry. The composition evokes mathematical order and perspective, a spatial manifestation of Renaissance ideals where structure and ornament coexist in perfect equilibrium;

A vast atrium with staircases, suspended walkways, and open terraces surrounded by glass curtain walls and vegetation. Sunlight filters through transparent façades, reflecting greenery and metal. The composition blends precision and organic growth, expressing contemporary ideals of sustainability, openness, and integration between architecture, light, and nature;

A sprawling futuristic cyberpunk cityscape of layered bridges, vertical staircases, and dense infrastructure illuminated by neon and fog. Towers and platforms overlap in endless depth, merging technology and chaos. The air shimmers with reflections and holographic light. A high-detail composition where architecture becomes both machinery and metropolis, vast and luminous.

In the first two cases (i.e. "renaissance" and contemporary style) the quality of the composition, in terms of architectural plausibility and photorealism increases with the AI models evolution however a significant consistency was observed among all the models, except SD1.5 which showed a significant variation. As in the previous cases, central perspective was overrepresented; moreover, blooming and volumetric effects are prominent in the most recent models and, especially, in Flux, up to decreasing the photorealism of the composition.

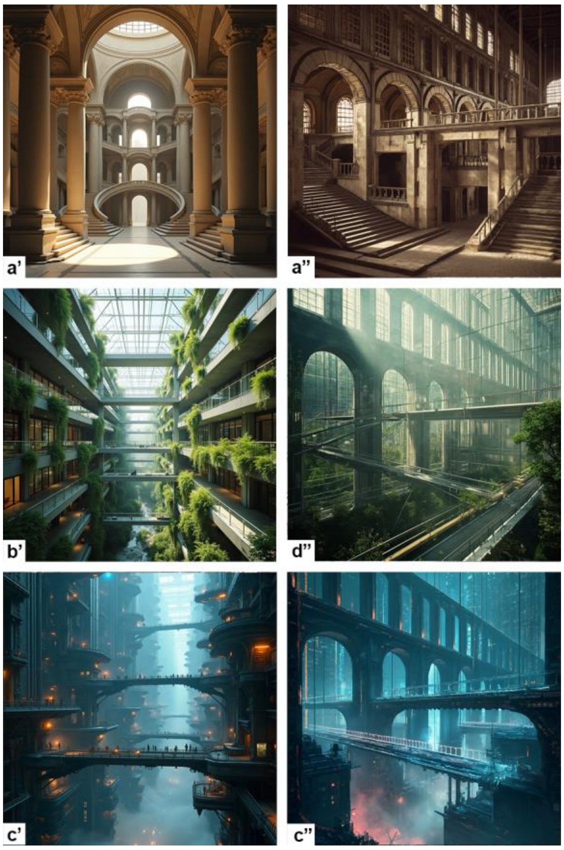

Figure 6: *Images generated with Flux and generalised prompts without and with depth ControlNet: a) "renaissance", b) contemporary and c) cyberpunk styles.*

In the case of the cyberpunk inspired prompt, however, significant stylistic differences emerged among the models tested. While in the two previous cases it was not possible to establish a likely stylistic reference or bias for each model and the results were mostly homogeneous in colours and similar in composition, Figure 7 shows how the interpretation of the prompt changed significantly in this case. This could be attributed to the different authors and styles most represented in the training datasets and, possibly, to the presence of licensed images in some datasets, which in this case could have a significant impact in the training. This results, for instance, in reminiscences of the drawings of Syd Mead and '80 oriented style in SD3.5 results or in the more modern and videogame-rendition of Hunyuan. Lastly, the similarities between the DALL·E composition in this and in Piranesi case (Figure 2) is surprising and notable and was confirmed providing the different prompts on different account, to avoid any possible interference of previous image generations from ChatGPT memory system.

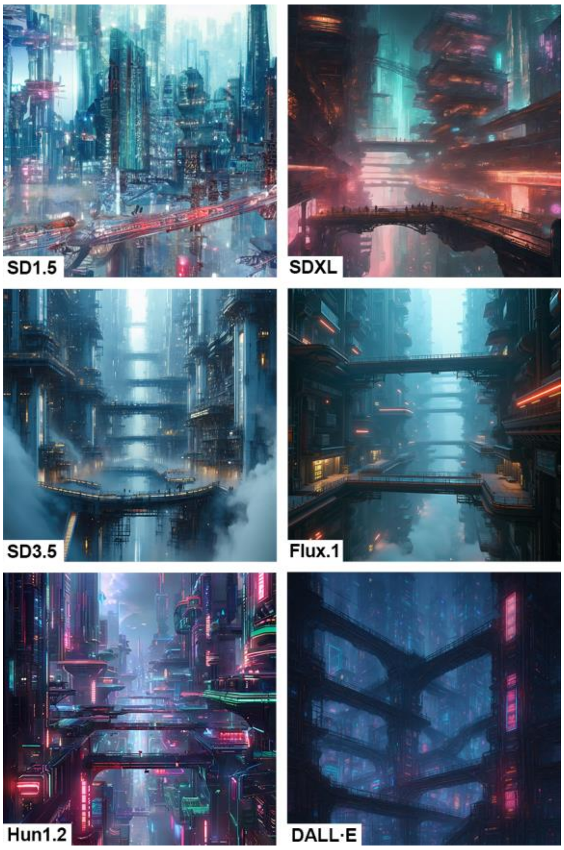

Figure 7: *Images generated with the cyberpunk prompt (c) highlighting different stylistic references.*

3. DISCUSSION AND CONCLUSIONS

The proposed workflow allows geometry-guided image generation connecting Blender and diffusion models using depth maps and ControlNet in ComfyUI. It proved to be capable of producing architectural visions coherent with the input compositions yet stylistically limited by the model adopted. It represents a technical framework for controlled non-proprietary image generation which could be used for rendering in early design stages. Furthermore, it could be extended with the use of LoRAs or specifically trained models to enhance the adherence to desired styles and purposes. Figure 8 represents a possible application in this regard which is generated with a highly specialized model tailored for modern architecture only.

Figure 8: *Example of Ai rendering generated with ArchitectureRealMix v11 [7].*

However, while this paper only presents a preliminary and qualitative analysis, it suggests that the use of general-purpose models to generate a specific style could introduce biases in composition, style, consistency and overall atmosphere. While the composition can be guided to obtain renderings of the desired architecture or scene, the qualities of the final rendering are model-dependent and less controllable. Moreover, they are deeply influenced by the curators of the training databases that directly shaped the visual quality of their models.

The comparisons between the pictures generated without depth control showed a significant bias toward recurring compositions like central perspective across all models, and specific biases like the prevalence of larger compositions and small details. SD3.5 and especially Flux, showed a significant recurrence of bloom, volumetric effects and high contrast rendering which convey a more "cinematic" look, but also cause a certain homogenization of the images generated which results in a short of model-specific visual quality. This is far less noticeable in SD1.5, which showed higher variability (or less prompt adhesion) and partially in SDXL.

Such observations suggest a broader reflection on AI's role in shaping our visual culture. Thought capable of producing new images and sometimes to successfully mix different inputs to create genuinely new content, AI can also shows significant signs of stereotyping which could lead to the overrepresentation of stylistic and visual tropes, the marginalization of underrepresented visual languages and the consequent narrowing of architectural imagination. The issue explored in this study, and the consequent envisioned threats, is that from the possibility of the *mechanical reproduction* that actually disseminated diverse visual culture, AI could cause a shift toward a *mechanical production* of images that nevertheless remain confined within the limits of their inherited data and risk to educate the public and the designer to an aesthetic homogenization which is, in fact, a "huge average" of a selected visual corpus adopted as training set.

Lastly, it should be noted that through careful prompt design, parameter tuning, and iterative generation, it is indeed possible to improve adherence to a desired style or compositional logic. Achieving this, however, often requires a high level of critical awareness and the ability to deconstruct a style into its constituent elements (i.e. lighting, details, texture, composition, perspectives, etc.) so that they can be explicitly articulated and incorporated into the prompt. Tools like ControlNet conditioning can assist in this process, but their effective use still depends on an informed, reflective engagement with the system.

This implies that while researchers and expert designers could be capable of leveraging such strategies to overcome some of the biases inherent in generative models, the majority of users (i.e. students, novices, or general practitioners) remain vulnerable to reproducing these limitations uncritically. Consequently, since all categories of users should engage critically with these technologies, balancing technical workflow strategies with reflective practice, there is a clear need for education and dissemination on the capabilities and constraints of AI tools, promoting awareness of how training datasets, model biases, and algorithmic tendencies shape the aesthetic outcomes of generated imagery.

4. REFERENCES

[1] Bianchini, Federico: Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12-15 Hine 2023, pp. 1493-1504

[2] Motoki, Fabio: More human than human: measuring ChatGPT political bias. *Public Choice*, vol. 198, pp. 3-23, August 2023.

[3] Piranesi, Giovan Battista: *Le Carceri d'Invenzione*, Plate 7, Rome, 1762. Accessible on Wikimedia Common at https://commons.wikimedia.org/wiki/File:Giovanni_Battista_Piranesi_-_Le_Carceri_d%27Invenzione_-_Second_Edition_-_1761_-_07_-_The_Drawbridge.jpg (accessed 29 September 2025)

[4] Bosch, Hieronymus: *The Garden of Earthly Delights*, Oil on wood, 1490-1510. Museo del Prado, Madrid. Accessibile on Wikimedia Common at https://en.wikipedia.org/wiki/The_Garden_of_Earthly_Delights#/media/File:The_Garden_of_earthly_delights.jpg (accessed 29 September 2025)

[5] Zhang, Lvmin: Adding conditional control to text-to-image diffusion models. Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 3836-3847.

[6] Rapp, Joanna Barbara: A geometrical analysis of multiple viewpoint perspective in the work of Giovanni Battista Piranesi: an application of geometric restitution of perspective. The Journal of Architecture, vol. 13, no. 6, pp. 701-736, 12/2008.

[7] Laizhende, (2024): ArchitectureRealMix [online], Available from: https://civitai.com/models/84958/architecturerealmix (accessed 4 August 2025).

AI Shakespeare: Mitigating Hallucinations in Co-Creative Animation

Hannes Rall, Alice Osinska and Aaron Zhi Qiang Lim

School of Art, Design and Media, Nanyang Technological University, Singapore,
Rall@ntu.edu.sg, alicja.osinska@ntu.edu.sg, aaron.limzq@ntu.edu.sg

ABSTRACT: Hallucination has long been a subject of research and scholarly debate. Based on an AI animation project inspired by Shakespeare, this paper adopts the concept of hallucination and frames it within an artistic context, specifically animation. Through an analysis of how hallucination has been defined in studies on AI and how it can be contextualised within animation, the paper identifies three major issues encountered during the animation process: stylistic inconsistency, narrative misalignment and motion disruption. The case study examines the style of the historical artist Eugène Delacroix, transforming his static illustrations into animated format. The process revealed several challenges arising from the incorporation of AI, which were mitigated through the use of reference images, hybrid storyboarding combining sketches and AI, and manual artistic intervention within the AI-generated animation. Framing the concept of hallucination within this field prompts a discussion on what is actually required from AI in artistic practice.

1. INTRODUCTION

AI advancements have shed light on the ethical and technical aspects of creative practice since the launch of Midjourney and ChatGPT in 2022. Since then, AI models have produced visually appealing outputs full of realism, often becoming difficult for humans to distinguish [1]. However, AI's hyperreality does not come without constraints, as its artistic variability remains limited. The shortcomings include misalignments, copyright issues, and unexpected generations, all of which have become subjects of public debate. As a result, the term hallucination has gained prominence, originating from the idea of something existing only in one's mind but not in the real world, and now describing AI's current inability to discern what is real or intended, existing solely within the state of AI itself.

The randomness of AI generation is largely dictated by an insufficient or faulty training process, which remains inaccessible for human intervention. Therefore, when creating with AI, detecting and mitigating hallucination becomes one of the essential tasks of the human author. The question is no longer whether AI should be used in the creative sector, but rather how it should be used.
While mitigating distortions in AI output has been addressed by some scholars [2], relatively little research has examined this issue in the context of artistic creation, particularly within animation practice. The overarching aim of this study is to extend the discussion on conceptualising hallucination beyond existing definitions such as: *'non-existent objects are erroneously detected or incorrectly localised at their anticipated positions'* [3], *'generative AI software systems generate fabricated or false information'* [4], and '*patterns or objects that are non-existent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate'* [5], and to move this discourse towards the domain of creative practice. Moreover, there may be more to hallucination than its pejorative connotation. Some authors interpret distorted output as a trigger for creativity and storytelling, emphasising its positive associations [6][7][8].

Our interest, however, lies in defining hallucination within the context of animation, with the aim of mitigating such outputs. For this reason, the paper investigates how AI can support the adaptation process within ethical constraints and how hallucination can be controlled in artistic practice to achieve specific outcomes. Following this statement, *how do we perceive hallucination in the animation field? And how can such a situation be mitigated by co-creative practices?*

In an ongoing Shakespeare animation project, we employ AI to translate classical illustrations into motion, demonstrating both the indispensable role of human direction and the potential of generative AI through a co-creative, sketch-based approach. This practice-based study not only addresses the question of how, but also provide a framework for animators working in the AI era. Consequently, through this case study, we seek to elucidate how GenAI can be integrated into creative practice without compromising artistic integrity.

2. DEFINING HALLUCINATION

The origins of the term ‚hallucination‘ refer to a person who constructs imaginative perceptions [9]. However, the emergence of AI has redefined the concept within a new framework. Terms such as ‘distortion‘, ‘AI fabrication‘, or the more commonly used ‘AI hallucination’ describe the information that is missing within images and is filled with random data by AI [10]. While the term initially arose from falsified information produced by Large Language Models (LLMs), it has recently gained more attention in broader studies of AGI, where it is conceptualised as *‘model outputs that do not align with the contemporary empirical realities of our current world’* and is categorised into various types depending on the field of study [11].

Some scholars interpret AI hallucination in a more explicit and straightforward manner, such as *‘instances where non-existent objects are erroneously detected or incorrectly localised at their anticipated positions’* [3], while others emphasise the term’s fluid nature, stating that *‘hallucination includes not only misalignment between text prompts and generated images but also the generation of factually defective images‘* [12]. Moreover, some definitions take a more less straightforward and direct approach, explaining that *‘hallucinations may seem to be plausible while they are in fact unacceptable’* [13], which may directly relate to misalignment at the artistic level.

Therefore, a creator with a specific concept in mind works toward a defined goal, yet within the AI setting, even when the output does not significantly distort the form or narrative, it may still diverge from the creator’s intended outcome. As a result, this paper seeks to take the discussion one step further by exploring the connotations of AI distortions, that is, AI hallucination in the context of animation.

2.1 HALLUCINATION IN ANIMATION

Originating from human experience, hallucination brings forward the relationship between human imperfection and AI hallucination, as humans also tend to falsify truths and manipulate content, often described as ‘subjectivity’, ‘bias’, or ‘stereotyping’ [14]. This behavioural pattern, observed in both humans and AI, suggests that continuous reflection and improvement are necessary, analogous to the supervision and art direction required for AI-generated output.

The issues at the intersection of animation and AI can therefore be summarised within the framework of hallucination. As demonstrated, hallucination does not only feature implausible situations but is also associated with unacceptable outputs. In some cases, it can be related to ‘caption hallucination’, described by as *‘producing descriptions or captions that either contain elements inconsistent with the image or omit important aspects of it’*. [2] Our case study encounters precisely this issue, as even with image-based generation, AI can still produce incorrect or unacceptable content. While the term hallucination has become well established in describing AI-induced distortions, its meaning and potential relevance to visual practice remain underexplored.

Following existing definitions, we divide hallucination into three categories which we encountered during the animation process:

- Stylistic inconsistency – when a model diverges from the given stylistic instruction
- Narrative misalignment – when AI misunderstands the atmosphere and reshapes the narrative
- Motion disruption – when behavioural aspects misalign with the rules of animation, whether in gestures, storyboarding, or scene timing

These three types of hallucination are examined through the project of readapting Shakespeare from static historical works into animation. The case study is grounded in the lithographs of Eugène Delacroix, the renowned nineteenth-century artist who illustrated Shakespeare’s play Hamlet in the lithographic medium.

Figure 1: *The examples of the famous Shakespeare illustrations by Eugène Delacroix serving as a base for animation readaptation.*

3. STRATEGIES TO MITIGATE THE HALLUCINATION IN ANIMATION

To begin with, it is worth introducing the reasons for hallucination encountered when utilising AI. An animator working with open-access models has limited control over the generation process, including stylistic consistency and the subtle gestures of characters. In a classical setting, the animator constructs the environment and, through the use of animation software, guides the narrative. This changes entirely in the context of AI. For instance, when animating with a single model, it may be difficult to achieve a specific outcome, as publicly available models differ significantly in their strengths and weaknesses. Moreover, the challenges vary depending on the stage at which AI is incorporated. During the ideation phase, ethical concerns arise since AI, trained primarily on internet-sourced data, can probe one's visual development. Additionally, the photorealistic bias in training data often fails to meet the expressive needs of animation. On the other hand, even with a well-prepared base input, hallucination may occur once the image is set in motion by AI. The project of animating Shakespeare with AI confronts precisely these challenges, including stylistic misalignment as well as narrative and motion distortion.

3.1 HISTORICAL REFERENCE

Shakespeare's plays have been adapted across diverse mediums, including animation and illustration. Although the practice of animating still images has been explored and is often viewed as an act of creative reinterpretation [15], its implications in relation to AI-induced distortions remain insufficiently examined. To align with principles of ethical AI use, this project adopts the strategy of working with public domain materials as a foundational source. By employing a model for readaptation rather than relying solely on AI-generated outcomes derived from training data, the approach mitigates risks of copyright infringement. Nevertheless, this process presents notable challenges, as stylistic distortion frequently alters the intended aesthetic once the image is animated.

Figure 2: *Comparative example demonstrating stylistic hallucination and its mitigation through prompt intervention. The original input (left) produces a hallucinated output (centre), which is subsequently corrected through prompt-based intervention (right).*

In this context, hallucination can be understood as the tendency of AI models to transform a distinct visual style into a three-dimensional aesthetic. This phenomenon corresponds to a form of 'stylistic inconsistency', in which the model disregards the provided stylistic parameters. Unlike a human animator who, when following a specific instruction, can refine and adjust decisions through critical judgment, AI lacks the capability for aesthetic evaluation inherent in a human artist. This absence underscores the indispensable role of the animator within AI-assisted workflows. In such cases, hallucination does not manifest through the creation of 'non-existent', 'nonsensical', or 'inaccurate' objects but rather through the deviation of the output from the intended stylistics toward a more realistic, film-like mode of representation. This form of stylistic inconsistency can be regarded as a deliberate manifestation of hallucination resulting from the biases embedded in training data. Within artistic practice, therefore, we define this occurrence as a type of hallucination in which AI modifies given instructions under the influence of its data-driven predispositions. Consequently, how to mitigate such a situation?

'Prompts' refer to the textual inputs provided by users. Their modification is conceptualised as an 'engineering' process. With the increasing availability of advanced and more accurate models for creative professionals, prompt engineering has become a highly valued skill, ensuring more precise outputs [16] and serving as one of the key methods for mitigating potential misalignments.

When text-to-image models such as Midjourney and DALL·E were first introduced, they astonished the creative community; however, their capabilities in generating specific styles at that time remained limited. Merely two years later, generative models now differ significantly in their training techniques and offer a greater variety of stylistic options. Therefore, beyond prompt engineering, knowledge of new models and their potential applications is essential for maintaining a smooth AI workflow. A relevant example is the growing practice of combining Depth and Canny within ControlNet. While Canny primarily focuses on edge detection, Depth preserves the spatial structure of the image. Through a combination of careful prompt design and the use of Canny, two-dimensional stylistics tend to produce more consistent results while avoiding unwanted realism.

One further aspect that must be emphasised is the role of source input as a base for readaptation. Minimal or absent human intervention is often criticised by practitioners as a form of AI theft [17]. This issue becomes particularly evident when visual outputs are generated solely from textual prompts, especially when these outputs attempt to represent recognisable characters or established narratives. In such cases, hallucination not only pertains to stylistic distortion or narrative misalignment but may also extend to ethical concerns of artistic appropriation. To address this, our case study employed public domain illustrations as the foundational inspiratoin base for creative readaptation.

3.2 HYBRID APPROACH

Online AI-generated videos are often recognised for their close resemblance to reality, which diverges significantly from what animators usually aim to achieve. But what happens when an animator works with limited source material? This particular situation arises as the original selected as an inspiration artist did not create illustrations for all scenes.

In *Act 4* of Shakespeare's *Hamlet*, the scene in which Hamlet is sent away to England is visualised through the animation of a sailing ship. Instead of generating an image directly from a textual prompt, the process first included a traditional hand-drawn sketch. This rough sketch served as a base, while a trained LoRA (Low-Rank Adapter), developed using all available lithographs by Eugène Delacroix, was employed to generate the appropriate style. From this image ideation process, another aspect emerges: filling the undefined space of a sketch may potentially enrich the overall storyboard process. Of course, it remains the artist's role to utilise such outputs and evaluate their aesthetic quality. This experimental observation suggests that hallucination, despite its negative connotations, may also carry creative potential.

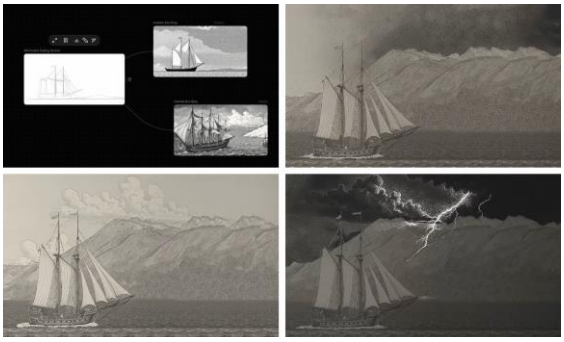

Figure 3: *Stages of visual development based on sketch generation with Canny.*

The situation changes, however, when generating images that depict characters or capture the specific atmosphere of a narrative. When provided with a sketch, AI often reshapes the narrative by adding additional characters, altering facial expressions, modifying age, or changing the overall tone of the scene. This example illustrates '*narrative misalignment'*, which can be framed within the concept of hallucination. In a traditional setting, an animator maintains consistency of character and scene, whereas AI may misalign previously established visual content.

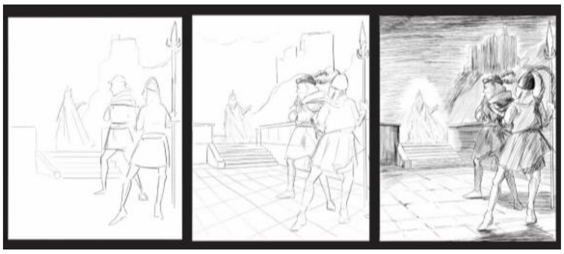

Figure 4: *Sketches used for AI generation, ranging from minimal detail (left) to complete drawing (right).*

Through the sketch-based experiment, we observed that the most effective AI style generation occurred when sketches contained only basic outlines, as fully detailed drawings with shading resulted in 'over-information' for the model. For this reason, further scenes development involved minimal linear drawing.

Figure 5: *Example of scene generation based on a sketch input, illustrating hallucination of environmental and character inconsistency*

Nevertheless, the inclusion of a sketch does not guarantee full consistency. On multiple occasions, AI generations based on sketch input produced various hallucinated outputs, including inconsistencies in style, character appearance, environmental setting, and even the number of characters within a scene. While in experimental practice such distortions might result in 'happy accidents', these evident flaws cannot persist in narrative-based animation, where they risk disrupting audience attention and undermining the coherence of the plot. For this reason, manual artistic intervention remains crucial in mitigating such occurrences, as explained in the following section.

3.3 CONTROLLING MISALIGMENT BY CO-CREATION

As outlined above, the sketch-based workflow does not ultimately guarantee successful output. Therefore, another strategy involves adjusting the scenes manually through the use of traditional animation software. The co-creation between fully automated AI processes and traditional tools not only enables greater control over the overall aesthetics but also mitigates hallucinated outputs. By combining AI-generated elements with manual storyboarding and animation techniques, the process remains within an ethical framework.

One significant case of misalignment was encountered in Act 2, where the character of Polonius sends a spy. While animation does not necessarily aim for complete realism in character movement, a professional animator is capable of establishing a baseline for cohesive motion. In this regard, AI does not suffice, as hallucination becomes evident in perspective shifts and motion timing.

Figure 6: *Example of motion hallucination persisting through multiple iterations, marked by inconsistencies in running speed, behavioural narration, and gesture alignment*

The chosen solution for this issue was artistic intervention, which did not eliminate AI input but rather integrated it with manual techniques. The intervention primarily addressed changes in perspective and running motion. In summary, the sketch base and running sequence were produced through manual animation, while the LoRA model generated the visual style of the base image and the secondary character animation. This process effectively resolved the motion misalignment that had originally appeared as hallucination in the form of distorted running speed and perspective.

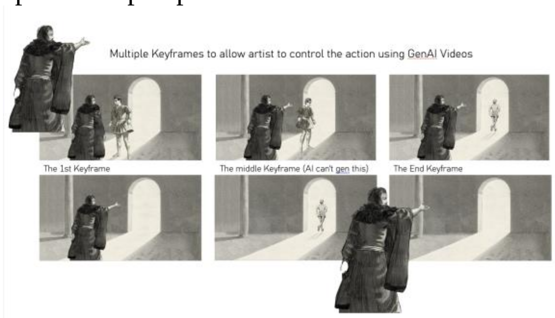

Figure 7: *Process of mitigating hallucination through a co-creative approach, where the overall scene remains AI-generated while motion is refined using manual animation techniques*

4. CONCLUSION

From the presented case study, three major conclusions can be drawn. First, the definition of AI hallucination has not yet been firmly established within the artistic context. What may be considered plausible in photorealistic image generation is not perceived in the same way within the arts, and particularly not within animation. By framing the concept of AI hallucination around three categories, this paper aimed to elucidate the notion of AI misalignment.

Second, AI-assisted animation requires continuous supervision of both its aesthetic and narrative outputs. Although the development of new models in recent years has contributed to a reduction in visual and motion distortions, these improvements do not fully meet the needs of artistic practice. Creators often move away from realism, favouring more experimental and aesthetically engaging approaches.

Lastly, the strategy of using historical references within the public domain, alongside a hybrid approach that integrates hand-drawn methods, AI generation, and manual intervention, ensures that AI animation remains within an ethical framework.

5. ACKNOWLEDGMENT

This research is supported by the Ministry of Education, Singapore, under the AcRF T2 grant Ethical Approaches to AI-Visualization in Shakespeare Adaptation (MOE-T2EP40223-0004).

6. REFERENCES

[1] Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., ... & Seli, P. Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive research: principles and implications*, 8(1), 42, 2023.

[2] Rani, A., Rawte, V., Sharma, H., Anand, N., Rajbangshi, K., Sheth, A., & Das, A. Visual hallucination: Definition, quantification, and prescriptive remediations. arXiv preprint arXiv:2403.17306, 2024.

[3] N. Maleki, B. Padmanabhan and K. Dutta, "AI Hallucinations: A Misnomer Worth Clarifying," 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, Singapore, pp. 133-138, 2024, doi: 10.1109/CAI59869.2024.00033.

[4] Christensen, J., Hansen, J. M., & Wilson, P. Understanding the role and impact of Generative Artificial Intelligence (AI) hallucination within consumers' tourism decision-making processes. *Current Issues in Tourism*, 28(4), 545–560, 2024. https://doi.org/10.1080/13683500.2023.2300032

[5] IBM. (n.d.). What are AI hallucinations? Retrieved September 21, 2025, from https://www.ibm.com/think/topics/ai-hallucinations

[6] Halperin, B., & Lukin, S. (2024). Artificial Dreams: Surreal Visual Storytelling as Inquiry Into AI 'Hallucination'. Proceedings of the 2024 ACM Designing Interactive Systems Conference, Denmark. https://doi.org/10.1145/3643834.3660685.

[7] Tonmoy, S., Zaman, S., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models, (2024), *ArXiv*, abs/2401.01313. https://doi.org/10.48550/arxiv.2401.01313.

[8] Jančařík, A., & Dusek, O. (2024). The Problem of AI Hallucination and How to Solve It. European Conference on e-Learning, Portugal. https://doi.org/10.34190/ecel.23.1.2584.

[9] El-Mallakh RS, Walker KL. Hallucinations, psuedohallucinations, and parahallucinations. *Psychiatry*. Spring;73(1):34-42, 2010. doi: 10.1521/psyc.2010.73.1.34. PMID: 20235616.

[10] Sun, Y., Sheng, D., Zhou, Z. et al. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanit Soc Sci Commun* 11, 1278 , 2024. https://doi.org/10.1057/s41599-024-03811-x

[11] Wang F. Lighthouse: A survey of agi hallucination. arXiv preprint arXiv:2401.06792. 2024 Jan 8.

[12] Lim, Y., & Shim, H. Addressing image hallucination in text-to-image generation through factual image retrieval, 2024. arXiv preprint arXiv:2407.10683.

[13] Neeloy Chakraborty, Melkior Ornik, and Katherine Driggs-Campbell. Hallucination Detection in Foundation Models for Decision-Making: A Flexible Definition and Review of the State of the Art. *ACM Comput. Surv.* 57, 7, Article 188 (March 2025), 35 pages. https://doi.org/10.1145/3716846

[14] Tlili, A., & Burgos, D. AI Hallucinations? What About Human Hallucination?! Addressing Human Imperfection Is Needed for an Ethical AI. *Int. J. Interact. Multim. Artif. Intell.*, 9, 68, March, 2025. https://doi.org/10.9781/ijimai.2025.02.010.

[15] Batra, K., Kavidayal, M.: Animating Still Images. Sept. 2022, arXiv preprint arXiv:2209.10497v2 , 2023.

[16] Federiakin, D., Molerov, D., Zlatkin-Troitschanskaia, O., & Maur, A. Prompt engineering as a new 21st century skill. *Frontiers in Education*. November, 2024, https://doi.org/10.3389/feduc.2024.1366434.

[17] Goetze, T. S. AI art is theft: Labour, extraction, and exploitation: Or, on the dangers of stochastic Pollocks. June, 2024. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, USA (pp. 186-196).

SESSION II

“Immersive Virtuality”

Moderation: Prof. Dr. Andreas Bienert
(form. Staatliche Museen zu Berlin)

Virtual Tour for Heritage Accessibility: The Case of Palazzo Carignano Museum

Roberta Spallone[a], Michele Russo[b], Marco Vitali[a], Fabrizio Natta[a], Enrico Pupi[e], Martina Casciola[b], and Martina Rinascimento[a]

[a]Department of Architecture and Design, Politecnico di Torino, Italy, roberta.spallone@polito.it, marco.vitali@polito.it, fabrizio.natta@polito.it, enrico.pupi@polito.it, martina.rinascimento@polito.it
[b]Department of History, Representation and Restoration of Architecture, Sapienza Università di Roma, Italy, m.russo@uniroma1.it, martina.casciola@uniroma1.it

ABSTRACT: The digital transformation in the cultural heritage sector has led to the widespread adoption of Virtual Tours (VTs), which are intended as interactive simulations of real-world locations that allow users to explore and navigate spaces remotely. This trend was significantly accelerated by the global pandemic, which compelled cultural institutions to reassess their accessibility models and has produced a diverse landscape of technological solutions, revealing a critical need to move beyond simple digital replicas toward more integrated and meaningful interactions. In this paper, VT has been created to offer the public new visualizations of rooms and panoramas that are partially or fully inaccessible inside the Palazzo Carignano Museum. Therefore, VT faced the central goal of reducing obstacles, inequalities, and gaps that limit citizens' participation in cultural life and heritage, favoring accessibility to rooms that are not accessible by the mobility-impaired or not visitable by the public due to safety reasons. A sequence of spherical images linked together by graphic elements that allow the photographs to be visited and scrolled makes up the VT. It will be available on the museum's website and usable on-site via totems and handheld devices.

1. INTRODUCTION

This proposal is rooted in a broader project aimed at enhancing the physical and cognitive accessibility of the Carignano Palace and Museum, which was commissioned by the Direzione Regionale Musei (DRM) - Museo di Palazzo Carignano. The project has been funded as part of the PNRR (National Recovery and Resilience Plan) in the Accessibility sector and concluded in February 2025.

The overall project is divided into two activities: analysis, conducted through an integrated digital survey, and interpretive modeling, of the museum and the XVII-century façades, performed using the H-BIM process. At the same time, communication strategies, methodologies, and tools ranged from reconstructive digital modeling and physical modeling to virtual tours (VT), virtual reality (VR), and augmented reality (AR). Within the scope of intervention, museum spaces, missing structures, and architectural elements are presented using certain above-mentioned techniques, selected on a case-by-case basis according to the project objectives. The phygital products will be integrated into the new installations and devices envisioned in the overall communication plan, as well as the new exhibit for the Palazzo Carignano Museum.

2. MAIN OBJECTIVES AND PROJECT OVERVIEW

VT faced the central goal of the project, which was to reduce obstacles, inequalities, and gaps that limit citizens' participation in cultural life and heritage, thereby favoring accessibility to rooms that are not accessible to the mobility-impaired or not visitable by the public due to safety reasons.

Another aim is to expand the number of visitors, not only by promoting accessibility, but also by communicating to tourists planning a visit to the city the quality of the Museum's interior spaces. Indeed, Palazzo Carignano, designed and built by Guarino Guarini from 1679, is an acknowledged masterpiece of Baroque architecture. The building was originally the residence of the cadet branch of the House of Savoy-Carignano. In 1848, the main hall was transformed into the chamber of the First Subalpine Parliament, while one room housed Camillo Cavour's study. The museum tour offers visitors the chance to admire 17th-century rooms decorated with frescoes, wood paneling, stucco work, and mirrors, and furnished with original pieces of furniture.

However, the palace is mostly explored from the outside, or by walking through its entrance hall, atrium, and courtyard, which are freely open to the public and connect two monumental squares, Piazza Carignano and Piazza Carlo Alberto.

Few are familiar with its vestibules, double staircase, and the precious apartments of Mezzanotte and Mezzogiorno, which housed the Princes of Carignano and are now home to the Museum of Palazzo Carignano.

Moreover, some rooms and open spaces, such as the lantern that covers the main hall, and the terraces on the roof, are not accessible to the public.

For these reasons, a virtual tour was created.

Three main phases characterized the work: the design of the spherical camera shoots and its implementation, the processing of the photographs to create a continuous path, and the creation of the user interface, equipped with a navigator.

A sequence of spherical images linked together by graphic elements that allow the photographs to be visited and scrolled makes up the VT. It will be available on the museum's website and usable on-site via totems and handheld devices.

Particular attention was paid, right from the initial stages of shooting with the spherical camera and then in the construction of the paths between interior and exterior spaces. Numerous meetings of the research group were devoted to examining the different possibilities of movement, with the aim of providing the public not with a paratactic visualization of the individual rooms, but with a syntactic understanding of the architectural complex in its morphological and decorative values. Equally, attention was paid to the display options for the main station points, route directions, access from ground level to upper floors, and points of interest accompanied by multimedia information.

Following a thorough benchmarking of international cases, VT is characterized by:

- continuity of the path, based on the principle of the intervisibility of station points,
- careful construction of the user interface, which offers alternative ways of routing while avoiding redundancies or loss of orientation,
- possibility of inserting hotspots through which multimedia content can be accessed.

3. VIRTUAL TOUR AND MUSEUMS: AN OVERVIEW

The digital transformation of cultural heritage has led to a widespread adoption of Virtual Tours, a trend significantly accelerated by the global pandemic, which forced cultural institutions to rethink their accessibility models [1]. This evolution has produced a diverse landscape of technological solutions, revealing a critical need to move beyond simple digital replicas toward more integrated and meaningful interactions [2]. An overview of the current state of the art reveals key trends and challenges, particularly concerning the representation of museums not just as collections, but as significant architectural heritage in their own right.

A primary distinction in the museum VT landscape lies in the technological platform. Many institutions, like the Musei Capitolini (Rome) and the British Museum (London), rely on third-party platforms such as Google Arts & Culture. This ensures broad visibility and a standardized user interface but often limits customization. Conversely, institutions such as the Louvre (Paris), the Rijksmuseum (Amsterdam), and the Museo Egizio (Turin) have developed bespoke platforms. These solutions offer greater control over the user experience, allowing for sophisticated navigation and deeper content integration. A third category utilizes specialized software like PANO2VR or Matterport, as seen in Turin with the Museo di Arte Orientale and the Apartment of Mezzogiorno in Palazzo Carignano. These tools offer advanced features, including photogrammetry-based 3D models and VR compatibility, providing a balance between custom development and off-the-shelf solutions [3].

Beyond the platform, the user's journey is defined by the navigation model, which becomes critical when the focus is the architectural "container". Many VTs favor a guided, cinematic experience – as seen in tours of the Palace of Versailles – that prioritizes storytelling over free exploration. While narratively effective, this can hinder a user's grasp of spatial relationships. A frequent challenge in VTs is poor navigation design, which can lead to user disorientation [4]. In contrast, the new project for Palazzo Carignano emphasizes spatial continuity, a crucial element for allowing users to assess the whole spatiality of the place. Unlike tours that present disconnected rooms, this approach is built on the intervisibility between observation points, enabling a coherent perception of the palace's architectural sequence. This method transforms the VT from a mere gallery into a tool for architectural comprehension, effectively communicating the building's spatial dimensions and proportions [5]. This focus on the building itself represents a significant step forward, using technology not only to overcome physical barriers but to offer a deeper, more holistic understanding of cultural heritage [6].

4. VIRTUAL TOUR TECHNICAL SPECIFICATIONS

The Virtual Tour (VT) of Palazzo Carignano was developed by balancing and optimising the acquisition of spatial data with the use of data visualisation tools, achieving a specific communicative effectiveness.

As mentioned in paragraph 3, the acquisition was planned initially by identifying the main trajectories with respect to the areas to be covered (Fig. 1). This preliminary plan was validated in situ and integrated with respect to the space morphological complexity. The latter was decisive in determining the number of interesting points, aiming to preserve the continuity of the route while also adapting the level of information from a specific perspective. Complex spaces, therefore, required more viewpoints, thereby allowing for a more exhaustive exploration. However, attention was also paid to optimising the data, limiting the viewpoint redundancy. It would have slowed down the visit and made the tool more complex to manage. In open or huge spaces, acquisitions were planned at greater distances, allowing for rapid movement while preserving the observation of the context.

The photo acquisition was conducted using an INSTA360 X4 camera, capable of simultaneously capturing two hemispherical images that are automatically stitched together into a 360° image (Fig. 2). The camera is equipped with a 1/2" sensor, an f/1.9 aperture, a focal length equivalent to 6.7 mm (35 mm) and the ability to capture images up to 72 MP (11904×5952 px). In the Palazzo Carignano VR, the maximum resolution was set, considering the average acquisition distance, to ensure an adequate level of detail (Ground Sampling Distance) for immersive navigation.

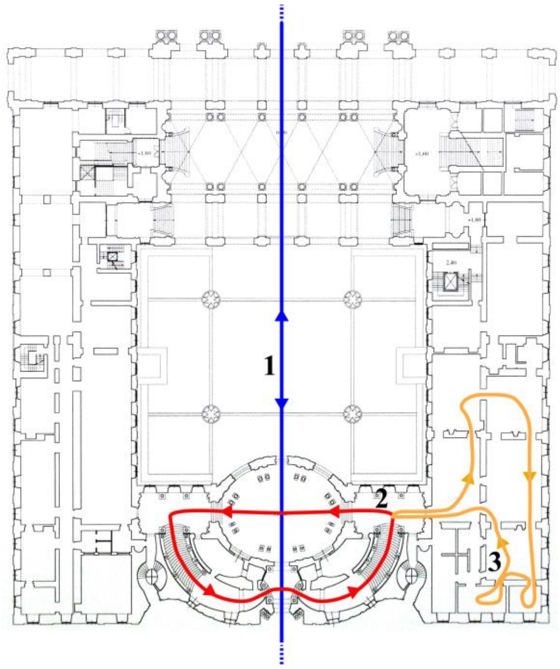

Figure 1: *Floor plan of the building with layout of some VT routes (Editing by M. Russo)*

A critical issue addressed in image acquisition was the variable lighting conditions. The need to define in advance set-ups suitable for specific conditions led to the identification of four main types of spaces: open (natural light only), semi-open (covered but with a predominance of natural light), closed with mixed light (natural and artificial) and closed (predominantly artificial light). Specific parameters for ISO, shutter speed and white balance were defined for each of these, ensuring colour uniformity between images. All 57 photographs were taken with the aid of a tripod.

5. VIRTUAL TOUR PROCESSING

The first step in post-production of the images was colour equalisation using Adobe Photoshop's Camera Raw plugin tools. It allows control of specific lighting issues while preserving the actual environmental conditions. The effects of different light sources, both natural and artificial, in interior spaces were mitigated, improving the overall readability and perceptual homogeneity of the spherical photographs.

The modified images (Fig. 2) were imported into the Virtual Tour PRO (3DVista) software, a platform that allows the user interface (UI) to be configured and customised (see paragraph 6) to adapt the virtual tour experience to different user needs. An initial assessment focused on simplifying and selecting the content to be included on the platform. As will be described in

detail in the following paragraph, the VT is the result of a delicate balance between the accessibility of the content, the clarity of the information, and the completeness of the data conveyed. In line with this process, the building's floor plans were simplified to provide VT *containers* that were immediately understandable in terms of positioning in the Palazzo system, highlighting only the spaces involved in the routes.
For the same reason, the number of points of interest has also been reduced in the key plan to avoid excessive overlapping of information (Fig. 3). Visitors, therefore, have the opportunity to explore new stations within the VT that are not included in the key plan but are helpful in the transition between stations.
This transition is supported by the inclusion of simplified icons, which must ensure recognisability while preserving graphic continuity. Some icons only appear when the mouse hovers over them, optimising the image visualisation.
The last development process focused on the portability of the VT. The interface was designed based on a Full HD (1920×1080) reference scheme.

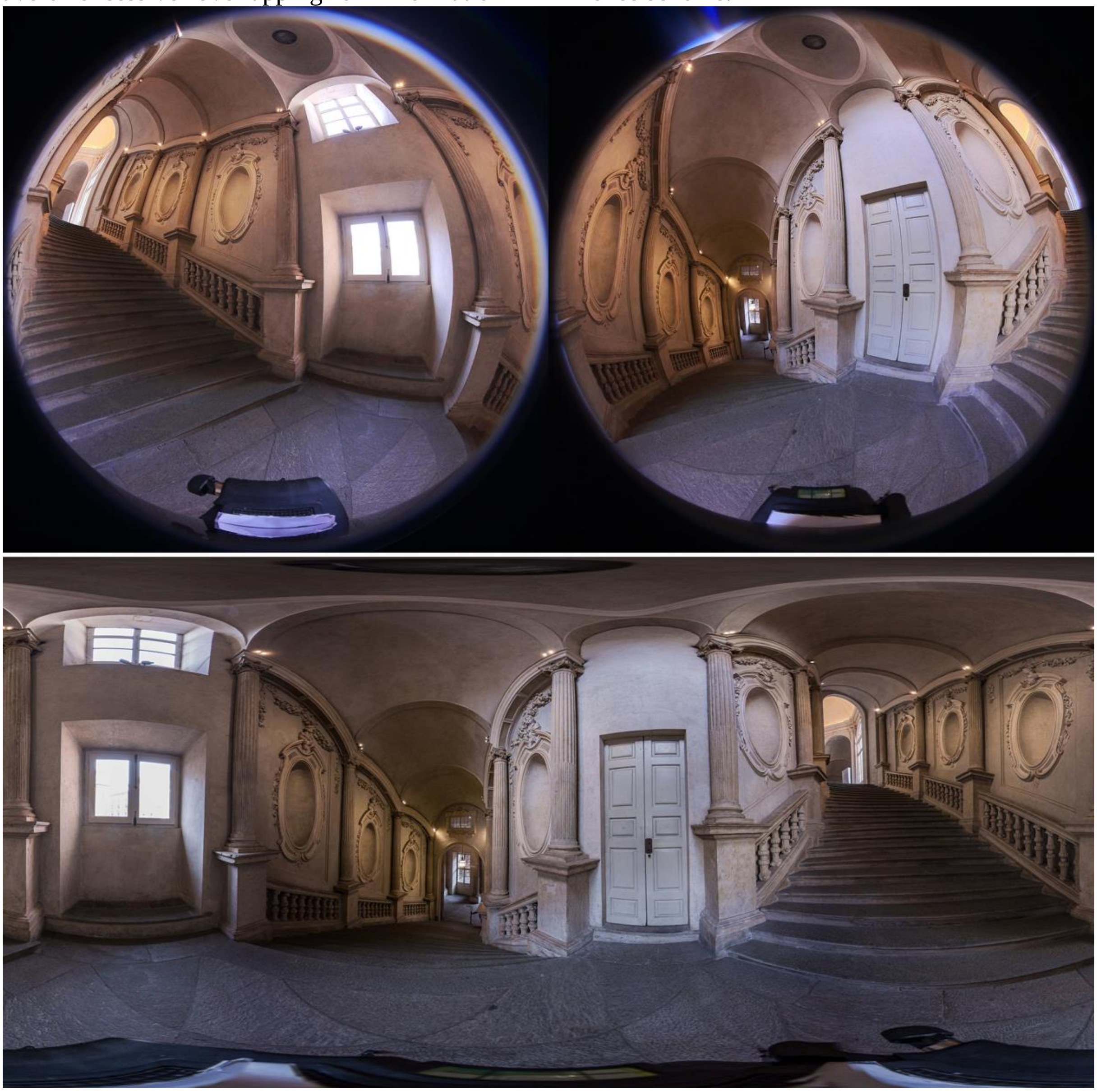

Figure 2: *Images from the two hemispherical photos and edited equirectangular image (Editing by M. Russo)*

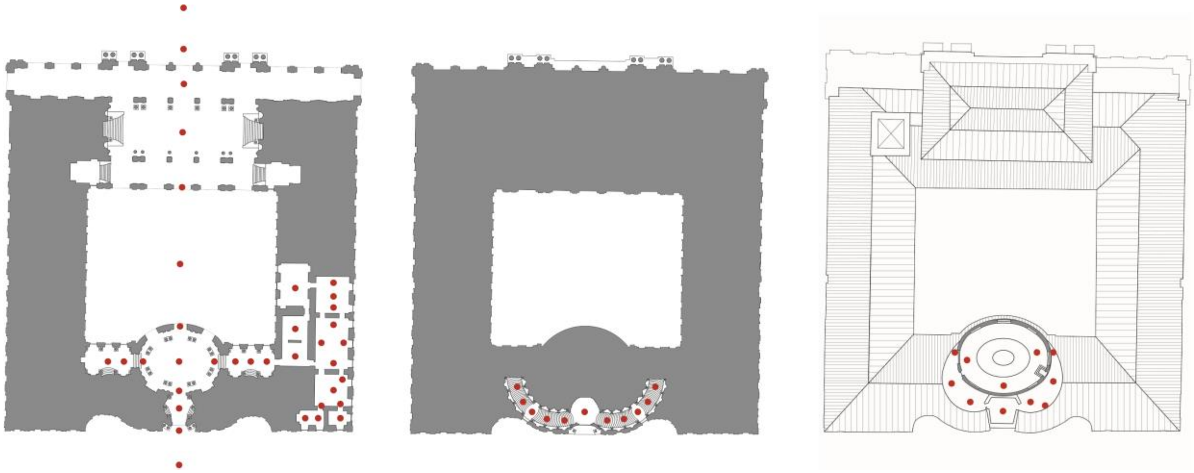

Figure 3: *Plans (ground floor, first floor, top floor) with the distribution of VT viewpoints (Editing by M. Russo)*

Figure 4: *Wireframe of VT interface. (Editing by M. Rinascimento)*

It is optimal for viewing on computers but can also be re-adapted to tablets and smartphones without modifying the original images. In these cases, the UI elements automatically adjust to the screen size, preserving both accessibility and interactivity of the experience. These elements can be further calibrated during the final implementation phase, once the target platform has been identified.

6. VIRTUAL TOUR USER INTERFACE

In order to make the virtual tour output as accessible and easily navigable as possible for the Museum broader target audience, some UX Design methodologies and good practices were followed so as to additionally guarantee the achievement of a User Centred output.

First, simple B/W visual schemes, technically referred to as wireframes, were sketched to represent a draft of the appearance of the digital interface of the virtual tour to-be (Fig. 4).

In this phase, it was crucial to convert the insights collected during the VT's benchmarking analysis, discussed in paragraph 3 of this contribution, in early design choices, namely spotting the elements and components of interest standardised for a virtual tour, like an interactive plan of the explorative area and control buttons for movement, and position them in the VT interface. This methodology helped the research and design team to imagine the digital interface layout, the displacement of content and ensure that the flux of user interactions with the digital system designed is intuitive and smooth. Coherently, the visual appearance of user experience components isn't relevant at this point, indeed placeholders recalling the shape and space occupied by interactive elements are placed. Following the design of the layout, a set of icons and visual components were designed and positioned in place of the correspondent placeholders (Fig. 5). Ensuring suitable movements dynamics resulted being essential to offer a pleasing and customizable exploration of the VT. Reasonably, the research team decided to design varied modalities to move in the digital space as both a set of movements and exploration icons and an interactive plan of the Museum building. The former was placed at the bottom of the interface in order to support the user to move across spherical images according to directions and zoom in or zoom out; the latter is formalized as a thumbnail of the plan of the building that visually tracks the position of the user in the VT and displays their optical cone, suggesting the user orientation in the map.

Another thought laying in accessibility and usefulness UX dimensions was on the possibility for users to freely explore the VT without being distracted by redundant visual indicators and components. For this reason, an eye-like icon was added onto the interface allowing users to enable or disable the rest of the icons, thus visually freeing the interface and facilitating a complete exploration of the spherical images. Finally, the definitive VT interfaces took shape (Fig. 6).

At this final stage, numerous tests were carried out to favour continuity along the route and inter-visibility between acquisition points that are placed along the pavement of the explorative area.

Users can in fact interact with both the observation points green-colored in the map and the acquisition points that can be found as users dive deep in the VT route. To furthermore enhance the exploration of the VT, an additional interactive component was added next to the building's plan, as so the icon displaying the mapped and explorable floors allowing users to directly teleport from one level to another just by clicking it.

***Figure 5:** Wireframe of VT interface with designed icons and interactive elements put in place (Editing by M. Rinascimento)*

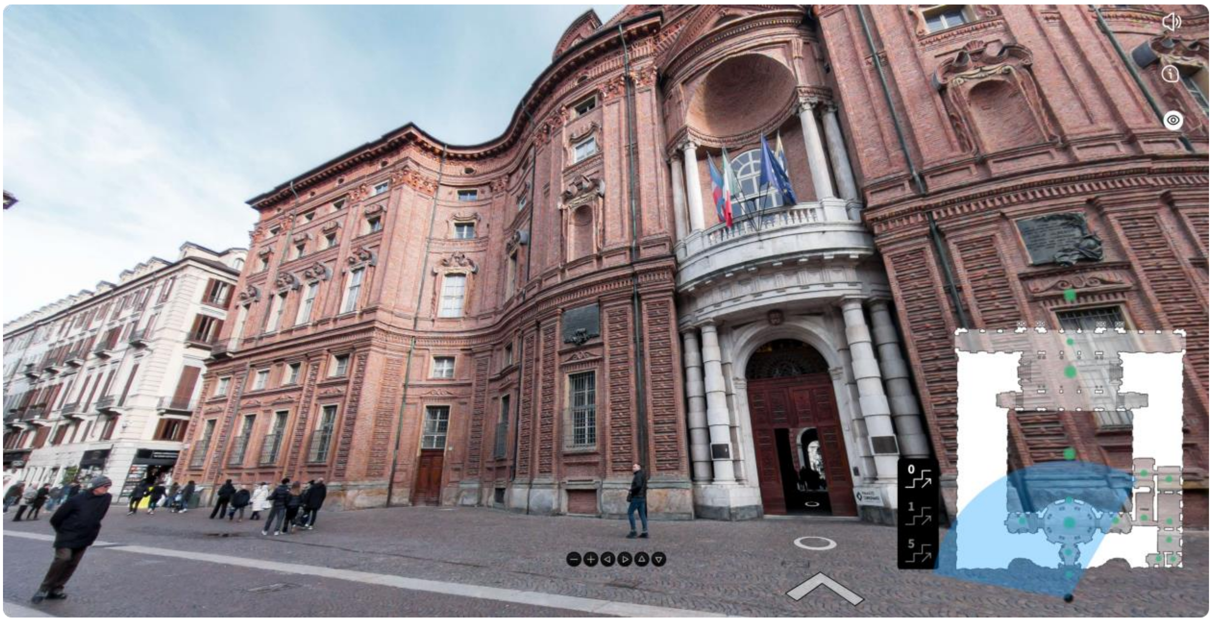

Figure 6: *Definitive VT interface displaying the façade of Palazzo Carignano (Development by M. Russo. UX Research and Design by M. Rinascimento)*

7. IMMERSIVE FRUITION OF EXISTING ARCHITECTURAL HERITAGE

The VT solution was identified as optimal for the Palazzo Carignano Museum. The choice of trade-off between accuracy and quality in cultural heritage documentation avoided the resource expenditure required by detailed geometric and morphological modeling with high-resolution textures, which was misaligned with respect to the necessary communicative output [7] [8]. Alternative approaches, although producing metrically accurate and morphologically detailed models, require articulated workflows for processing, optimization, and integration of heterogeneous datasets, with consequently significantly higher computational and temporal resource requirements [9].

VT based on spherical images, however, constrains public fruition to predetermined station points from the acquisition phase, limiting spatial continuity and the perception of motor freedom. These are essential elements for a robust sense of presence, understood as the product of immersion and coherence [10]. Nevertheless, although this feature correlates with navigation fluidity, user control, and sensory feedback [11], it can be elevated even with interaction limitations, being associated with better cognitive performance and spatial abilities [12].

The immersiveness of the experience can be significantly enhanced through the use of Head-Mounted Displays (HMDs), which substantially improve the sense of presence compared to desktop monitor visualization. HMDs provide a wide field of view, three-degrees-of-freedom (3-DoF) tracking, and sensory isolation from the surrounding environment, thereby favoring perceptual immersion. Comparative studies demonstrate that 360° fruition via HMD produces superior perception of control and immersion compared to desktop screens, despite the limitations of constrained navigation [13] [14].

The 3DVista Virtual Tour PRO software enables the conversion of the VT into an immersive experience for HMDs through the insertion of a VR icon in the UI and the enabling of an export option that optimizes content through foveation techniques and visual quality calibration for different hardware architectures, including Meta Quest, Pico, and HTC Vive. The platform implements two interactive paradigms: gaze-based interaction, which reduces cognitive complexity and mitigates the risk of Visually Induced Motion Sickness (VIMS) [14], and controller-based interactions, which expand interactive possibilities by enabling the activation of embedded multimedia contents.

HMDs generate multiple benefits: they enhance cognitive and emotional engagement, improving information retention; facilitate cognitive accessibility by creating controlled environments advantageous for users with neurodiversity; and support wayfinding strategies through intuitive three-dimensional graphic elements. However, the spatial restrictions of constrained navigation and potential discomfort factors, such as visual fatigue and VIMS, persist, intensifying between 10 and 20 minutes of exposure, although some users demonstrate good adaptive

capacity [15]. The immersive experience is therefore optimal for a heterogeneous audience for sessions of approximately 15 minutes and targeted fruitions, while the desktop version remains advisable for extensive explorations and users sensitive to VIMS.
For the case study of the Palazzo Carignano Museum, integration with web-based platforms extends the dissemination potential by enabling remote access through consumer-grade devices, transforming the VT into a communicative tool adaptable to different modalities and contexts of fruition, in alignment with the project's objectives of inclusion and accessibility to cultural heritage.

8. FUTURE EXPERIENCES IMPLEMENTATIONS

At the current stage of development, the Virtual Tour constitutes an explorable spatial database, as described in the previous paragraphs, offering different levels of immersion and a high degree of user freedom. It effectively provides a basis for the implementation of additional content aimed at enriching the user experience through different media:
Audio tracks (music, background noises, voices) that evoke, through movement within the virtually visited places, scenarios of use of the spaces in relation to their intended use in a specific historical period, or that illustrate, through multilingual narration, architectural and spatial characteristics, historical notes, curiosities, or more.
• Multilingual subtitles to ensure maximum inclusion of different types of audiences in support of the audio tracks,
• Videos in Italian Sign Language (LIS) for brief descriptions and illustrations of the contents and highlights of the tour
• Points of Interest (POI), i.e. sensitive points anchored to specific elements of the spherical image, designed as links to specific insights into the architecture visited or objects of particular interest contained therein: these insights may include texts, historical images, archive drawings, documents, three-dimensional models navigable in VR or multimedia content of various kinds.
The design of content related to virtual tours is a particularly delicate matter, requiring the convergence and harmonisation of numerous skills: the content must strictly adhere to a narrative project that intertwines the cultural demands of museum curators, the expertise of communication and User eXperience specialists, and the know-how of those who produce three-dimensional and VR models, audio and video content. The harmonisation of this wide range of content and media must aim to ensure easy, fluid, and enjoyable use, while meeting standards of scientific accuracy. On the one hand, therefore, the content must be presented in an appealing and easy-to-explore way, carefully balancing the different types of content and offering customisable viewing options based on the interests and time available for the virtual visit. On the other hand, in line with the curatorial approach, it must be presented using language that is cognitively inclusive, accessible and understandable to the different segments of the audience for whom the experience is intended.
Based on these assumptions, the research group aims to continue its collaboration with the Palazzo Carignano museum to develop content that complements and supports the virtual visit experience.

9. CONCLUSION

In the field of museum visits, virtual tours have, as we have seen, experienced impressive growth during the pandemic. This has led to their use as a substitute for in-person visits, in which, however, the connection with the architecture has been lost in many cases. A museum as unique as Palazzo Carignano, a building of great artistic value, is not typically used as an exhibition space for the visual arts, but rather as a testament to the historical events that preceded the unification of Italy.
The focus of this work on the use of architectural space, the specific nature of its rooms, and the functions it has assumed over time, as well as its furnishings, characterizes the results of the research, together with the main objective of making spaces available that are not open to visitors or accessible to everyone.

10. ACKNOWLEDGMENT

This paper, whose authors shared the methodological framework, was written by R. Spallone (par. 2), F. Natta (par. 3), M. Russo (par. 4) M. Casciola (par. 5), M. Rinascimento (par. 6), E. Pupi (par. 7), and M. Vitali (par. 8).
The authors wrote together pars. 1, 9.

11. REFERENCES

[1] Tomasella, Noemi; Camagni, Flavia; Ippoliti, Elena: Il ruolo dei Virtual Tour per l'accessibilità e l'inclusione del patrimonio museale. In: Empler, T. et al. (eds), *DAI - Il Disegno per l'Accessibilità e l'Inclusione*, PUBLICA, Alghero, 2024, pp. 508-527.

[2] Bono, Jacopo: The cultural role of Digital Platforms for the museum heritage experience. In: Carlevaris, L. et al. (eds), *èkphrasis. Descrizioni nello spazio della rappresentazione*, FrancoAngeli, Milano, 2025, pp. 2335-2358.

[3] Meloni, Alessandro: *Percepire e rappresentare lo spazio oltre la visione*, FedOAPress, Napoli, 2025.

[4] Li, Jia; Nie, Jin-Wei; Ye, Jing: Evaluation of virtual tour in an online museum: Exhibition of Architecture of the Forbidden City. *PLoS ONE*, vol. 17, no. 1, pp. 1-17, January 2022.

[5] Paladini, Alice, et al.: Impact of virtual reality experience on accessibility of cultural heritage. *GEORES 2019 – 2nd International Conference of Geomatics and Restoration*, Milan, Italy, 8-10 May 2019, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-2/W11, pp. 929-936.

[6] Lasorella, Margherita; Cantatore, Elena; Fatiguso, Fabio: Smart approaches to enhance technical knowledge of in/low-accessible heritage. *SCIRES-it*, vol. 11, no. 2, pp. 97-112, 2021.

[7] Fascia, Rebecca; Barbieri, Federico; Gaspari, Federica; Ioli, Francesco; Pinto, Livio: From 3D survey to digital reality of a complex architecture: a digital workflow for cultural heritage promotion. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-2/W4-2024, pp. 205-212, 2024.

[8] Gaspari, Federica; Barbieri, Federico; Fascia, Rebecca; Ioli, Francesco; Pinto, Livio: An Open-Source Web Platform for 3D Documentation and Storytelling of Hidden Cultural Heritage. *Heritage*, vol. 7, no. 2, pp. 517-536, 2024.

[9] Rodríguez Navarro, P.; Gil Piqueras, T.; Verdiani, G.: Comprehensive Methodology for Documenting the Defense Towers of the Valencian Coast (Spain). In: *Defensive Architecture of the Mediterranean: XV to XVIII centuries. Vol. I*, Editorial Universitat Politècnica de València, Valencia, 2015, pp. 321-328.

[10] Skarbez, Richard; Smith, Missie; Whitton, Mary C.: Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, vol. 2, 2021.

[11] Thorp, Sebastian Oltedal; Rimol, Lars Morten; Lervik, Simon; Evensmoen, Hallvard Røe; Grassini, Simone: Comparative analysis of spatial ability in immersive and non-immersive virtual reality: the role of sense of presence, simulation sickness and cognitive load. *Frontiers in Virtual Reality*, vol. 5, 2024.

[12] Palombi, Tommaso; Galli, Federica; Giancamilli, Francesco; D'Amico, Monica; Alivernini, Fabio; Gallo, Luigi; Neroni, Pietro; Predazzi, Marco; De Pietro, Giuseppe; Lucidi, Fabio; Giordano, Antonio; Chirico, Andrea: The role of sense of presence in expressing cognitive abilities in a virtual reality task: an initial validation study. *Scientific Reports*, vol. 13, no. 1, p. 13396, 2023.

[13] Koeva, Mila; Luleva, Mila; Maldjanski, Plamen: Integrating Spherical Panoramas and Maps for Visualization of Cultural Heritage Objects Using Virtual Reality Technology. *Sensors*, vol. 17, no. 4, p. 829, 2017.

[14] Argyriou, Lemonia; Economou, Daphne; Bouki, Vassiliki: Design methodology for 360° immersive video applications: the case study of a cultural heritage virtual tour. *Personal and Ubiquitous Computing*, vol. 24, no. 6, pp. 843-859, 2020.

[15] Zielasko, Daniel; Rehling, Ben; von Dawans, Bernadette; Domes, Gregor: Do Not Immerse and Drive? Prolonged Effects of Cybersickness on Physiological Stress Markers And Cognitive Performance. *arXiv preprint*, arXiv:2506.11536, 2025.

Virtuality as Knowledge: Reclaiming Unbuilt Architecture through Immersive Archives

Nicola D'Addario

Faculty of Architecture, University of Porto, Portugal, nic.daddario@gmail.com

ABSTRACT: What does it mean to archive architecture that was never built? This paper explores how immersive technologies can transform the historiography of architecture by reclaiming unbuilt or lost projects as epistemic spaces. Traditional archives reduce architectural knowledge to drawings and texts, but digital simulations enable a phenomenological reconstruction of spatial experience—an embodied encounter with the "unrepresentable" dimension of architecture. Building on theories of perception, Gestalt, and architectural phenomenology, this study examines virtual reconstructions as performative interfaces that generate knowledge through experience rather than representation. By addressing the notion of "hypothetical heritage," it argues that virtuality does not merely simulate the past: it reactivates it, converting absence into a critical mode of presence. The paper situates immersive archives within the broader discourse of digital heritage, proposing virtuality as a legitimate domain of architectural understanding in the age of AI and computational creativity.

1. INTRODUCTION

Architecture has always been mediated through representation. Plans, sections, and models are not mere technical devices but epistemic instruments through which space is conceived, transmitted, and remembered. Yet, the conventional architectural archive—composed of drawings, photographs, and documents—inevitably reduces architecture to its visual and textual fragments. As Bruno Zevi famously argued, architecture is not seen in plan or elevation, but in the lived experience of space [1, 2]. This phenomenological claim exposes a critical paradox: how can we know architecture that no longer exists, or that was never built at all?

In the age of digitality and artificial intelligence, this question acquires renewed urgency. While tools like 3D modelling, photogrammetry, and photorealistic rendering are now widely used for architectural representation—though more often for design and analysis than for historical understanding—immersive visualization and interactive simulation enable forms of engagement that transcend static media. These new tools allow users to inhabit digital spaces and to experience architecture as temporal, sequential, and embodied: an encounter closer to Architecture itself. Within this context, unbuilt and destroyed works—long confined to the margins of historiography, despite being substantial to the advancement of the discipline—can be reactivated as experiential archives.

This paper proposes to examine digital virtuality as a form of knowledge: a means to partially reconstruct architectural presence through phenomenological simulation rather than material conservation of analog traces. The term knowledge here does not refer merely to the accumulation of information about architectural form, but to the process through which meaning is constituted in perception. Following Michael Polanyi's notion of tacit knowledge [3], architectural understanding arises from embodied engagement—the "knowing how" that precedes any explicit "knowing that." Likewise, Husserl's *Erfahrung und Urteil* [4] situates experience as the precondition of all conceptual judgment. Within this framework, immersive archives produce not descriptive data but experiential insight: they operationalize knowledge through appearance, orientation, and movement. Virtuality thus reconfigures architectural historiography from a discourse of representation to one of experiential cognition. By situating this approach within the broader debates on digital heritage, post-representational historiography, and computational creativity, the paper argues for an expanded notion of the architectural archive—one that acknowledges virtual reconstructions as critical and epistemic spaces.

2. MAIN ASPECTS

2.1 FROM REPRESENTATION TO SIMULATION

Architectural knowledge has historically been mediated through systems of representation: drawing, perspective, photography, and, more recently, digital modelling. Each of these has produced a specific epistemology of space, defining what can be known, preserved, and transmitted. This problem of translation—from the abstract language of the drawing to the phenomenal reality of the building—was thoroughly explored by Robin Evans [5]. As Zevi and Evans demonstrate, and as phenomenology has long argued, architectural representations dissect, but they do not synthesize the lived experience.
It is critical to clarify that the "virtual" is not synonymous with the "digital." In Deleuze's ontology, the virtual designates that which is real but not yet actual—an intensive field of potentialities structuring the conditions of emergence [6]. This aligns with the classical etymology of the word, which is rooted in potentiality. Pierre Lévy later extended this to the digital domain, describing virtuality as "a mode of being" that demands actualization rather than reproduction [7].
This apparent overlap is not sufficient to consider the terms interchangeable. We should, perhaps, speak of Digital Virtuality: a state where the digital is the medium of application and virtuality is the "mode" of potentiality being actualized. Seen through this lens, virtual reconstructions do not imitate architecture; they activate its latent possibilities. The immersive archive proposed here, therefore, operates as a site of actualization, where the conceptual potential of unbuilt works becomes phenomenally accessible.
This challenges the logic of the modern archive, which, as Foucault suggests, is not a neutral repository but a dispositive that orders knowledge through visibility and classification [8]. In architecture, this logic privileges the image of the building, or its quantifiable data, over its experience.
The emergence of digital and immersive technologies confronts this representational paradigm. While early computer visualization replicated the conventions of drawing, contemporary simulation tools—such as interactive VR environments or HBIM platforms—transform the act of viewing into a performative and temporal process. The archive becomes an interface: not merely a place where information is stored, but a field where spatial knowledge is enacted.
In this sense, virtual reconstruction is not a derivative of the real but a mode of re-presencing: the digital reenactment of spatial conditions, atmospheres, and perceptual rhythms otherwise lost to history. This shift invites a redefinition of the archive itself. As Derrida argued in Mal d'archive, every act of archiving is haunted by the desire both to preserve and to destroy—a tension he calls "archive fever" [9]. Hal Foster later reframed this ambivalence as an "archival impulse," in which contemporary theorists and artists transform archival practice into a generative, performative act [10]. The immersive ar-

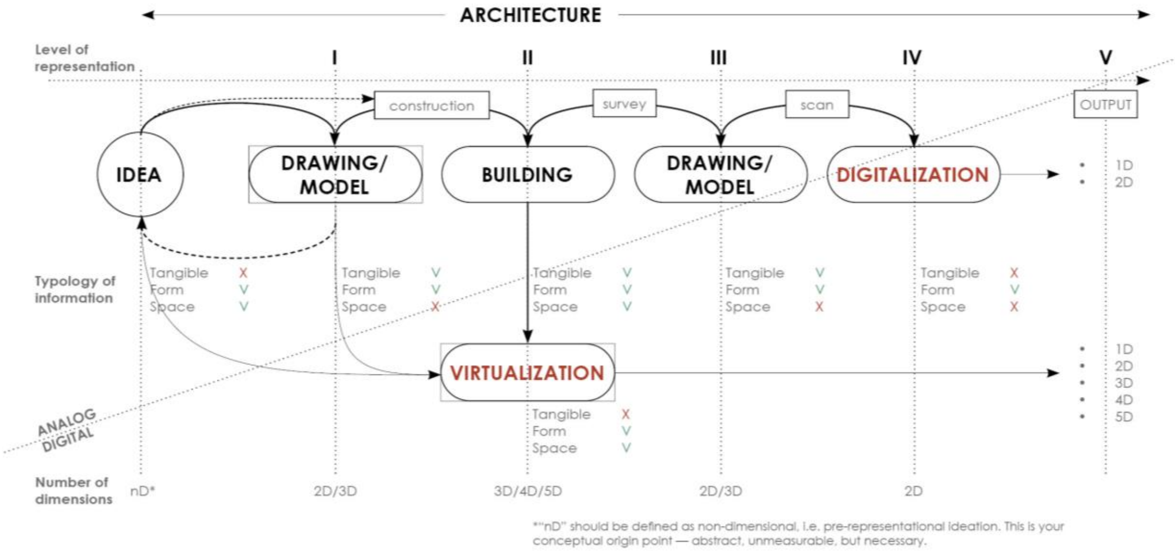

***Figure 1:** This framework distinguishes "Digitalization" (IV)—a static 2D output—from "Virtualization." The matrix (left) shows how virtualization is the only mode of representation that, like the physical "Building" (II), can actualize the phenomenological dimension of "Space," which is lost in all other media*

chive continues this trajectory: it no longer safeguards architectural traces, but performs their reactivation. What emerges is a critical form of post-archivality, in which knowledge is produced through the encounter, not merely stored for retrieval.

2.2 THE HYPOTHETICAL HERITAGE

We propose the notion of "hypothetical heritage" based on the recognition that cultural value often exceeds material survival. Architectural history abounds with projects that, though unbuilt, have shaped discourse and imagination: from Piranesi's visions to the utopian projects of Ledoux and Sant'Elia. Traditionally, these works were transmitted as static drawings—"paper architectures" that testified to ideas rather than inhabitable forms, yet still exerted enormous influence.

This concept applies equally to lost architectures that survive only through fragmentary representations. A key example is Paul Letarouilly's monumental survey of the Vatican and Rome's Renaissance buildings, which—acting simultaneously as survey and reconstruction—preserved the spatial logic of structures now long-altered or destroyed. Following the 2003 UNESCO Convention on Intangible Heritage and the 2005 Faro Convention, heritage studies have increasingly acknowledged immaterial practices, memories, and knowledges as integral to cultural identity. Extending this framework to architecture, hypothetical heritage refers to those non-built or lost works whose conceptual potency continues to influence spatial culture. Their preservation, therefore, relies not on material conservation but on the potential to reactivate their experiential dimension through digital means.

Immersive reconstructions, therefore, become instruments of critical remembrance, transforming absence into a productive category of cultural resilience. Yet the reactivation of hypothetical heritage also raises ethical and political questions. As Rodney Harrison reminds us, heritage is never neutral—it is a "mode of cultural governance" that frames how societies relate to their pasts [11]. In the digital realm, issues of authorship, authenticity, and ownership become even more complex: who "owns" the reconstruction of an unbuilt project? What happens when algorithmic processes mediate cultural memory? These questions compel us to treat the immersive archive not simply as a technical achievement but as a site of negotiation, where competing claims to history, authorship, and identity are staged anew.

2.3 PHENOMENOLOGY AS CRITICAL LENS

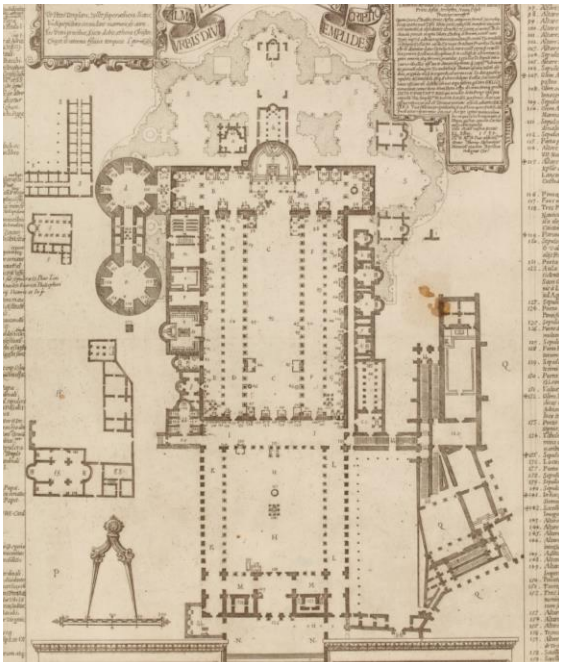

Figure 2: *Paul Letarouilly, plan of St. Peter's Basilica (c. 1840-1860). This plate exemplifies "hypothetical heritage" as it applies to lost architecture. It functions as both a fragmentary record (survey) and a critical interpretation (reconstruction), preserving a spatial logic that has since been altered or destroyed.*

To evaluate the epistemic potential of this newly defined immersive archive, the paper adopts a phenomenological perspective. From Husserl's notion of intentionality to Merleau-Ponty's embodied perception [12], phenomenology has emphasized that knowledge of space arises through experience rather than abstraction. In architectural thought, this lineage informs the work of Norberg-Schulz, Holl, and Pallasmaa, for whom architecture is understood as a lived horizon of orientation, movement, and atmosphere [13].

Immersive simulations invite a re-reading of this tradition in the digital age. They do not replace the body but extend it—creating what Mark Hansen describes as a new corporeal interface [14]. Don Ihde's post-phenomenology provides a useful framework for understanding this extension. In *Bodies in Technology* [15], Ihde describes how technological mediation neither alienates nor disembodies the subject, but multiplies perceptual relations between self and world. Similarly, Böhme's aesthetics of atmosphere [16]—a concept with clear parallels to Gestalt theory—situates perception as a coproduction between the sensing body and its spatial milieu, where the “whole” is perceived before its parts.

The immersive archive exemplifies this relational ontology: it produces atmospheres that are both digitally computed and phenomenally lived, generating what might be called a mediated corporeality—a body expanded by the very technologies that once seemed to distance it from architectural experience. Within virtual environments, spatial phenomena such as light, sound, and materiality can be reconstituted as dynamic perceptual events, allowing users to experience the previously "unrepresentable" qualities of architecture through mediated presence. This constitutes a form of digital phenomenology: not a substitute for the real, but a way of knowing through appearance, engagement, and affective resonance.

2.4 CASE ILLUSTRATION: THE VIRTUAL DANTEUM

To exemplify how immersive reconstruction can function as epistemic practice, this section presents an ongoing project. The virtual reconstruction of Giuseppe Terragni’s *Danteum* (1938) serves as a concrete experiment in the epistemic potential of immersive archives. Conceived as a spatial translation of Dante’s *Divina Commedia*, the original project remained unbuilt, surviving only through a limited set of drawings and model photographs. Its radical abstraction—geometry as allegory, procession as narrative—has long rendered it an emblem of architecture’s speculative dimension. Indeed, its pure rationalist language, lacking familiar references or elements, makes it nearly impossible to grasp scale or perception from the plans alone.

The present immersive reconstruction reanimates this latent architecture through a fully navigable 3D environment built in Unreal Engine 5. The choice of a so-called “game engine” is deliberate: it highlights the proximity between the gaming experience and the pedagogical one we propose. This links the work to the classical concept of ludus (play) as a tool for learning, an idea explored by thinkers from Plato to the Roman orator Quintilian. Rather than adopting the first-person “visitor” perspective (common to both VR and many screen-based games), this third-person configuration positions the user as both observer and participant, producing a doubled mode of perception that mirrors the poem’s own dialectic between vision and narration.

This choice of a third-person perspective, while seemingly at odds with first-person phenomenology, is a deliberate methodological decision. It addresses the “body problem” not through haptic simulation, but through representation of scale and presence. The visible avatar acts as a human proxy or testimone (witness), allowing

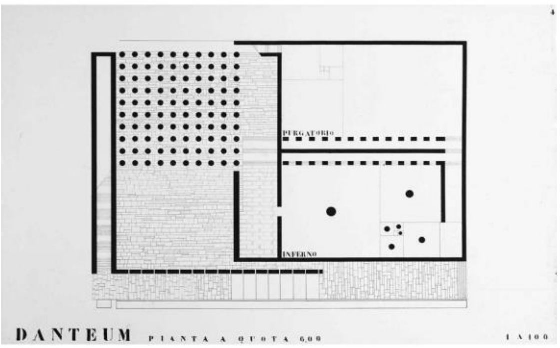

Figure 3*: Giuseppe Terragni, plan for the Danteum (1938). The primary document of this “paper architecture.” Its pure abstraction and lack of familiar references render the project's scale and phenomenological qualities—such as the disorientation of the pillar forest (top left)— "unrepresentable” in two dimensions.*

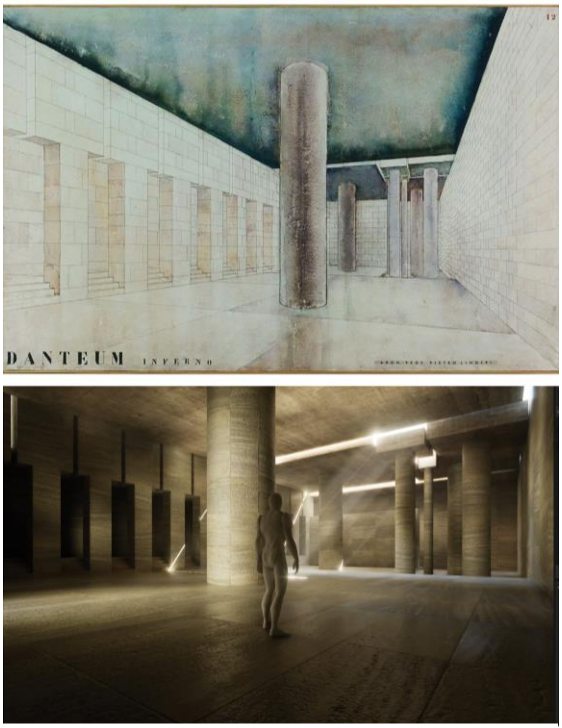

Figure 4: *Comparison of historical representation (Inferno) with the immersive archive. Top: Terragni's original perspective rendering. Bottom: The UE reconstruction. The addition of the avatar and the simulation of dynamic light and shadow demonstrate the method's ability to 're-presence' the intended phenomenological script of the space*

Figure 5: *Comparison of historical representation (Purgatorio) with the immersive archive. Top: Terragni's original perspective rendering. Bottom: Corresponding view from the UE reconstruction. The visible avatar, acting as a testimone (witness), transforms the static scene into a performative interface, providing an intuitive grasp of the monumental scale*

the user to observe a body in relation to the architecture. This provides an immediate, intuitive grasp of proportion, scale, and the spatial relationship between the human form and Terragni’s monumental voids—a form of “vicarious” embodiment. It also serves a practical function, allowing the experience to be shared and discussed by a group, rather than isolating a single user in a VR environment.
Within this environment, the Danteum unfolds not as a static monument but as a phenomenological script: the progression from the Inferno’s darkness to the Paradiso’s luminosity becomes a lived, temporal sequence mediated through spatial thresholds, acoustic cues, and material atmospheres. The reconstruction thereby transforms Terragni’s unrealized project into an experiential hypothesis—a performative enactment of his architectural poetics rather than a mere visualization. This immersive enactment yields specific epistemic insights unavailable from the static plans. For instance, while the drawings suggest the “Inferno” is dark, the simulation allows one to *experience the phenomenological script* of light's variation during the threshold transition. More critically, the “forest” of pillars at the entrance, often read as a simple formal device, is revealed in the simulation as a powerful instrument of disorientation; the user experiences this loss of direction in real-time, as the original poem suggests and the architect probably intended. This is not the static, “retinal” architecture that phenomenologists rightly critique. Rather, it is a dynamic, sequential engagement of vision—the change in light and exposure during movement—that is fundamental to ancestral embodied cognition. As such, the Virtual Danteum demonstrates how immersive archives can operate as instruments of historical inquiry, enabling researchers to test and interpret the perceptual logic of unbuilt works. What emerges is not a simulation of Terragni’s architecture, but a critical re-presencing: an interpretative space where historical imagination and digital phenomenology converge.

3. CONCLUSION

The progressive convergence of digital technology, artificial intelligence, and architectural historiography compels a reconsideration of what it means to preserve, study, and transmit architectural knowledge. This paper has argued that virtual reconstruction should not be understood as a form of illusion or simulation of the real, but as an epistemic practice capable of reactivating the experiential dimension of architecture. By transforming the archive from a static repository into an immersive interface, digital environments enable a performative understanding of architectural heritage—one grounded in perception, temporality, and embodied cognition.
Through the notion of hypothetical heritage, the study repositions unbuilt or lost projects within the field of digital heritage, proposing that absence itself can become a source of knowledge. The virtual archive thus operates as a site of critical reconstruction.
This reconstruction is, fundamentally, a hermeneutic act. Given that the records of hypothetical heritage are often partial or fragmentary, the work is never a neutral technical task but an “interpretation”. This task demands a specific, disciplinary *saper fare* (know-how). It is not archaeology, which typically builds upon physical traces, nor is it traditional history, which is often more strictly bound to textual sources. Neither is it the work of a 3D artist, who retains the freedom to invent without disciplinary constraints. This reconstruction is a job for the architect, the architectural historian, and the theo-

rist, as it is akin to developing a "project" (*progetto*)—a critical proposal that synthesizes fragmentary evidence into a coherent, experiential hypothesis.

Phenomenology offers a coherent lens through which to interpret this transformation. The advent of generative AI further complicates this epistemic terrain. Algorithmic reconstruction does not merely automate modeling; it introduces a new form of computational imagination, where the system learns spatial logics and generates plausible but non-existent architectures. As Mario Carpo notes, this constitutes a "third digital turn," in which machines participate in design cognition [17]. The immersive archive thus becomes dialogical: a site where human (interpretive) and machine (generative) intelligences co-produce architectural meaning. The question shifts from what we can reconstruct to how we can collaborate with algorithmic processes in re-enacting architectural memory.

Immersive archives do not replicate reality but stage it anew, revealing how spatial understanding arises from the dynamic interplay between body, perception, and technology. The Virtual Danteum, for instance, underscores the proposition that virtuality, when understood as epistemic reactivation, can transform architectural historiography from an act of recovery into one of creation. By re-staging Terragni's unbuilt vision as a navigable experience, the immersive archive operates not only as a vessel of memory but as a site of thought—an experiment in how architecture continues to think through us, even when it was never built. This points toward a future phenomenology of the digital, one that conceives virtuality not as a surrogate of the physical but as an autonomous and legitimate mode of knowing architecture, where the archive itself becomes a living thought-space.

4. REFERENCES

[1] Zevi, Bruno: *Saper vedere l'architettura*, Einaudi, Torino, 1948.

[2] Zevi, Bruno: *Architettura in Nuce*, Einaudi, Torino, 1960.

[3] Polanyi, Michael: *Personal Knowledge: Towards a Post-Critical Philosophy*, University of Chicago Press, Chicago, 1958.

[4] Husserl, Edmund: *Experience and Judgment: Investigations in a Genealogy of Logic*, Northwestern University Press, Evanston, 1973 [orig. 1939].

[5] Evans, Robin: *Translations from Drawing to Building and Other Essays*, MIT Press, Cambridge (MA), 1997.

[6] Deleuze, Gilles: *Difference and Repetition*, Columbia University Press, New York, 1994 [orig. 1968].

[7] Lévy, Pierre: *Qu'est-ce que le virtuel?* La Découverte, Paris, 1995.

[8] Foucault, Michel: *The Archaeology of Knowledge*, Pantheon Books, New York, 1972.

[9] Derrida, Jacques: *Archive Fever: A Freudian Impression*, University of Chicago Press, Chicago, 1996 [orig. 1995].

[10] Foster, Hal: "An Archival Impulse," *October*, vol. 110, pp. 3–22, 2004.

[11] Harrison, Rodney: *Heritage: Critical Approaches*, Routledge, London, 2013.

[12] Merleau-Ponty, Maurice: *Phenomenology of Perception*, Routledge, London, 2002 [orig. 1945].

[13] Pallasmaa, Juhani: *The Eyes of the Skin: Architecture and the Senses*, Wiley, Chichester, 1996.

[14] Hansen, Mark B.N.: *New Philosophy for New Media*, MIT Press, Cambridge (MA), 2004.

[15] Ihde, Don: *Bodies in Technology*, University of Minnesota Press, Minneapolis, 2002.

[16] Böhme, Gernot: "Atmosphere as the Fundamental Concept of a New Aesthetics," *Thesis Eleven*, no. 36, pp. 113–126, 1993.

[17] Carpo, Mario: *The Second Digital Turn: Design Beyond Intelligence*, MIT Press, Cambridge (MA), 2017.

Holographic Artifacts for the Enhancement of Academic Heritage: The Curioni Collection at the Polytechnic University of Turin

Maurizio Marco Bocconcino (MMB), Mariapaola Vozzola (MPV), and Emanuele Ricchiello (ER)

Department of Structural, Building and Geotechnical Engineering, Politecnico di Torino, Italia, maurizio.bocconcino@polito.it

ABSTRACT: This paper develops a methodological reflection on strategies for enhancing the wooden models created by professor Giovanni Curioni in the second half of the nineteenth century, aimed at investigating the structural principles underlying architectural and engineering forms. The study examines the academic collection preserved at the Department of Structural, Building, and Geotechnical Engineering of the Politecnico di Torino, which includes over 140 scale models derived from the books L'arte di fabbricare (1867–1885), conceived as didactic instruments for the emerging discipline of construction science, in which Curioni was a pioneer in Italy. Within the framework of the university's Third Mission, the research promotes the creation of a Virtual Museum to disseminate and valorize polytechnic collections through digital platforms and interactive interfaces. The integration of holographic technologies enables the transformation of manuals and models into dynamic, three-dimensional projections, fostering new modes of knowledge transmission. Preliminary applications to selected models from the Curioni Collection highlight the interpretive and narrative potential of holographic representation as a medium for contemporary scientific and cultural communication.

1. INTRODUCTION (MMB)

Research has been carried out on Giovanni Curioni's collection of wooden models, which is preserved at the Politecnico di Torino's Department of Structural, Building and Geotechnical Engineering (DISEG). This research explores the role of Drawing as a cognitive and tactile infrastructure for knowledge. An operational system for the enhancement of academic heritage has been developed through an integrated process of surveying, modelling, digitisation, 3D printing and holographic projection. The project emphasises the importance of striking a balance between digital mediation and physical tangibility, which includes returning the original models to educational settings. Representation is interpreted as a cyclical process in which drawing connects material memory and technological innovation to generate shared, inclusive knowledge.

This contribution offers a methodological reflection on the enhancement of analog models created in the second half of the 19th century. His work aimed to explore the structural essence of architectural and engineering forms. The collection includes over 140 scale models derived from illustrations in Curioni's seminal work L'arte di fabbricare (1867–1885), conceived as a scientific aid for teaching construction science, a field in which Curioni is considered a pioneer in Italy.

The research, aligned with the goals of the university's Third Mission, seeks to promote and disseminate knowledge of polytechnic collections through the creation of a Virtual Museum with the aim of stimulating and encouraging a return to reading physical models and a deeper and more conscious material contact with them. This museum will host information models and digitized documents, accessible through multiple platforms - web repositories, physical and digital interfaces - to reach a broad and diverse audience. The Curioni collection is thus reinterpreted as a contemporary communication tool, where manuals and models are transformed into

"virtual and dynamic projections" through new representational technologies.
Central to this transformation can become the use of holography, a technology based on light diffraction and interference that produces highly realistic three-dimensional images. Holographic devices such as tables, display cases, and projectors enable immersive environments in which images and stories are projected, creating new experiential realities. This new research phase investigates the potential of holographic reproduction, user interaction with holographic artifacts, and how visual representation techniques support this intangible but spatially and materially grounded form of communication. A preliminary application to selected models from the Curioni Collection highlights the narrative and interpretive opportunities enabled by holographic display.

2. ORIGINS, MOTIVATIONS AND KNOWLEDGE PATTERNS FOR THE RESEARCH (MMB)

The research on the Curioni Collection was developed within the domain of Survey and Drawing, which is concerned with the "generation, construction and analysis of drawings, images and models as the results of scalar representations of existing or designed realities" and the "visual translation of concepts, ideas and narratives".
In this epistemic framework, Drawing is conceptualised as a cognitive language and a medium for mediating the relationship between the tangible world and its informational translation.
The collection of wooden models created by Giovanni Curioni in the 19th century as educational aids for construction constitutes a technical and educational heritage that is unique in terms of quality and completeness (Figure 1). The objective of the research was to restore the models to their original function as tools of knowledge. This was achieved by reinterpreting them in the contemporary digital context and placing them at the centre of a process of representation that combines documentation, communication and, ultimately, teaching.
The activity forms part of a multi-year process of enhancing the technical and scientific heritage of DISEG, which commenced with the Mosca Library [1] and continued with the Porcheddu Archive [2]. These experiences have facilitated the experimental development of methodologies for the surveying and digitisation of historical technical documents, thereby establishing the theoretical and operational foundations for the more complex digitisation of the Curioni Collection.
The most recent phase of the process was represented by the experimentation conducted on the Betta-Bardelli Archive [3], which saw the consolidation of the methodology and its expansion to the field of the representation of constructive memory.

Figure 1: *Classification of models in the Curioni collection.*

In this continuum, the Curioni Collection constitutes the central axis of a research project that has confirmed drawing into a cognitive and disciplinary infrastructure, capable of connecting the history of technical knowledge to its contemporary digital reinterpretation. The objective was twofold: firstly, to document, and secondly, to reactivate the models as formative and perceptive devices. In doing so, drawing, in the broadest sense of the term and in its original form, was placed at the centre of the process of knowledge.

2.1 METHODS AND OPERATIONAL TOOLS (MMB)

The operational process involved three primary methods of surveying and modelling, which were executed concurrently [4].

The acquisition of the subject was undertaken with the employment of a precision laser scanner, with the objective of producing high-density three-dimensional point models for the purpose of conservation documentation.
The utilisation of smartphones equipped with LiDAR cameras facilitates low-cost surveying, a technique that is conducive to the expeditious and replicable acquisition of data for educational and informational objectives.
The reconstruction of the artefacts was simultaneously informed by Curioni's original drawings, which are contained in the volumes entitled L'arte di fabbricare (The Art of Manufacturing). The geometric reconstruction was based on historical sources.
These methods were complemented by traditional direct surveying techniques, utilising instruments such as squares, calipers, and metres, which served as a training aid to instruct students in the requisite knowledge for accurate measurement and representation. The educational approach commenced with a reduced scale model of the artefacts, fostering a foundation for students to develop proficiency in measurement and representation at a smaller representation scales.
The plurality of approaches adopted has enabled the consolidation of the concept of drawing as a comparative process, which integrates tools, scales and languages to facilitate an integrated understanding of the object (Figure 2).
Operationally (see. Cap. 3), the processing was carried out using Polycam Pro, Rhinoceros 3D, Blender and Revit for geometric and informational modelling, while the publication of the models on Sketchfab allowed them to be disseminated on the web with metadata and descriptive sheets.
The 3D printing of the digitised models represented the phase of returning to the material: a cognitive and educational act rather than a reproductive one. This made it possible to verify the geometric consistency and restore the physical perception of the form.
Concurrently, the reintroduction of the original wooden models to the students' desks in the courses taught by professors Ursula Zich and Martino Pavignano (third year of the Architecture degree programme) served to reinforce the connection between tactile experience, observation and representation, thereby restoring drawing to its original function as a sensory cognitive practice.

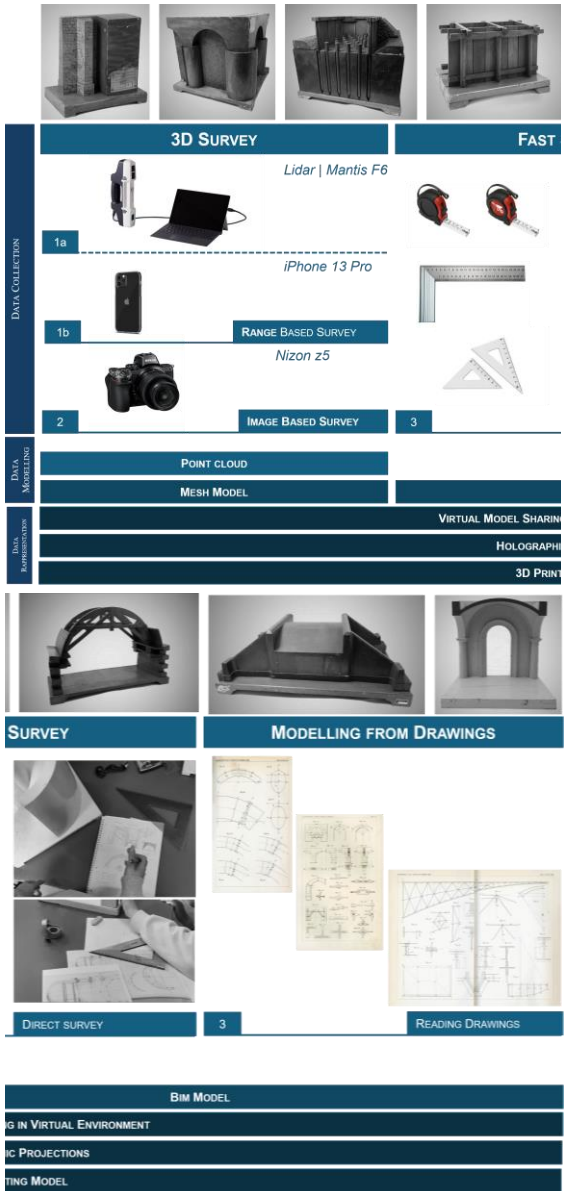

Figure 2: *Survey processes: a matter of integrative techniques in a knowledge system.*

As a preliminary conclusion to the research process, commercial low cost disseminastion of holographic projection technology has paved the way for a novel form of immersive and synaesthetic representation, characterised by the utilisation of digital models. Despite the absence of weight, the hologram restores the three-dimensionality of the form in real space and strengthens the perceptual relationship with the physical object, posing significant problems of scale representation and graphic codes for perception and interaction with projected models, through appropriate interfaces. This phase thus brings the methodological cycle of the research to a conclusion, confirming that digital mediation does not negate matter, but rather serves to amplify its understanding and cognitive value [5].

2.2 EXPERIMENTATION WITH HOLOGRAPHIC IMAGES (MMB)

The most recent updated outcome of the research process was experimentation with holographic images, which served as a phase of reflection on the very nature of representation.

The hologram, understood as a three-dimensional image generated by light and perceived in real space, takes the form of a complex projection, in which the visible form does not exist in itself, but is reconstructed by the observer through an active perceptual process.

It is evident that two primary classifications of holography exist, which differ in terms of their fundamental nature, underlying principles, and methodological value.

The first, more widespread, is that based on retinal memory: the persistence of the image on the retina allows visual perception to reconstruct the continuity of a figure that does not actually exist materially.

This phenomenon is exemplified by rotating LED devices (3D Hologram Fan), which project a sequence of two-dimensional images in rapid succession into space.

The visual system, incapable of perceiving temporal discontinuity, integrates successive projections into a stable figure that appears suspended and three-dimensional.

The holographic image thus arises as a perceptual event, rather than a physical object, and its three-dimensionality is the result of a mental construction based on visual memory.

From a disciplinary perspective, this form of holography is considered to be part of the tradition of drawing, referred to as a "synthetic vision". This is defined as an act of image recomposition through the perception of time and movement.

The perceptual hologram is not a geometric projection; rather, it is a cognitive reconstruction that renders visible the ability of drawing to translate the dynamics of form into visual experience.

The second mode is of a geometric-constructive nature and is distinct in that three dimensionality is not attributed to retinal persistence, but rather to the spatial recomposition of multiple simultaneous projections.

In this instance, holographic reconstruction is based on a principle analogous to that of descriptive projections: a three-dimensional object is represented by four images, placed at the vertices of an optical tetrahedron.

The arrangement of the four views, set at 90° and reflected on semi-transparent surfaces, combines visually at a point of intersection, thereby rendering an actual spatial configuration.

In this case, the hologram is not perceived as an imagined figure, but as a real geometric figure in space, generated by the convergence of orthogonal or perspective projections.

The methodological shift is of pivotal significance: we transition from the act of projection to that of object, from the drawing that serves as a representation of form to the image that meticulously reconstructs it.

From a disciplinary perspective, this signifies that holography transcends its traditional role as a mere advanced form of representation, instead emerging as a three-dimensional synthesis of projections. This assertion positions it as a significant model within the broader history of theories of vision and the construction of space.

The distinction between perceptual and constructive holography provides a methodological framework for interpreting drawing as a medium that transcends conventional boundaries, thereby facilitating the transition from surface to light, from two-dimensional plane to three-dimensional volume, and from the act of drawing to its representation as an image.

In the initial case, representation is considered to be a perceptual act, whereby form manifests itself in the retinal continuity of movement.

In the second, it is a geometric act: form is generated by the convergence of projections, as in descriptive construction or informative modelling.

The two modes articulate the interdisciplinary and cognitive essence of drawing.

The perceptual hologram is the consequence of the synesthetic and temporal dimension of representation, whereby the image is the result of active vision and a bodily experience of space.

Conversely, the constructive hologram serves to reinvigorate the tenets of descriptive geometry, translating the theory of projections into a dynamic three-dimensional system, wherein sections and shadows metamorphose into planes of light interference.

From this standpoint, holographic representation can be regarded as a novel form of projection and section: light substitutes the line, transparency becomes the plane of intersection, and space itself becomes the support of representation.

The section is no longer a static cut, but rather a field of light crossing; the projection is no longer the reduction of form on the plane, but its expansion in real space.

The configuration of the holographic image as an act of three-dimensional drawing is therefore achieved, resulting in a visual construction that facilitates the establishment of a unified perceptual space for both the observer and the object.

3. DEMATERIALISATION OF CURIONI MODELS (MPV&ER)

The process of surveying and digitising wooden models is part of a research project dedicated to the enhancement and digital preservation of the Curioni Collection. The aim of this project is to define a method for managing and archiving three-dimensional models based on the FAIR principles (Findable, Accessible, Interoperable, Reusable), with a view to ensuring the interoperability, traceability and long-term preservation of digital data [6]. The activity was therefore focused on the documentation and conservation of the artefacts, allowing them to be consulted by the public and enjoyed online, with the aim of creating a virtual museum and transforming them into tools for widespread and participatory knowledge. Digital technologies generate new ways of accessing and relating to cultural heritage, fostering a broader dialogue between people and objects and redefining the way the public engages with collections [7].

The first phase involved three-dimensional scanning of the models using the Polycam Pro application, based on a photogrammetric process that combines high geometric accuracy with an efficient and replicable procedure. The acquisition was carried out through automatic video recording, from which the software extracted sequences of images at regular intervals, processed to generate the point cloud and three-dimensional mesh.

The use of mobile devices and photogrammetry and videogrammetry techniques allows three-dimensional models of good accuracy to be obtained with rapid and automated procedures (Figure 3), thanks to the photogrammetric processing of video frames without the need for complex interventions by the operator [8].

Figure 3: *Visualization of the digital model in a mobile environment.*

For each survey, the maximum detail parameters were set, enabling the "use object masking" option to isolate the model from the support surface and surrounding elements. Uniform lighting and a constant shooting distance ensured the photometric homogeneity necessary for the correct texturing of the model. Once the mesh generation was complete, cleaning and finishing were performed directly within Polycam, using the 'Crop Box' command to remove background residue or unwanted portions and obtain a clean, centred geometry (Figure 4).

This preliminary optimisation phase allowed the digital model to be refined and prepared for the modelling and publication phase. Following cleaning, the files were exported in '.obj' and '.fbx' formats and stored in a dedicated folder structure, which also contained the photographs taken during the survey and the cropped versions. Each model was accompanied by a direct link to the relevant Historical Collection of the Polytechnic University of Turin, so as to create a correspondence between the physical object, its digital replica and the archival documentation.

The subsequent operations were carried out in Rhinoceros 3D, where the meshes were imported to create the support base using the surface extrusion command (Figure 5). The base has a dual function: on the one hand, it closes the lower face, which was not photographed and therefore not reconstructed by the mesh, and on the other hand, it provides a unified dimensional reference for all reproductions.

At this stage, a transparent virtual display case was also added, with a protective and museum enhancement function, as well as a graphic scale indicating the main measurements, useful for restoring the actual proportions of the artefact.

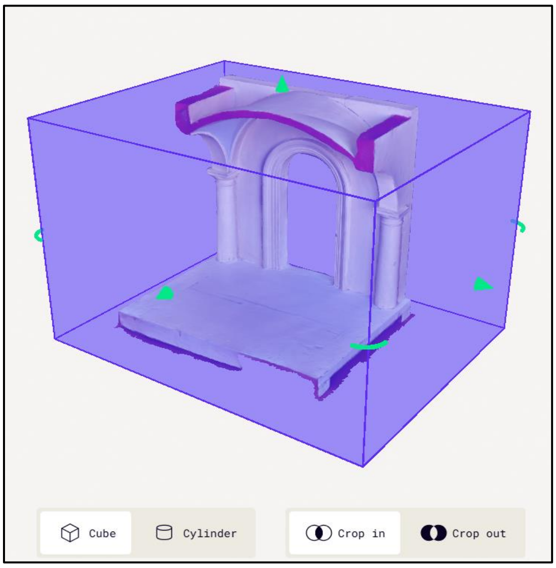

Figure 4: *Visualization of the digital model in Polycam – the "Crop Box".*

Once the complete group, consisting of the wooden model, the base and the display case with measurements, had been defined, it was exported in '.fbx' format, suitable for online publication. The platform chosen for dissemination was Sketchfab, which allows interactive three-dimensional models to be viewed directly from a browser, while maintaining textures and geometric information (Figure 6).

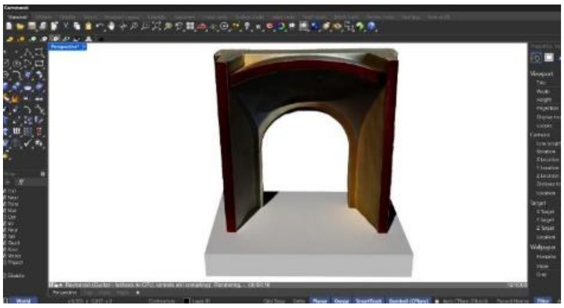

Figure 5: *Visualization of the digital model in Rhinoceros 3D.*

After uploading, the models underwent a visual optimisation process, which included adjusting the lighting, colouring the base and creating descriptive annotations using interactive labels containing technical information, high-resolution photographs and direct links to the Politecnico's Historical Collections website. The integrated annotations make the model not only a three-dimensional object, but also a real information support, through which it is possible to combine spatial representation with the consultation of detailed data and images.

Figure 6: *Visualization of the digital model in Sketchfab.*

The publication of Curioni's models on open platforms such as Sketchfab can be interpreted as a form of distributed digital archive, in which visual documentation, descriptive metadata and provenance information help to ensure the authenticity, traceability and scientific accessibility of the digitised heritage, prolonging its usability and cognitive function over time [9].
Following online publication, the digitisation project was extended to include three-dimensional holographic visualisation using a 3DHologramFan, a device based on high-speed rotating LEDs capable of generating images suspended in space with a depth effect. After modelling and publication, an orbital video was created with the model rotating around its own axis, useful for representing the object in 360 degrees. This video was imported into the SpinDisplay application, software dedicated to managing the contents of the holographic fan, which allows for its synchronisation and projection in real space. In this way, the collection has become not only a digital archive, but also accessible to the public through a visual and interactive installation capable of rendering

the three-dimensional perception of the model and projecting it in holographic form.

4. OUTCOMES, IMPACT AND PROSPECTS (MMB)

The project's methodological elements are characterised by a diversity of skills, ranging from digital representation to diagnostics, documentation to geometric surveying, and information modelling, construction engineering to visual communication [10]. This diversity aligns with the interdisciplinary approach characteristic of the Drawing disciplinar sector, where graphic representation is conceptualised as a scientific, cognitive and communicative language, adept at integrating diverse forms of knowledge into a unified cognitive and representational process. The research has yielded substantial results across multiple domains, establishing a robust and replicable methodological framework for the documentation and enhancement of academic assets [11].

From a scientific perspective, the triangulation between laser scanners, LiDAR and modelling from historical drawings has enabled the comparison of accuracy and survey times, thus establishing a validated protocol that can be transferred to other archives and collections. The incorporation of thermographic investigations has led to the establishment of a novel diagnostic level, thereby demonstrating the potential for drawing to extend to the domain of material and structural knowledge.

From a disciplinary standpoint, the research reinforces the notion of drawing as a dynamic instrument for the acquisition of knowledge and the interpretation of ideas. The dialectic between the digital and the tangible, as evidenced by the transition from modelling to 3D printing, and the reintroduction of wooden models in lessons, demonstrates that representation does not merely substitute for matter; rather, it serves to revitalise its perceptual and cognitive significance. The act of drawing, therefore, becomes a cyclical process: from the real to the digital and back again, in a continuous dialogue between observation and reconstruction.

The experience has resulted in the establishment of an integrated teaching model at educational and cultural levels. This model combines technical knowledge with direct experience of the artefact, thereby facilitating a holistic learning approach. The integration of conventional surveying techniques, digital modelling methodologies and holographic visualisation has been instrumental in cultivating an understanding among students of representation as a critical, measured and interpretative act.

The research on the Curioni Collection, in conjunction with earlier studies of the Mosca Library and the Porcheddu Archive, and ongoing research in the Betta-Bardelli Archive, establishes a cohesive trajectory. In this trajectory, the act of drawing is recognised as a medium for the transmission of knowledge and as an integral component of the disciplinary infrastructure that underpins technical and scientific heritage. Representation, in its complete cycle from the real to the virtual and back again, becomes the locus where science, culture and memory meet, thus restoring drawing to its original function as a tool for understanding and transmitting knowledge. It is evident that, through these experiments, design reaffirms its ability to adapt to contemporary languages of representation, while maintaining its epistemic nature. Drawing, as a discipline, is one that, while transforming its tools, continues to deal with the relationship between space, vision and knowledge.

Holography, in its various forms, thus becomes the contemporary heir to projection and section, not as a substitute for descriptive geometry, but as its luminous and perceptive evolution, capable of translating the theory of drawing into the logic of light and vision.

The digitisation of the Curioni Collection has enabled the development of a methodological model that integrates the precision of surveying, the geometric consistency of modelling, and the communicative power of digital representation. The transition from documentation to interaction — from 3D surveying to holographic projection — marks an evolution of drawing as a cognitive language capable of extending from the represented space to the perceived space (Figure 7).

Representation is no longer regarded as a closed product (as it seems in archives and collections), but rather as a shared process, in which the digital model becomes a node in an information network that can be consulted, commented on and reproduced.

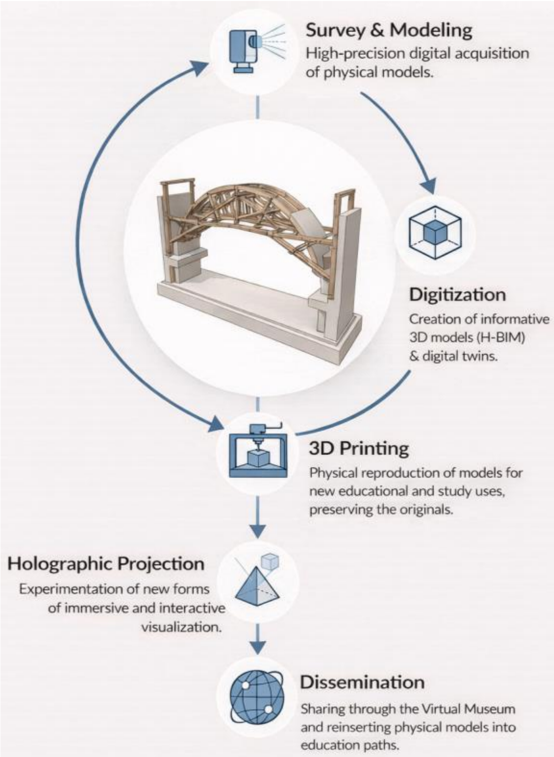

Figure 7*: An operating system for heritage valorization has been developed to manage analog models into modern communications tools. This cyclical process connects tangible memory with digital innovation.*

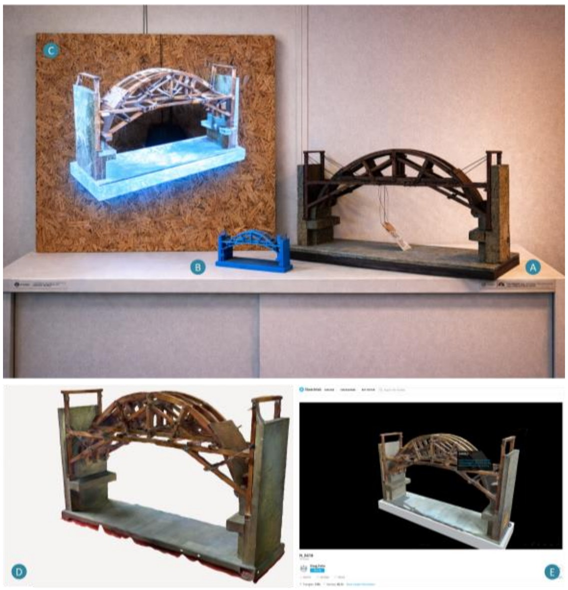

Figure 8: *Example of a model of a complete suit of armour, known as suspended armour, with wooden fittings (Modello di una armatura completa, detta sospesa con ferramenta in legno), 40x76x20 cm, Inv. 47 (1865-1887, DISEG); A –original wooden model; B - 3D printed model; C – holographic model; D – polycam model; E – sketchfab model).*

The utilisation of cost-effective technologies and open platforms, such as Sketchfab, has rendered models accessible and interoperable, thereby translating the FAIR principles into an operational paradigm of openness and traceability (Figure 8).

This finding serves to substantiate the assertion that drawing functions as a conceptual instrument, serving both as a conduit for comprehending form and, concomitantly, as a medium for the edification of knowledge surrounding it. Experimentation with holographic projection has engendered a novel paradigm in the realm of disciplinary reflection in this applied research.

This experimentation serves to restore the original function of drawing as an instrument of integral knowledge, capable of mediating between sensory experience and theoretical construction.

In terms of future prospects, the project offers a replicable model for the construction of dynamic digital archives and interactive virtual museums. In such a scenario, three-dimensional and holographic representation would become a common language of access, study and dissemination.

The Curioni Collection provides a compelling illustration of how the digitisation process, guided by the meticulous delineation of drawings, can not only accurately restore the physical form of models but also their original function as instruments for thought, comprehension and the dissemination of knowledge, in this case pertaining art of construction.

5. ACKNOWLEDGMENT

The working group involved in the project to enhance and digitise the Curioni Collection is composed of a diverse set of skills and disciplinary profiles from different areas of academic research. It brings together figures from the sciences of representation, surveying and construction, creating a cross-sector collaboration that has made it possible to integrate different methods, tools and objectives into a unified research framework:

Scientific responsibility and coordination: Maurizio Marco Bocconcino; coordination (surveying and modelling): Mariapaola Vozzola; coordination (historical research and educational applications): Martino Pavignano; Ursula Zich, for educational aspects; Marco Piras and Paolo Dabove, for geomatics and measurement skills; Professor Mauro Borri Brunetto, for construction sciences; architect Margherita Bongiovanni and Dr Francesca Gervasio, for activities related to the University's cultural and scientific heritage; engineer Nives Grasso and DISEG technician

Pierluigi Guarrera, for support with acquisitions and metric processing; engineers Luca Gioberti, Federica Bonino, Larisa Semis, Tommaso Verdier, Muhammad Daud, José Luis Reyes Mesias, for assistance with surveying and modelling; junior engineers Emanuele Ricchiello, Roberto Cagliero and architect Salvatore Tartaglia, involved in acquisition, information retrieval and experimentation with three-dimensional and holographic models. The group is also collaborating with the thermographic analysis team coordinated by Monica Volinia, in recent cooperation with the CNR in Padua, for non-invasive diagnostic investigations on wooden models.

6. REFERENCES

[1] Bocconcino, Maurizio Marco; Novello, Giuseppa: Itinerari grafici tra le carte della Biblioteca Mosca. FortMed 2018 Proceedings, Messina, 2018.

[2] Bocconcino, Maurizio Marco; Vozzola, Mariapaola; Pavignano, Martino; Garzino, Giorgio: Itinerari digitali tra carte e disegni dell’archivio Porcheddu. AISI 2020 Proceedings, 633–645, 2020.

[3] Bocconcino, Maurizio Marco; Vozzola, Mariapaola; Pavignano, Martino: Drawing as witness: The Betta-Bardelli Archive between historical memory and digital innovation. DisegnareCon, vol. 18, no. 34, pp. 1–8, 2025.

[4] Bocconcino, Maurizio Marco; Vozzola, Mariapaola; Pavignano, Martino: Giovanni Curioni’s Digital Museum (1/2): Comparative Survey Techniques. ISPRS Archives, vol. XLVIII-4, pp. 85–92, 2023.

[5] Bocconcino, Maurizio Marco; Vozzola, Mariapaola; Pavignano, Martino: Giovanni Curioni’s Digital Museum (2/2): Possible Strategies for Access and Inclusion. ISPRS Archives, vol. XLVIII-4, pp. 93–100, 2024.

[6] Bocconcino, Maurizio Marco; Vozzola, Mariapaola; Pavignano, Martino; Gioberti, Luca: Giovanni Curioni’s Digital Museum (2/2): Possible Strategies for a Data Management Plan. 29th CIPA Symposium “Documenting, Understanding, Preserving Cultural Heritage: Humanities and Digital Technologies for Shaping the Future”, Firenze, 2023.

[7] Newell, J.: Old objects, new media: Historical collections, digitization and affect. Journal of Material Culture, vol. 17, no. 3, 2012.

[8] Musicco, A.; Rossi, N.; Verdoscia, C.: Valutazione dell'accuratezza della videogrammetria basata su smartphone per il processo di documentazione del patrimonio culturale. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2023.

[9] Koller, D.; Frischer, B.; Humphreys, G.: Research challenges for digital archives of 3D cultural heritage models. Journal on Computing and Cultural Heritage, vol. 2, no. 3, 2009.

[10] Bocconcino, Maurizio Marco; Pavignano, Martino; Vozzola, Mariapaola: BIM Model Reconstruction of Architectural Archives: A Case Study in Turin, Italy. ISPRS Archives, vol. XLVIII-M-9, pp. 157–164, 2025.

[11] Bocconcino, Maurizio Marco; Cumino, Caterina; Pavignano, Martino; Vozzola, Mariapaola; Zich, Ursula: Il patrimonio accademico nella triplice missione universitaria: esperimenti di accessibilità e inclusione for all. Atti del Convegno DAI 2025, Alghero, 2025.