remote sensing

MDPI

Article

A Knowledge-Grounded AI Engine for Digital Twins: Supporting Conservation Through Long-Term Monitoring

Luigi Barazzetti [1,*], Rasoul Eskandari [1], Abdullah Kadah [1], Mattia Previtali [1] and Fabio Roncoroni [2]

[1] Department of Architecture, Built Environment and Construction Engineering (ABC), Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy; rasoul.eskandari@polimi.it (R.E.); abdullah.kadah@mail.polimi.it (A.K.); mattia.previtali@polimi.it (M.P.)
[2] Polo Territoriali di Lecco, Politecnico di Milano, Via Previati 1/c, 23900 Lecco, Italy; fabio.roncoroni@polimi.it
* Correspondence: luigi.barazzetti@polimi.it

Highlights

What are the main findings?

- A knowledge-grounded AI engine integrates monitoring data, technical reports, and expert knowledge to support the interpretation of structural behaviour.
- The integration of ground-based and satellite monitoring within a 3D Digital Twin enables context-aware analysis of structural behaviour and deformation trends.

What are the implications of the main findings?

- Digital Twins for complex buildings can evolve into knowledge-driven systems supporting structural assessment and conservation decisions.
- AI can make accumulated engineering expertise more accessible, supporting evidence-based interpretation while preserving the essential role of human specialists.

Abstract

This paper presents a knowledge-grounded Digital Twin framework designed to support the interpretation of long-term structural monitoring data. The approach combines AI reasoning models, project-specific knowledge retrieval, numerical observations, and spatial information within an interactive environment where specialists can explore monitoring data and formulate analytical requests. The framework is demonstrated through the Milan Cathedral case study, using ground-based measurements, satellite interferometry data, and a selected subset of technical reports covering more than two decades. A Cesium-based 3D environment, including gaussian splatting visualization, provides spatial access to the monitoring information. The application illustrates how specialists can use AI to retrieve relevant knowledge, guide computational analyses, and examine relationships between monitoring observations and previous engineering interpretations. New forms of interaction among specialists, AI, and monitoring information are illustrated and discussed, highlighting the complementary roles of accumulated knowledge, AI-supported analysis, and human expertise in the interpretation process.

Keywords: geospatial visualization; Cesium; Digital Twin; knowledge-grounded AI; gaussian splatting; structural monitoring

Check for updates

Academic Editor: Massimiliano Pepe

Received: 28 September 2026
Revised: 13 November 2026
Accepted: 15 November 2026
Published: 18 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. From Digital Twins to Knowledge-Driven Heritage Conservation

Digital twins have already demonstrated their potential to improve efficiency and support decision-making across a wide range of industries, including manufacturing, energy, automotive, healthcare, and smart cities [1–5]. Originally conceived as a key enabling technology for smart manufacturing and Industry 4.0, the digital twin has evolved over the past fifteen years into a versatile approach in which a virtual replica of a physical asset exchanges data with its real-world counterpart in real time [6]. This flow of information makes the technology a powerful driver of data-driven decision-making, predictive maintenance, and lifecycle management [7].

Digital twins have achieved significant success in industrial sectors, largely because they are typically developed for well-defined systems characterized by known geometry, specific material properties, established operating conditions, and predictable physical behavior. By integrating continuous sensor streams directly with mathematical models, these frameworks enable real-time monitoring, simulation, optimization, and predictive maintenance [8,9]. Consequently, the primary objective in these contexts is to maintain an accurate digital representation of the physical asset to facilitate informed operational decision-making [10]. Furthermore, these frameworks excel in industrial applications by enabling a continuous flow of information that facilitates proactive fault detection and real-time operational optimization [11,12]. By synthesizing high-frequency sensor data with high-fidelity physics-based simulations, digital twins allow operators to perform sophisticated analyses, simulate complex operational scenarios, and optimize lifecycle performance, which collectively reduce costly downtime and significantly enhance overall system reliability within well-defined, predictable environments [13].

However, cultural heritage challenges the traditional concept of a digital twin. Unlike industrial systems, where a digital twin often represents an asset with well-defined behavior, heritage assets are inherently complex, evolving, irregular, and stratified, and require multidisciplinary approaches [14–16]. Its conservation depends not only on geometric or sensor data, but also on historical documentation, material and damage characterization, structural monitoring, expert interpretation, and decades of accumulated experience [17–19]. Heritage buildings and sites are the result of continuous transformations that have occurred over time. They often have complex, irregular geometries, consist of heterogeneous materials, and are affected by various forms of deterioration and structural damage. Consequently, specialists often work with an incomplete understanding of their geometry, materials, construction history, and current condition, unlike in more standardized industries, where comprehensive and structured information about an asset is typically available. In this context, a digital twin can be more than a pure digital representation of a heritage asset. It becomes a framework that transforms heterogeneous data into knowledge, knowledge into understanding, and understanding into informed action (Figure 1). The true value of a digital twin for heritage therefore lies not in its ability to replicate reality, but in its ability to support a deeper understanding of reality and to enable more informed conservation and management decisions.

Today, Artificial Intelligence (AI) is expanding this paradigm, enabling new ways for humans to access, interact with, reason about, and learn from complex knowledge systems. For instance, advanced machine learning algorithms, when integrated with Internet of Things (IoT) sensors, have fundamentally shifted cultural heritage preservation from reactive, periodic inspections to proactive monitoring models [20]. This paradigm shift is underpinned by the implementation of Digital Twins, which synthesize real-time sensor data with semantic heritage models to simulate complex degradation patterns and structural behaviors [16,21]. By using these interconnected systems, stakeholders can execute

predictive maintenance strategies that mitigate decay while simultaneously optimizing the decision-making processes for conservation interventions. Furthermore, the integration of these systems can facilitate a multidisciplinary approach that enables the simultaneous regulation of structural and environmental conditions to support long-term conservation strategies [22]. Such integrated systems allow conservationists to predict the long-term impact on material integrity, thereby optimizing the scheduling of interventions [16,23]. Beyond mere data aggregation, researchers now envision the integration of ontological models to structure this information, effectively minimizing AI-induced errors and enhancing the transparency of automated decision-making processes [24].

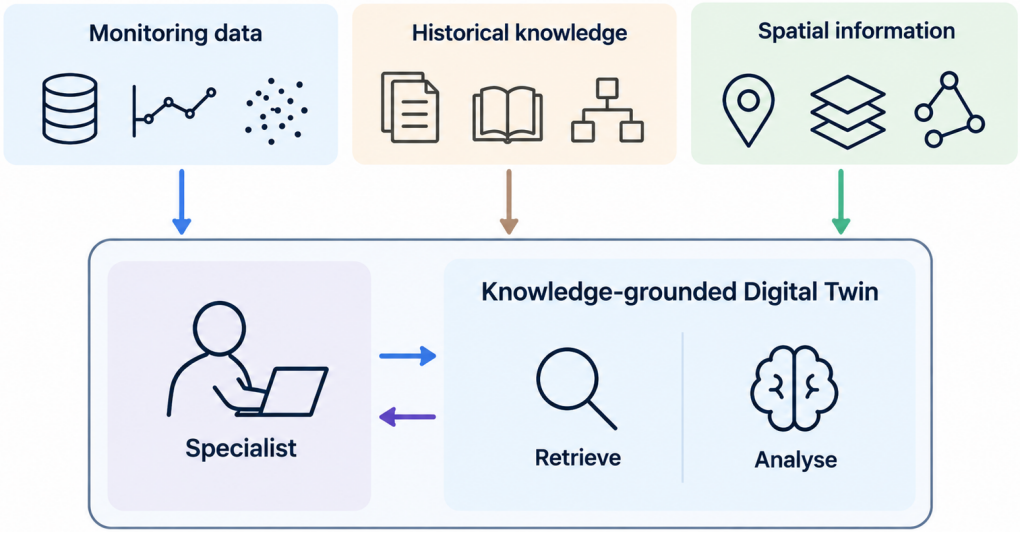


Figure 1. Conceptual overview of the knowledge-grounded Digital Twin, connecting monitoring data, historical knowledge, and spatial information with expert-guided retrieval and analysis.

Given the multifaceted nature of cultural heritage, no universal digital twin exists. Instead, a heritage-oriented digital twin functions as a dynamic knowledge framework, integrating diverse perspectives tailored to specific conservation objectives. In this work, we focus on one of these perspectives: structural monitoring. We investigate how decades of monitoring data can be transformed into actionable knowledge through digital twins and artificial intelligence, supporting the interpretation of structural behaviour and enabling more informed conservation decisions. This approach uses multimodal data fusion to move beyond simple threshold alarms. By deploying predictive models, these systems can distinguish between normal environmental fluctuations and critical structural anomalies that necessitate intervention [25].

Various sensors are now available and can be installed to capture real-time information on a wide array of structural and environmental parameters, incorporating devices such as accelerometers for dynamic vibration analysis, inclinometers for tilt measurement, and linear variable differential transformers for high-precision tracking of crack propagation and structural settlement [26–28]. These sensing networks, when coupled with high-accuracy recording and modeling techniques, facilitate the continuous monitoring of structural and environmental variables. Additional data can be gathered by instruments and tools such as robotic total stations, GNSS receivers, levels, and other advanced monitoring equipment [29–31]. Movement data can also be acquired through InSAR, which offers remote, non-contact sensing capabilities to detect millimeter-scale displacements over extensive architectural surfaces or historical buildings [32–37].

Thus, useful data can be collected both by sensors permanently installed on the structure and through periodic monitoring campaigns. These heterogeneous datasets, ranging

from high-frequency vibrations to long-term displacements, necessitate computational procedures to effectively correlate environmental stressors with structural responses [38]. Generative artificial intelligence is becoming increasingly relevant for converting raw data streams into useful information [39]. Indeed, although acquiring good data is crucial, transforming such data into useful information to support long-term conservation activities is more complicated, often requiring expert users to bridge the gap between technical data acquisition and interpretative diagnostics. Recent developments in generative AI offer sophisticated computational frameworks that could enhance how human operators interact with, interpret, and derive actionable insights from intricate datasets.

1.2. Bridging Artificial Intelligence and Human Expertise

Working with heritage buildings and sites requires active engagement with both the monuments and their surrounding environment, together with an ongoing process of building knowledge to support informed decision-making based on reliable evidence. Nevertheless, decisions cannot always be driven solely by quantitative data, as they may also involve qualitative assessments, expert judgment, and a degree of subjectivity.

The conservation and monitoring of buildings involve several interconnected activities, ranging from the acquisition and analysis of measurements to the examination of existing documentation, on-site inspections, and the interpretation of structural behaviour. Understanding a complex building requires more than quantitative observations. It also depends on direct perception, knowledge of construction techniques and materials, awareness of previous interventions, and experience gained through long-term interaction with the structure.

These different dimensions of knowledge are closely connected in professional practice. Specialists continuously combine measurements, documentary evidence, field observations, and engineering judgment to develop and refine their understanding of the building. Artificial intelligence can support this process by making large amounts of information more accessible, identifying relationships between different sources, and assisting with analytical tasks. However, the interpretation of results still requires contextual knowledge and critical assessment.

The proposed system was developed to support this interaction between human expertise and the growing amount of information available for structural monitoring. Through a natural-language interface, specialists can access monitoring observations, historical reports, and technical documentation, formulate analytical questions, and explore relationships between different sources of evidence. The aim is to make this information easier to retrieve and interpret while keeping specialists actively involved in the analysis and assessment process.

The proposed system is a knowledge-grounded generative AI engine designed to support the interpretation of structural monitoring data. The proposed framework is grounded in a curated repository of domain-specific information, including historical monitoring reports, scientific publications, inspection records, and technical documentation. Through the integration of retrieval mechanisms with a large language model (LLM), the engine generates responses that are informed by the available engineering evidence, enabling consistent interpretation of structural behavior, historical trends, and monitoring results. As the knowledge repository expands with new reports and observations, the system can draw on a broader context without changes to the underlying generative AI model.

Beyond supporting information retrieval, the proposed system aims to assist the reasoning process. In structural monitoring, identifying an anomaly is only the first step; understanding why it occurred is essential. Effective decision-making therefore requires not only detecting these effects but also formulating plausible explanations of their causes

by considering historical behavior, environmental conditions, construction details, previous interventions, and expert knowledge. Explaining the potential causes of an anomaly enables specialists to address the origin of the problem rather than merely reacting to its symptoms.

The main contribution of this study is the exploration of a knowledge-grounded digital twin in which structural monitoring observations, existing documentation, and spatial information are brought together to support interpretation. The proposed approach goes beyond data visualization and document retrieval by allowing specialists to formulate questions, introduce constraints based on their experience, and examine the results in relation to the available evidence. Unlike approaches primarily centered on data visualization, automated monitoring, or AI-based prediction, the proposed framework emphasizes the interaction between spatially selected monitoring observations, accumulated engineering knowledge, and expert-guided interpretation. This combination represents the specific contribution investigated in the present study.

A central aspect is the interaction between the knowledge accessible to the AI system and the expertise of the user. The framework is not intended to replace engineering judgment or provide autonomous structural diagnosis, but to explore how existing AI reasoning capabilities can support specialists in understanding complex monitoring information.

2. Materials and Methods

2.1. Overview of the Proposed Knowledge-Grounded AI Engine

This study introduces a prototype digital platform that uses AI and 3D geospatial information to support the structural monitoring of complex architectural heritage. While Milan Cathedral is presented here as the primary case study, the platform is concurrently being explored through other experimental implementations. The key idea is an AI system that integrates semantic reasoning to synthesize multi-modal sensor data and historical documentation, thereby bridging the gap between raw diagnostic outputs and structured knowledge representation. The workflow brings together monitoring observations, existing documentation, and spatial information within a common environment, allowing specialists to explore their relationships and interpret structural behaviour through AI-supported interaction.

Figure 2 illustrates the overall architecture of the proposed knowledge-grounded analysis system for structural monitoring. The system enables two complementary modes of interaction: natural-language prompts, through which users can ask questions, compare monitoring campaigns, request summaries, or explore trends, and 3D model navigation, allowing users to select specific structural elements or spatial regions directly within the digital model. Both interaction modalities are integrated into a central analysis system that combines heterogeneous information sources.

A key feature of the proposed system is that its capabilities are derived not only from numerical monitoring data but also from more than two decades of monitoring reports. The Cathedral's monitoring archive extends over more than 50 years, as documented in [40]. However, the knowledge repository used in the present study includes only a selected subset of technical reports spanning more than two decades. These reports capture the accumulated expertise of specialists, including observations, interpretations, hypotheses, and explanations of the structural behavior over time. Rather than simply identifying anomalies, specialists describe their likely causes, discuss the influence of environmental conditions, construction history, maintenance interventions, and previous events, and justify their conclusions. By incorporating this expert knowledge into the AI engine, the system can support the contextual interpretation of monitoring observations. In this way, the accu-

mulated experience of experts becomes an integral part of the decision-support process, allowing future users to benefit from decades of documented knowledge.

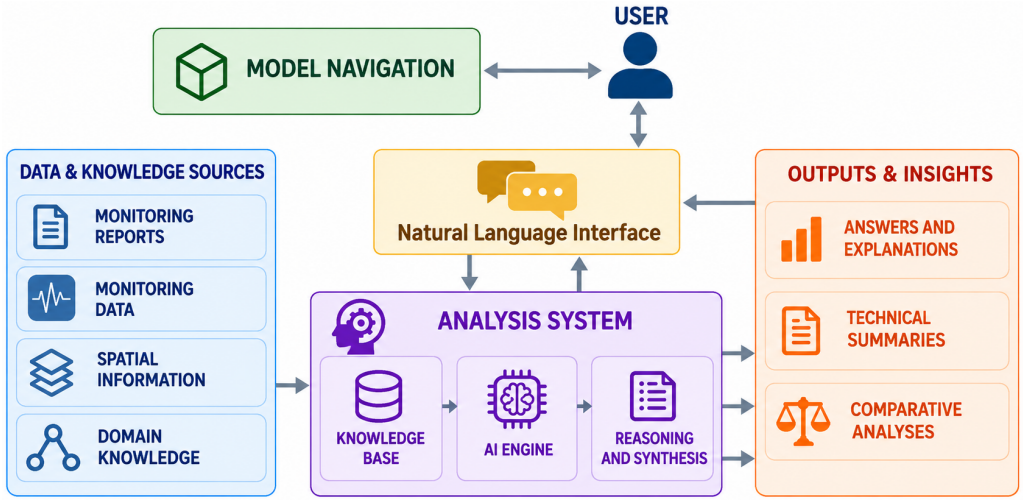


Figure 2. Conceptual configuration of the proposed framework, illustrating some of the operations that can be performed using the proposed approach.

The historical reports represent more than documents: they record the accumulated reasoning of specialists. The reports are merely the vehicle through which expertise has been recorded over more than two decades. Their real value lies in the embedded knowledge: interpretations of structural behavior, explanations of anomalies, hypotheses, and the rationale behind inspection and conservation decisions.

The knowledge-grounded framework combines three complementary functions: knowledge retrieval from historical reports and technical documentation, AI-supported exploration and analysis of monitoring data, and reasoning and synthesis to assist specialists in interpreting structural behaviour. The results are presented through interactive visualizations, technical summaries, comparative analyses, and context-aware explanations based on the available evidence. The system also supports an iterative, user-in-the-loop interaction, allowing specialists to refine their queries, introduce additional constraints, and critically assess the generated results.

An essential aspect of the proposed framework is the integration of monitoring data with the spatial and structural representation of the asset. Sensor locations and measurement points should not be treated as isolated identifiers but linked to their actual position within the structure, the corresponding structural elements, and a common georeferenced coordinate system. Figure 3 presents the graphical user interface (GUI), which enables interactive exploration through both the 3D visualization and a natural language prompt window.

The spatial context enables the engine to associate observations with the physical asset, correlate measurements acquired from different monitoring techniques, and reason about the relationships between neighbouring elements, structural subsystems, and environmental conditions. By combining semantic knowledge with georeferenced structural information, the system provides a spatially contextualized interpretation of monitoring data, supporting both natural language interaction and direct exploration through the digital model.



Figure 3. The GUI enabling interactive exploration through both the 3D visualization and a natural-language prompt window (in red).

2.2. Data Description: From Space All the Way Down to the Ground

To validate this proposed approach, we implement the framework as a proof-of-concept prototype applied to the Cathedral of Milan, a complex historical asset requiring continuous monitoring and interventions. It is beyond the scope of this paper to provide a detailed description of the monitoring activities in the Cathedral, which requires continuous maintenance and restoration activities [41]. Instead, this case study serves as a pilot application to demonstrate the system's capacity to integrate 3D data, diagnostic/monitoring data and historical records into a unified and interactive knowledge-driven environment.

This section introduces the datasets used to develop the proposed framework. The available information originates from both ground-based monitoring systems and satellite observations, providing complementary sources of information. The monitoring system at the ground level of the Milan Cathedral is described in detail in [40], which provides an overview of the long-term structural monitoring program, the measurement techniques, and the archive of observations. More recently, the monitoring system has been complemented by satellite-based measurements using Persistent Scatterer Interferometry (PSI) [42].

The satellite-based deformation dataset was generated from a stack of 96 COSMO-SkyMed (CSK) SAR images acquired along a descending orbit between 5 May 2019 and 20 November 2025. A small scene of approximately 700 m × 400 m, centered on the Milan

Cathedral, was extracted to restrict the analysis to the area of interest. The SAR images were co-registered using the ESA SNAP Graph Processing Framework (GPF). A single-master interferometric network configuration was then established using the acquisition of 29 March 2023 as the reference epoch, with all remaining acquisitions as slave images, forming interferometric connections with this master image. This configuration was adopted considering the predominantly urban character of the study area, where permanent scatterers are expected to maintain phase coherence even over relatively long temporal baselines. Furthermore, the single-master scheme retains a common phase reference across the interferometric stack for the subsequent multi-temporal parameter estimation. The resulting interferograms were corrected for both flat-Earth and topographic phase contributions using precise orbit information and the SRTM 1 arcsec digital elevation model (DEM). A high-quality pixel characterized by a main-lobe scattering response, high spatial coherence, high amplitude stability (i.e., low amplitude dispersion) was selected as the reference point (RP). Subsequently, candidate pixels (CPs) were identified applying a combined quality criterion: pixels for which the sum of spatial coherence (ranging on [0, 1]) and amplitude stability index (ranging on [−Inf, 1]) was greater than or equal to 0.85.

The processing was formulated within a differential multi-temporal InSAR framework. Considering the limited spatial extent of the investigated scene, atmospheric phase delays were assumed to be spatially correlated and approximately uniform over the area with respect to the RP, adopting a small-area perspective. For each CP, the unknown parameters were represented by: (i) residual height, linear deformation velocity, and thermal response. Their corresponding phase contributions were related respectively to the perpendicular baseline (derived from the satellite orbit geometry), the temporal baseline (determined from the acquisition dates), and a thermal baseline (constructed from external mean daily temperature observations). Following parameter estimation, only CPs exhibiting a temporal coherence $\gamma \geq 0.6$ were retained as successfully processed point (SPP) and used in the subsequent analyses. It should be noted that the relatively long temporal baselines associated with some master-slave pairs within the adopted interferometric configuration may increase temporal decorrelation, particularly for PSs exhibiting changes in their backscattering characteristics. However, this limitation was addressed through the combined initial quality-based selection of CPs and the subsequent temporal coherence threshold.

Subsequently, the deformation time series of each accepted SPP was reconstructed by combining the modeled deformation components (i.e., temporal and thermal responses) with the residual phase contribution remaining after parameter estimation. Then, the SPPs were geocoded by refining the initial SRTM elevation with the estimated residual height and transforming the resulting radar coordinates into geographic coordinates through Doppler-Range-Ellipsoid (DRE) equations.

The large volume and heterogeneous nature of the collected data do not directly provide actionable knowledge. These data must be processed, integrated, and interpreted to extract meaningful information about the structure's condition and behavior. This transformation from raw measurements into useful information is a fundamental step in supporting predictive analyses, informed conservation, and maintenance decisions within a Digital Twin framework. The proposed knowledge-grounded AI system was developed with the aim of integrating newly available satellite observations with traditional ground-based sensors, which have provided—and continue to provide—monitoring data over several years. The objective is to progressively incorporate an increasing variety and volume of data within a coherent and reasoned framework, rather than simply accumulating additional raw measurements. Indeed, collecting large amounts of data is fundamentally different from extracting meaningful information from them. The value of the proposed approach therefore lies not in the quantity of data that can be incorporated, but in its

ability to contextualize, correlate, and interpret heterogeneous observations to support the generation of useful knowledge for long-term monitoring and conservation.

2.3. Monitoring Data Validation Prior to Knowledge-Grounded AI Integration

Before being used to support analyses within a knowledge-grounded AI system, monitoring data must undergo a validation process to ensure their quality, reliability, and consistency. The performance of AI-based systems strongly depends on the quality of the input data, making data validation a fundamental step in a Digital Twin workflow. In the present study, the ground (in situ) monitoring data originate from a long-established monitoring system that has been continuously maintained and verified over several years as part of the Cathedral's routine monitoring activities. Their quality and reliability have therefore already been assessed through operational use.

Conversely, the persistent scatterer interferometry data represent a new source of information introduced within this project. For this reason, a dedicated validation procedure has been carried out before integrating them into the platform. The validation has two main objectives. First, it ensures that the PSI measurements satisfy the quality requirements needed for reliable AI-based analyses and subsequent structural interpretation. Second, it provides a validated observational basis for the subsequent AI-supported analyses, allowing the system to operate on monitoring information whose consistency has been independently assessed.

For the sake of clarity, this paper presents the validation process through an illustrative case study limited to a specific area of the Cathedral. A dedicated analysis was performed on a subset of 23 persistent scatterers extracted from the descending COSMO-SkyMed dataset within the monitored façade sector of the Milan Cathedral. The selected scatterers were identified on the basis of their spatial distribution and their location within the portion of the structure covered by an independent ground-based monitoring system. The objective of this selection was to investigate the deformation behaviour of a representative façade-scale area rather than relying on the displacement history of a single target.

Individual persistent scatterers may be affected by local conditions. By considering a group of spatially distributed scatterers, these local effects can be reduced and the resulting signal becomes more representative of the actual behaviour of the monitored sector. Consequently, the analysis focuses on the common component shared by multiple targets rather than on the potentially anomalous behaviour of an individual scatterer.

To investigate the common deformation signal, three aggregate time series were computed, namely the arithmetic mean of the selected persistent scatterers, the median displacement series, and a weighted mean displacement series. The arithmetic mean provides an estimate of the average displacement behaviour of the monitored area, while the median series reduces the influence of isolated outliers.

The comparison between the three aggregate series revealed a high degree of consistency, indicating that no individual scatterer dominates the overall response. This result is particularly important because it demonstrates that the deformation pattern observed in the data is spatially coherent across the monitored façade sector. The similarity between the mean, median and weighted-average series suggests that the selected scatterers capture a common displacement behaviour and that the aggregate signal is robust with respect to the specific averaging strategy adopted.

To investigate the temporal characteristics of the observations, a harmonic model was fitted to the aggregate time series. The adopted model combines a long-term component with a seasonal oscillation. The amplitude of the seasonal oscillation was computed from the fitted harmonic coefficients and used to quantify the magnitude of the recurring deformation pattern. To further isolate the oscillatory component of the signal, a detrended

version of the selected dataset was generated. In this procedure, the linear trend estimated for each individual persistent Scatterer was removed prior to the aggregation step. This approach preserves the seasonal behaviour while eliminating long-term displacement tendencies, thereby facilitating the comparison with the independent monitoring measurements. The detrended aggregate series therefore represents the average seasonal response of the monitored façade sector.

The resulting time series were subsequently compared with the independent monitoring measurements acquired at Points P33, P34, P35, and P36 (Figure 4), which, according to the traditional naming convention, correspond to the central columns of the façade. These monitoring points are located within the same structural sector represented by the selected scatterers and therefore provide an independent reference for evaluating the deformation behaviour identified through the satellite observations.

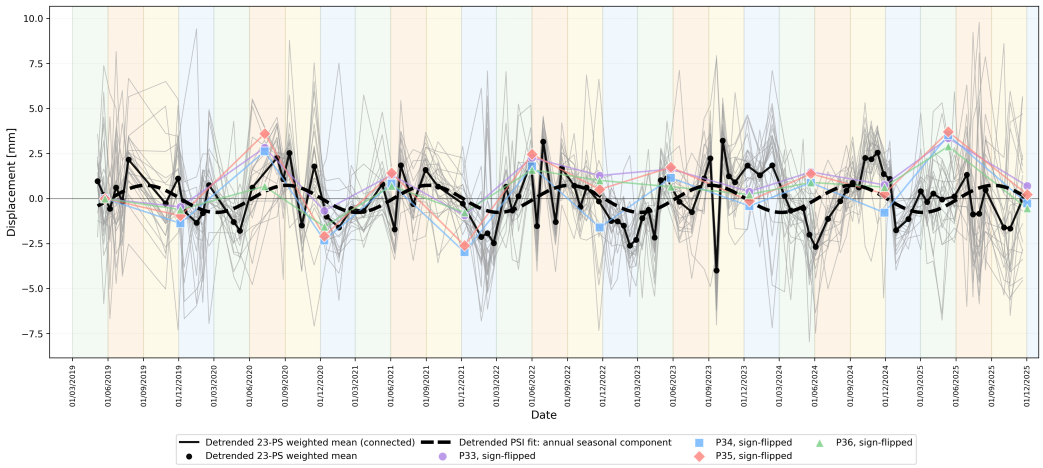


Figure 4. Comparison between satellite time series and independent monitoring measurements at Points P33, P34, P35 and P36.

The geometry of the satellite measurements deserves particular consideration. Interferometric SAR does not directly measure horizontal or vertical displacements. For the selected COSMO-SkyMed dataset, the average incidence angle is approximately 26.75°. For the present analysis, it was assumed that the monitored portion of the façade experiences predominantly East–West movements, while vertical and North–South components are negligible. This assumption is physically plausible because the monitored façade is approximately aligned in such a way that thermal expansion and contraction are expected to generate mainly horizontal displacements. This hypothesis is further supported by independent in situ monitoring data based on high precision leveling. The North–South inclination measurements recorded by ground-based sensors exhibit only small, periodic variations, with the annual cycle closing consistently over time, indicating the absence of cumulative deformation. Likewise, the vertical displacements measured by the leveling network are very limited, confirming the overall stability of the monitored area.

Consequently, the LOS displacement measured by PSI can be interpreted as a scaled representation of the East–West motion of the façade. The comparison between the two datasets reveals several common characteristics. First, both the ground measurements and the PSI observations indicate displacement amplitudes of only a few millimeters. The ground-level measurements show a total variation of approximately 5 mm over the monitoring period, with average values ranging from about +1.8 mm to approximately −3.4 mm. These amplitudes are more than one order of magnitude larger than the instrumental precision.

Second, the ground measurements exhibit a clear cyclic behavior. Positive and negative phases alternate throughout the observation period, producing oscillations with typical seasonal amplitudes. The recurrence of this pattern over several years strongly suggests the presence of a seasonal component. Because all four monitored points respond coherently, the seasonal signal can be interpreted as a global response of the monitored portion of the facade. The time series shows a larger scatter than the in-situ observations, which is expected given the intrinsic uncertainty associated with interferometric measurements and the complexity of the monitored architectural surface. Nevertheless, despite the higher noise level, the satellite observations also display alternating positive and negative phases. When considered at a seasonal or annual timescale, the measurements exhibit a temporal behaviour that is consistent with the cyclic pattern observed in the ground data.

The agreement between the two independent monitoring techniques is particularly significant because the measurements are based on completely different physical principles. In-situ observations provide direct geometric measurements with sub-millimeter precision, whereas PSI measurements are derived from data acquired from space. The fact that both datasets indicate recurrent millimeter-level oscillations supports the interpretation that the detected movements are real and not artefacts of a particular measurement technique.

From a physical perspective, the observed behavior is consistent with environmentally driven deformation. The upper façade is continuously exposed to seasonal variations. Thermal expansion and contraction of the facade elements and supporting structural components can therefore generate reversible East–West displacements of a few millimeters. In contrast, the evidence for a significant long-term trend is comparatively weak. Although some differences between individual years are visible, these variations remain small compared with the recurring annual cycle. Therefore, the seasonal component should be regarded as the principal deformation signal observed in the monitored portion of the façade.

Overall, the comparison demonstrates that the in-situ data and the satellite observations provide mutually consistent information. The ground-level measurements precisely quantify the amplitude of the seasonal movements, while the satellite data independently confirm the existence of the same cyclic behavior over a larger set of observation points. The results indicate that the upper façade is primarily affected by reversible seasonal East–West movements of a few millimeters, whereas evidence for a persistent long-term deformation remains limited.

3. System Design and Implementation

The main outcome of the proposed methodology is an experimental knowledge-grounded Digital Twin that brings together monitoring observations, existing documentation, and spatial information within an interactive environment. The following sections present the developed platform, its main functions, and the integration of the different sources of information. Particular attention is given to how these components support the exploration of monitoring data and interaction between specialists and AI.

3.1. System Architecture and Geospatial 3D Data Integration

The proposed platform was developed as a web-based prototype integrating 3D geospatial visualization, multi-source monitoring data, and knowledge-grounded AI interaction within a unified digital-twin environment (Figure 5). The system follows a client–server architecture. The client side is responsible for the interactive visualization of the monitored area, the rendering of geospatial and 3D layers, the selection of monitoring points, the generation of graphical outputs, and the interaction with the natural-language assistant. The server side manages application configuration, access to AI services, commu-

nication with the vector store, document ingestion, and the protected operator interface used to update the project knowledge base.

The frontend is based on CesiumJS and provides an interactive geospatial environment in which heterogeneous monitoring information can be visualized in its spatial context. The digital representation of the monitored area combines an orthophoto or map background, georeferenced PSI points, OpenStreetMap-derived contextual information, and a 3D visualization of the monument based on gaussian splatting (GS), which is loaded as a 3D asset. Particular attention was given to the alignment of the splat model, since the geodetic reference of the SAR data, the terrain model, and the 3D reconstruction should coincide. These datasets act as the spatial backbone of the system, allowing the monitored asset to be interpreted not as an isolated object, but as part of its surrounding urban environment. This is particularly important for the Milan Cathedral, where the interpretation of monitoring data is influenced by the relationship between the monument, the square, adjacent buildings, and the local urban morphology.

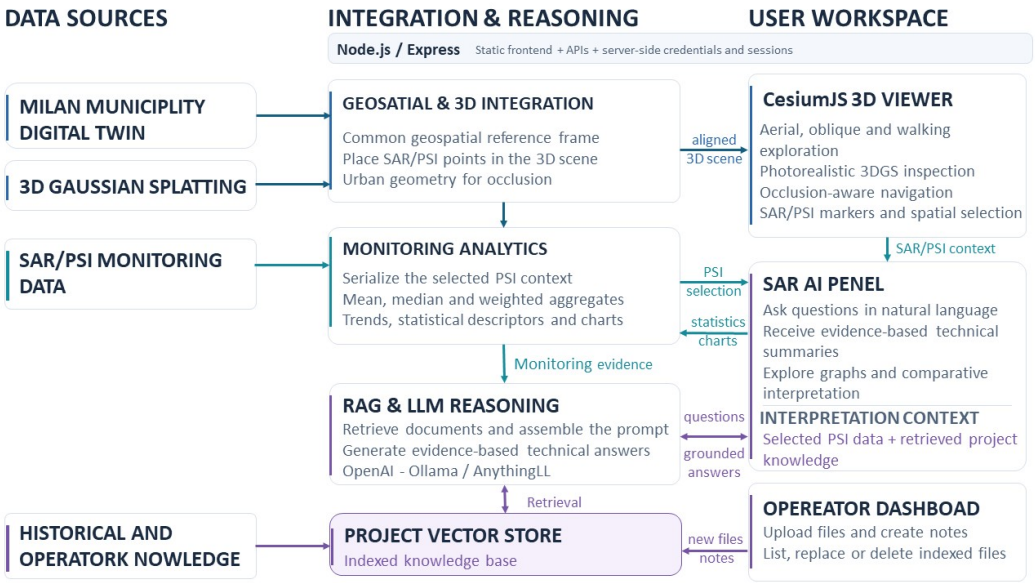


Figure 5. Architecture of the developed platform: Data Sources, Backbone and Integration Layer and User Interaction.

The gaussian splatting representations used in the experimental platform were obtained through two complementary acquisition strategies. The first used a SHARE S20 SLAM system operating in RTK mode, providing directly georeferenced GS data. The second used an XGRIDS PortalCam. Its reconstruction was subsequently aligned to the SLAM dataset by a three-dimensional transformation (translation, rotation, and scale), based on manually identified corresponding points. Residual discrepancies of several centimeters were observed in exterior areas after this transformation. These are observed transformation residuals, not an independently established georeferencing accuracy or a registration RMSE.

In the present prototype, the gaussian splatting representation is used as a photorealistic interface that facilitates navigation, visual recognition of building elements, and rapid access to associated monitoring information without requiring a detailed geometric model. It is not used to obtain structural displacements or other additional structural measurements. The observed alignment discrepancies were therefore considered acceptable for this demonstrative visualization purpose, but the model should not be used for precise metric assessment.

The monitoring layer is structured around persistent scatterer interferometry measurements, which are represented as georeferenced entities in the Cesium scene. Each scatterer is associated with metadata, including its identifier, spatial coordinates, mean velocity, coherence or quality indicators when available, and displacement time series. The viewer supports both single-point inspection and group-based selection. In the latter mode, users can define a spatial subset of PSI points directly in the 3D/map interface, for instance by selecting a façade sector or a portion of the Cathedral square. Once selected, the platform computes aggregate time series from the selected scatterers, including arithmetic mean, median, and coherence-weighted mean where applicable. These aggregate signals are then visualized as interactive plots and can be exported or passed to the AI component as contextual evidence. A specific design choice was to avoid treating monitoring points as isolated records. Instead, each selected point or group of points is interpreted as a spatially embedded observation. This allows the platform to associate the numerical signal with the physical position of the scatterers, the corresponding architectural sector, and the surrounding built environment. Such spatial contextualization is particularly relevant for PSI data, because the interpretation of line-of-sight displacement measurements depends on acquisition geometry, façade orientation, visibility conditions, and possible occlusions caused by adjacent buildings.

The backend is implemented as a Node.js/Express application. It serves the Cesium-based frontend, exposes runtime configuration endpoints, handles the AI requests, and manages the operator-side functions for document upload and vector-store maintenance. Sensitive parameters are stored in an environment configuration file rather than being hard-coded in the source code. These include the AI API key, the vector-store identifier, operator credentials, Cesium Ion token, 3D asset identifiers, and default visualization parameters. This architecture allows the same software prototype to be deployed with different datasets, credentials, asset identifiers, and monitoring configurations without modifying the application source code.

The resulting architecture separates three complementary levels of the digital-twin platform. The first level is the geospatial visualization layer, which provides the urban and architectural spatial context. The second level is the monitoring layer, which contains PSI observations, time series, attributes, and derived indicators. The third level is the knowledge-grounded AI layer, which connects the current spatial and analytical context to a repository of technical documents and historical monitoring knowledge. The integration of these levels transforms the 3D scene from a passive visualization environment into an operational interface for structural monitoring, data exploration, and evidence-based interpretation.

3.2. Multimodal Visualization and PSI Data Exploration

The platform supports multiple visualization and interaction modes in order to address different analytical needs (Figure 6). The first mode is a standard geospatial view, in which the user explores the monitored area from an aerial or oblique perspective. This mode is useful for understanding the overall spatial distribution of PSI points, their relationship with the Cathedral, and their position with respect to the surrounding urban context. From this view, the user can identify clusters of scatterers located on the monument, in the square, or on neighbouring buildings, and can evaluate whether the spatial distribution of measurements is sufficient for a façade-scale or sector-scale interpretation.

A second visualization mode emphasizes the photorealistic 3D gaussian splatting reconstruction. In this mode, the GS model becomes the primary visual reference, allowing users to inspect the monument with a level of visual detail closer to direct field observation. The user can relate monitoring points to architectural elements, façade portions, towers,

spires, or other structural components. This is particularly useful when the objective is not only to visualize a displacement value, but to understand where that value is located on the physical asset. The GS representation therefore supports a more intuitive interpretation of the monitoring data by linking numerical observations to visually recognizable portions of the structure.

A third mode provides a pedestrian or walking-navigation perspective. Instead of navigating only from an aerial point of view, the user can move through the scene from a street-level viewpoint. This mode is intended to approximate the experience of field inspection and is useful for evaluating the relationship between the monument, the square, the surrounding buildings, and the visible distribution of monitoring points. In this configuration, the user can inspect the monument from viewpoints that are closer to those used during on-site surveys, while still maintaining access to the georeferenced monitoring information and AI-based interpretation tools.

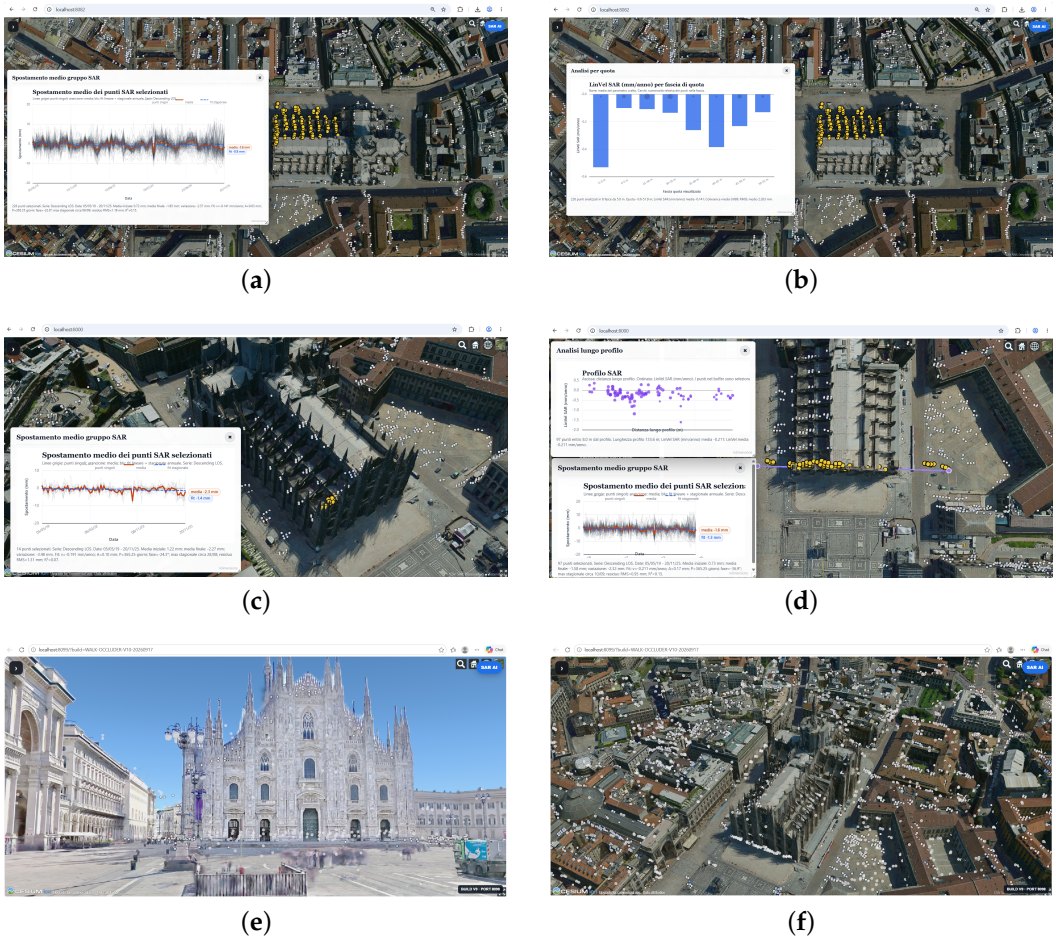


Figure 6. Query of PSI data in the platform: (**a**) temporal analysis of the PSI in the northern area of the cathedral, (**b**) analysis of PSI data organized for elevation, (**c**) temporal analysis of selected PSI data, (**d**) analysis of PSI data along a section, (**e**) visualization of PSI data and gaussian splatting, (**f**) visualization of PSI data and municipal layer of digital twin of the city of Milan.

The need for different visualization modes arises from the heterogeneous nature of the integrated data. PSI observations are spatial measurements acquired from a satellite line of sight, while the GS model is a photorealistic reconstruction optimized for visual rendering. The municipal digital-twin data provide explicit geospatial structure, while the monitoring time series provide quantitative deformation information. No single visualization mode is sufficient to represent all these aspects. The aerial view supports global spatial reasoning, the oblique 3D view supports architectural interpretation, and the walking mode supports

inspection-oriented exploration. The user can therefore move between complementary representations according to the question being investigated.

A specific occlusion layer was introduced to improve the consistency of the GS-based visualization. Although gaussian splatting provides highly realistic visual results, it does not always provide an explicit and robust geometric structure for visibility reasoning. In particular, PSI markers may remain visible even when they are located behind the monument, behind surrounding buildings, or outside the portion of the scene that should be visible from a specific viewpoint. This can generate misleading visual interpretations, especially in street-level navigation, where the user expects a behaviour similar to physical inspection. To address this issue, the platform uses a simplified occlusion layer derived from explicit urban geometry, such as OpenStreetMap or municipal building footprints. These geometries act as proxy volumes for visibility management and marker filtering.

The simplified occlusion geometry is used to hide markers visually when they lie behind the building from the current viewpoint, avoiding a visually confusing display of points through the 3D representation. This affects their display, not the underlying monitoring observations or their availability for analysis. The occlusion layer is more operational and visual: it improves scene readability, reduces misleading marker visibility, and helps the user understand the relationship between the monitoring points and the surrounding urban morphology. This distinction is important because the GS model is used as a photorealistic visualization layer, whereas the occlusion layer provides simplified geometric support for visibility reasoning. In the walking-navigation mode, the occlusion layer can be automatically activated and updated as the camera moves, allowing the visualization of PSI markers to remain more consistent with the user's current viewpoint.

PSI data are represented in the Cesium scene as georeferenced monitoring entities. The user can inspect a single point to visualize its displacement history or can select a group of points to analyse the collective behaviour of a façade sector or urban area. This second option is particularly important for PSI data, since the interpretation of a single scatterer may be affected by local scattering conditions, residual phase noise, or changes in radar reflectivity. A group-based analysis reduces the influence of isolated targets and allows the user to focus on the common component shared by multiple scatterers.

When a group of PSI points is selected, the platform computes aggregate descriptors of the corresponding displacement time series. The selected group is then summarized through a graphical representation showing the temporal evolution of the aggregate displacement signal. Additional indicators, such as the number of selected points, temporal coverage, average displacement, minimum and maximum values, and linear trend estimates, can also be computed. The graph is designed to provide a rapid visual assessment of whether the selected sector exhibits stable behaviour, seasonal oscillation, progressive displacement, or anomalous variability. This interaction design allows the user to move progressively from spatial exploration to quantitative analysis. First, the user identifies an area of interest in the 3D scene. Then, a set of PSI points is selected directly in space. The platform extracts the corresponding time series and produces a graphical summary.

Finally, the same selected context can be transferred to the AI assistant for further interpretation. In this way, the spatial selection becomes an analytical object that can be visualized, summarized, compared, and discussed through natural language. The platform therefore supports a multimodal workflow in which the user alternates between visual navigation, spatial selection, time-series analysis, and AI-assisted reasoning.

The same mechanism can be used for comparative exploration. The user may select two different groups of scatterers, for instance one located on the façade and another located in the surrounding square or on a neighbouring sector. The system can compare their aggregate displacement histories, evaluate whether their trends are similar, and identify

differences in amplitude, variability, or temporal behaviour. This type of comparison is useful for moving beyond point-wise analysis and towards spatially distributed structural interpretation. In a structural digital-twin context, the objective is not only to know how a single monitoring point behaves, but to understand whether different parts of the asset share the same deformation regime or whether one area behaves differently from the surrounding context.

Overall, the multimodal visualization environment allows the monitoring data to be explored as spatially situated evidence. PSI observations are not presented as disconnected numerical records, but as measurements located within a complex architectural and urban scene. The GS model supports visual interpretation, the municipal digital-twin data provide geospatial consistency, the occlusion layer improves viewpoint-dependent readability, and the plotting tools allow the user to inspect the temporal behaviour of selected points or groups. This combination provides the basis for the AI-assisted interpretation described in the following subsection.

3.3. Knowledge-Grounded AI Integration and Operator-Managed Knowledge Base

The AI component of the platform was designed as a knowledge-grounded layer connected to both the 3D visualization environment and the documentary archive of the monitoring project. The objective is not to use the Large Language Model as a generic chatbot, but to provide it with a controlled project-specific context composed of monitoring data, technical documentation, spatial selections, and operator-generated notes. For this reason, the system follows a retrieval strategy in which the LLM receives, at query time, the information that is relevant to the current user request and to the selected area of the digital twin (Figure 7).

Several categories of information are integrated into the AI workflow. The first category consists of textual and semi-structured documentation. This includes historical monitoring reports, technical notes, inspection records, scientific documents, maintenance documentation, spreadsheets, and operator-uploaded files. This allows the system to retrieve semantically relevant passages even when the user does not use the exact terminology contained in the original documents. In this way, decades of accumulated monitoring knowledge can be queried through natural language rather than through manual document search.

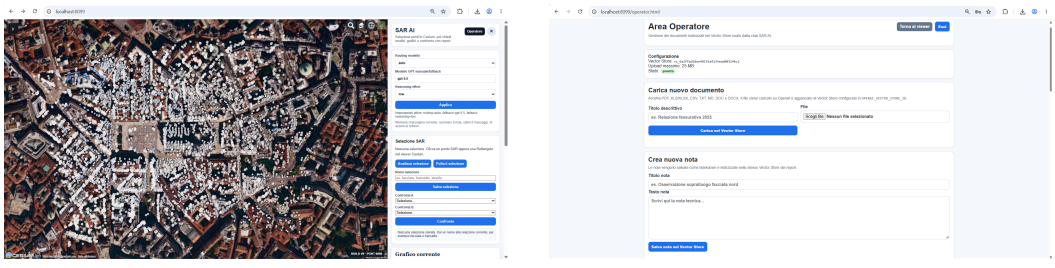
The second category consists of numerical and geospatial monitoring data. In the current implementation, PSI observations play a central role. Each PSI point displayed in the Cesium viewer is associated with displacement time series and metadata. When the user selects one or more points in the 3D scene, the frontend extracts the corresponding monitoring context and sends it to the backend together with the user's natural-language prompt. This context may include the list of selected points, their identifiers, the number of observations, the time interval covered by the data, aggregate displacement statistics, trend estimates, and other descriptors derived from the selected time series.

The third category consists of interaction-derived context. The platform does not treat a query as an isolated text message. Instead, each query is interpreted together with the current state of the viewer. For example, if the user selects a group of scatterers on a façade and asks whether "this area" is stable, the system associates the expression "this area" with the active spatial selection. Similarly, if an aggregate graph has been generated from the selected scatterers, the corresponding statistical summary and time-series descriptors are attached to the prompt. This mechanism allows the LLM to reason over the specific data currently being explored, rather than producing a generic answer about structural monitoring.

At query time, the backend combines the user prompt, the selected monitoring context, and the retrieved documentary evidence. The workflow can be summarized as follows. First, the user interacts with the Cesium viewer by selecting one or more monitoring points or by asking a question in the AI panel. Second, the platform serializes the current analytical context, including spatial selection, time-series descriptors, and relevant metadata. Third, the backend performs semantic retrieval from the vector store to identify technical reports, notes, or document fragments related to the query. Document indexing relies on the OpenAI vector store service, using the default chunking configuration of 800 tokens per chunk with a 400-token overlap. Fourth, an augmented prompt is constructed and submitted to the LLM. This prompt contains the user question, the selected monitoring evidence, the retrieved documentary evidence, and instructions requiring the model to distinguish between measured data, retrieved information, and interpretative hypotheses. The generated answer is then returned to the frontend and displayed in the AI panel.

This approach makes it possible to formulate different types of queries. Descriptive queries can be used to ask which points are selected, what time interval is covered by the data, what the average displacement is, or whether the selected points show coherent behaviour. Analytical queries can ask the system to identify trends, compare two groups of scatterers, evaluate seasonal behaviour, or summarize differences between mean, median, and weighted displacement series. Interpretative queries can ask whether an observed pattern is compatible with previously documented behaviour, environmental forcing, thermal expansion, or known structural mechanisms. Documentary queries can ask the system to retrieve information from historical reports, summarize previous inspections, or identify whether similar behaviour was observed in the past.

Examples of possible prompts include: “Summarize the behaviour of the selected PSI points”; “Does this façade sector show a significant long-term displacement trend?”; “Is the observed signal mainly seasonal or progressive?”; “Compare this group of scatterers with the historical monitoring reports”; “Which documents mention movements in this area?”; “Are these displacements compatible with thermal effects?”; or “Generate a technical note about the current selection.” These examples illustrate the transition from data-centric interrogation to structure-centric reasoning. The user does not need to know the exact database fields, file names, or report titles. The spatial selection and semantic retrieval process provide the missing context required to formulate an evidence-based answer.



(**a**) (**b**)

Figure 7. AI implementation in the platform: (**a**) the AI interface integrated in the platform with the possibility of selecting the AI model, (**b**) the operator interface allowing for uploading documents or notes in the project vector store.

Two alternative LLM implementation strategies were tested. The first strategy used a cloud-based ChatGPT/OpenAI configuration. In this configuration, the platform relies on the OpenAI API and a project-specific vector store. The cloud-based implementation employed OpenAI GPT-5.4 and GPT-5.5 reasoning models through the API, with reasoning effort set to high. The project knowledge base used OpenAI document indexing and retrieval services, without fine-tuning the underlying language models. Document

ingestion, embedding generation, semantic retrieval, and LLM inference are handled through the cloud-based infrastructure. This configuration supports document retrieval and synthesis across heterogeneous sources, while allowing relatively straightforward integration with an updateable project-specific vector store. It is particularly suitable for a demonstrator in which the objective is to evaluate the interaction paradigm and the ability of knowledge-grounded AI to support monitoring interpretation. Its main limitation is that prompts, retrieved contexts, and selected metadata are processed by an external service, which requires careful consideration of privacy, confidentiality, and institutional data-governance policies.

The second strategy was based on a local architecture using Ollama and AnythingLLM. In this configuration, Ollama provides the runtime for local open-source language models, while AnythingLLM supports document ingestion, workspace organization, retrieval, and conversational interaction. This approach offers greater control over data locality because documents and model inference can remain on local hardware or within an institutional network. It is therefore attractive for scenarios in which historical reports, technical assessments, or maintenance records cannot be processed by external cloud services. However, the quality and robustness of the responses depend strongly on the selected local model, the available hardware, the context-window size, and the retrieval configuration. Smaller local models may be adequate for simple summarization or document lookup, but they may be less reliable for complex engineering interpretation and multi-source reasoning.

For the final demonstrator, the cloud-based OpenAI configuration was adopted as the primary implementation because it supported the required integration between the Cesium viewer, the PSI selection workflow, the operator-managed document repository, and the knowledge-grounded chat interface. Nevertheless, the tests with Ollama and AnythingLLM demonstrated that the proposed architecture is not conceptually tied to a single LLM provider. The essential component is not the specific model, but the surrounding infrastructure that supplies the model with validated monitoring data, retrieved technical documentation, spatial context from the 3D viewer, and explicit instructions for evidence-based interpretation.

The platform also includes a protected operator interface for maintaining and updating the knowledge base. Authenticated operators can upload new documents, create textual notes, inspect the files currently associated with the vector store, delete obsolete material, or replace existing documents with updated versions. Newly uploaded material is indexed in the same vector store used by the AI assistant and becomes available for subsequent queries without retraining the underlying language model. This mechanism is particularly relevant for long-term monitoring, where new inspection reports, measurement campaigns, maintenance records, and expert observations are continuously produced.

The operator-managed knowledge base transforms the AI component from a static demonstrator into an incrementally updateable monitoring knowledge system. Each new document or note can enrich the context available to future users. For example, after an inspection campaign, an operator can upload a technical report. Subsequent processing can then ask whether a newly observed displacement pattern is consistent with that inspection record. In this way, the AI system supports continuity between monitoring campaigns and helps preserve institutional knowledge that would otherwise remain distributed across separate reports, files, and individual expert experience.

From an engineering perspective, the AI-generated responses should not be interpreted as autonomous diagnoses. The system is intended to support expert reasoning by retrieving relevant evidence, summarizing complex information, identifying possible correlations, and formulating hypotheses. Final interpretation and decision-making

remain the responsibility of domain specialists. This distinction is fundamental for monitoring applications, where the consequences of incorrect interpretation may be significant. The value of the AI layer lies in reducing the effort required to access and integrate heterogeneous evidence, while maintaining a clear separation between measured data, retrieved documentation, and model-generated interpretation.

The integration of a Large Language Model within a Cesium-based monitoring platform can significantly enhance the way users interact with complex monitoring datasets. However, it is important to understand that the quality and consistency of the responses do not depend solely on the language model itself, but rather on the overall system architecture.

A standard LLM accessed through an API is inherently stateless. Each request is processed independently unless additional context is explicitly provided. Consequently, if a user asks a question such as "How is the Cathedral square behaving?" the model has no intrinsic knowledge of the Duomo monitoring project, the available reports, the historical measurements, or previous conversations. The answer will therefore be based only on the information contained in the current request. To reproduce the behavior of an intelligent assistant capable of supporting long-term monitoring activities, it is necessary to provide the model with contextual information. This can be achieved by storing and retrieving previous conversations, technical reports, monitoring data, and metadata describing the monitored structure. In practice, the language model becomes only one component of a larger ecosystem.

Long-term continuity can be further improved by introducing a persistent project memory. Instead of relying exclusively on conversational history, the system stores structured knowledge describing the monitored asset. Such knowledge may include the identification of monitoring points, the location of structural elements, known anomalies, historical interventions, and recurring patterns of behavior. When a future query is submitted, the platform can automatically retrieve this information and provide it to the language model. As a result, questions asked months or even years apart can be interpreted within a consistent project context.

Achieving responses comparable to those obtained in an interactive project-oriented environment requires more than simply connecting an LLM through an API. It requires the combination of a monitoring database, document repositories, retrieval mechanisms, conversational memory, and structured project knowledge. When these components are integrated, the resulting system is capable of supporting informed decision-making, historical analysis, anomaly investigation, and long-term monitoring of complex cultural heritage assets.

4. Results and Discussion

4.1. Multimodal Exploration Combining 3D Interaction and Text Prompts

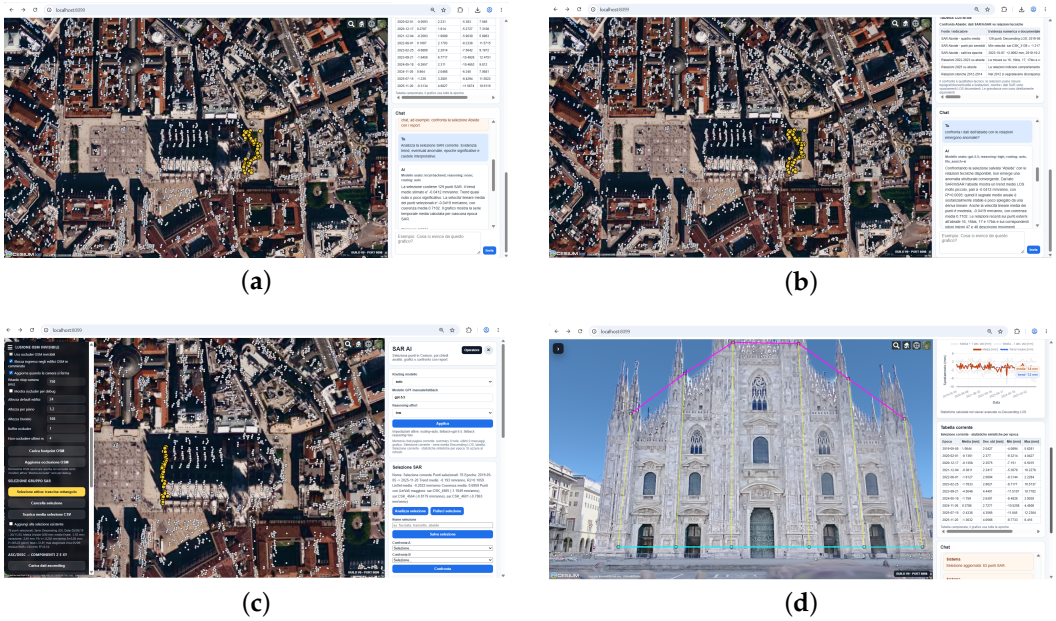
A representative use case of the proposed system concerns the analysis of monitoring measurements over the façade sector of the Milan Cathedral. The workflow starts from the Cesium-based viewer, where the user navigates the georeferenced scene and visually identifies the area of interest. The façade, the surrounding square, the PSI scatterers, the map background, and the gaussian splatting reconstruction are displayed in a common spatial reference. This allows the user to interpret the distribution of the scatterers not only as a set of numerical records but also as observations physically located on or around the monument.

The user can select a group of persistent scatterers directly in the viewer, for example by drawing a selection region over the monitored façade or over a subset of scatterers located in the Cathedral square. Once the selection is completed, the system automati-

cally extracts the time series associated with the selected points and computes aggregate descriptors. These include the number of selected scatterers, temporal coverage, mean displacement series, median series, trend estimates, and basic statistics such as minimum, maximum, and average displacement values. The corresponding graph is generated in the interface and provides an immediate visual representation of the collective behaviour of the selected group.

This group-based approach is particularly useful for interpretation. As mentioned, by selecting a spatially coherent group of scatterers, the user can reduce the influence of isolated targets and focus on the common deformation component shared by multiple observations. In the demonstrator, the selected points are therefore not interpreted individually only, but can be aggregated into a façade-scale or sector-scale signal. This strategy is consistent with the validation logic adopted in the monitoring analysis, where multiple scatterers are combined to obtain a more robust representation of the structural behaviour of a monitored sector.

The selected SAR context is then transferred to the AI assistant. The side panel dedicated to AI receives a structured description of the current selection, including the identifiers of the selected points, their spatial distribution, the computed statistics, and the relevant time-series descriptors. The user can then ask a natural language question referring implicitly or explicitly to the selected region. For example, after selecting the façade scatterers, the user may ask: "Is this portion of the façade showing a significant displacement trend?" or "Is the observed behaviour compatible with seasonal deformation?" The system interprets the question in relation to the active selection and combines the numerical evidence from the viewer with information retrieved from the knowledge base (Figure 8).



(**a**) (**b**)

(**c**) (**d**)

Figure 8. Example of data query using the AI model: (**a**) data selection and simple query (summary of selected PSI data trend) for the AI model in the apse, (**b**) a more complex query for the AI model involving selected PSI data and information coming from the historical reports in the apse area, (**c**) selection of the facade area PSI, (**d**) facade area PSI and the ground monitoring system.

The response generated by the AI assistant is therefore grounded in two types of evidence. The first is the live analytical context derived from the selected PSI data, including the current graph and statistical indicators. The second is the documentary context retrieved from the vector store, which may include historical monitoring reports, previous technical interpretations, inspection notes, and descriptions of known deformation mechanisms.

The assistant can consequently provide an answer that does not merely describe the graph but relates the observed signal to previous knowledge about the monument. For instance, it can distinguish between a weak long-term trend and a recurring seasonal oscillation, or indicate whether the magnitude of the displacement is consistent with previously documented millimetric movements.

An important aspect of this interaction is that the user does not need to manually formulate a database query or search through reports before asking a question. The spatial selection made in the 3D viewer becomes part of the prompt context. This creates a multimodal interaction paradigm in which the user first defines the object of interest geometrically and then reasons about it linguistically. The same mechanism can be used to compare different portions of the structure. For example, the user can select one group of scatterers on the façade and another group in a neighbouring area, then ask the assistant to compare their displacement behaviour. The system can compute and present differences in average displacement, trend, temporal variability, and seasonal amplitude. The AI assistant can then discuss whether the two groups appear to behave coherently or whether one sector shows a distinct pattern. Such comparison is useful for moving from point-wise interpretation to spatially distributed structural reasoning, where the objective is not only to detect movement but to understand whether different parts of the monument share the same deformation regime.

The user is not required to switch between separate software, plotting tools, document archives, and AI interfaces. Instead, the selection of monitoring data, the generation of graphical summaries, the retrieval of historical documentation, and the formulation of technical explanations occur within a single integrated environment.

4.2. Natural Language Interaction for Advanced Operations

One of the most interesting aspects of integrating monitoring reports into the AI system is the ability to interact with it using natural language prompts. This enables users to communicate with an AI assistant in a conversational way, as if consulting a knowledgeable colleague capable of accessing, integrating, and analyzing heterogeneous datasets in real time.

This approach represents a shift from data-centric queries to structure-centric reasoning. Instead of interrogating individual measurements or time series (e.g., "Is there a discontinuity in this time series?"), users can ask higher-level questions about the structure itself, such as "Is the façade behaving normally?" or "Are there any areas that require attention?" The AI interprets these requests by identifying the relevant structural components, retrieving information from monitoring data, historical reports, inspection records, and other available sources, and synthesizing the evidence into a comprehensive assessment. In this way, the interaction becomes focused on understanding the condition and behaviour of the monument rather than on analyzing isolated datasets.

The AI responses are not limited to natural language explanations. When appropriate, the system can automatically generate additional outputs, such as time-series graphs, trend analyses, statistical summaries, and other visual representations that support its conclusions. This combination of conversational interaction and on-demand analytical outputs allows specialists to move between general questions and detailed supporting evidence, which can then be examined and verified.

The overall user experience is therefore similar to collaborating with a (virtual) colleague that has immediate access to large volumes of heterogeneous information and can rapidly identify relevant evidence across multiple data sources. Such a system can assist users by highlighting relationships, patterns, or historical information that might otherwise require years of domain expertise or extensive manual investigation to uncover.

At the same time, the AI system should not be regarded as an error-free specialist or an authoritative source of truth. Rather, its responses should be interpreted as informed hypotheses or analyses that require critical evaluation by domain experts, much as suggestions and opinions exchanged during discussions between human specialists are subject to verification and validation. It provides a more natural and intuitive way of interacting with complex monitoring datasets through conversational language. This conversational paradigm lowers the barrier between users and the underlying information, allowing experts to focus on interpreting structural behavior rather than on navigating databases or formulating complex technical queries.

From this perspective, the main contribution of the proposed system is not that the AI behaves like a human expert, but that it fundamentally changes the interface between specialists and complex monitoring data. Specialists can reason directly about the physical structure using natural language, while the AI manages the retrieval, integration, and interpretation of the underlying heterogeneous datasets. This shift from data-centric queries to structure-centric reasoning provides a more direct way to interact with long-term monitoring information, while retaining the critical judgment of domain experts.

AI-generated responses can be critically reviewed by specialists familiar with the monitored structure and checked against the available measurements and documentation. Such expert review is important in practice, but it should not be confused with a formal, independent validation of analytical accuracy. A possible direction for future validation is the systematic comparison of AI-generated analyses with results independently obtained by domain experts using conventional methodologies. Such an evaluation could assess both the quality of the generated responses and their consistency with established engineering interpretations.

As a representative example, the AI system was asked to perform a cluster analysis of the monitored columns using the available long-term monitoring data. The results are shown in Figure 9. The prompt explicitly requested the identification of $k = 4$ clusters, consistent with the methodology adopted in the previous study by [43]. The previous study was not included in the project knowledge base. However, this does not establish complete independence from information potentially available through the model's pretraining.

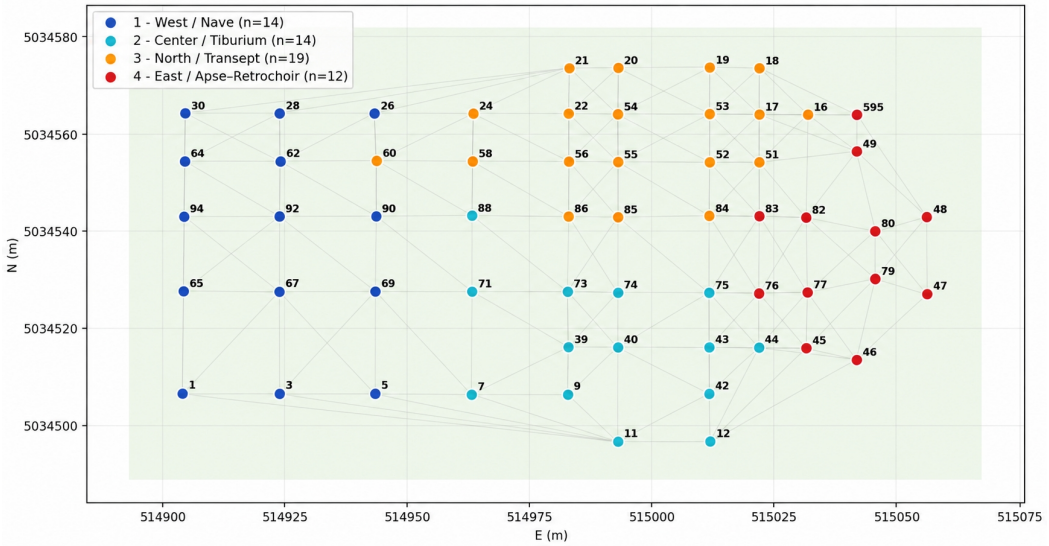


Figure 9. Results of the analysis obtained using the AI-based system, showing the spatial distribution of the identified clusters.

Furthermore, the prompt instructed the AI to account not only for the similarity of the displacement time series but also for the spatial distribution of the columns within the

cathedral, recognizing that neighboring structural elements are expected to exhibit correlated behavior. The resulting clusters showed similarities with those previously identified through expert analysis. This qualitative comparison illustrates how temporal monitoring data and spatial constraints can be considered together within an AI-assisted analytical workflow. The fact that the number of clusters was defined a priori by the user can be viewed from two perspectives. On the one hand, the AI was given less autonomy, and its decisions were, to some extent, guided towards the results expected by the operator. On the other hand, this confirms the importance of allowing such systems to incorporate input from expert users rather than operating entirely independently.

This example illustrates how AI-supported analysis can extend beyond document retrieval by combining numerical observations, spatial information, and constraints introduced by the specialist. The resulting interpretation remains part of an interactive process in which engineering knowledge and critical assessment play an essential role.

4.3. Interpretation, Expert Interaction, and Limitations

The proposed framework illustrates how monitoring observations, existing documentation, and spatial information can be brought together to support the interpretation of building behaviour. Its main contribution lies in the interaction between the knowledge made accessible through AI and the experience of specialists, rather than in the independent performance of the underlying language models. Specialists contribute knowledge of construction characteristics, previous interventions, field observations, and the monitoring context. They can use this experience to formulate questions, introduce constraints, and critically assess the engineering significance of the results. A statistically optimal solution is not necessarily the most meaningful from an engineering perspective.

In the clustering example, the number of groups was deliberately fixed to four to facilitate comparison with an earlier expert interpretation, while spatial proximity was introduced as a user-defined condition. The example is therefore an exploratory, expert-guided analysis, not an independent validation of autonomous clustering. Quantitative agreement and stability tests would be needed to assess the reliability of the classification.

The reliability of AI-assisted interpretation also depends on the available evidence. The knowledge base was initially built from monitoring reports prepared by experienced specialists over many years, documenting repeated observations and their interpretation. These reports provide an established foundation, but may contain omissions, inconsistencies, or subjective assessments. Retrieved information does not eliminate the possibility of unsupported AI-generated conclusions. Unexpected results, inconsistencies, and even hallucinations may prompt useful questions and further verification when recognized by a knowledgeable specialist, but they must not be accepted as evidence without checking the original measurements and documentation.

Three related challenges deserve particular attention. First, increased access to information may lead to information overload: more data and retrieved material do not necessarily produce better understanding or decisions, and may distract from the most relevant observations. Second, effective interaction requires cognitive effort from specialists, who must formulate meaningful requests, evaluate competing interpretations, and verify the evidence. Third, these systems need to be considered in relation to established professional practice. Their value depends not only on what they can retrieve or calculate, but also on whether the information improves the way monitoring activities and engineering decisions are carried out.

The framework could be adapted to other monitored buildings provided suitable observations, documentation, and spatial information are available. In this context, knowledge grounding should be understood as a means of providing relevant documentary evidence,

rather than as a guarantee of factual correctness. The distinction between measured observations, retrieved information, and AI-generated interpretations remains essential for maintaining expert control over the assessment process. However, the present study does not include controlled comparisons of AI configurations or a quantitative assessment of analytical accuracy. The examples illustrate the proposed integration and interaction approach, rather than superior performance over conventional monitoring methods. Accordingly, the evidence presented concerns the feasibility and functionality of the integrated workflow, rather than the statistical performance of individual AI components. Quantitative benchmarking and ablation studies would address complementary questions concerning retrieval accuracy, model reliability, and the relative contribution of individual components.

5. Conclusions

This study explored the development of a knowledge-grounded Digital Twin for long-term structural monitoring, combining numerical observations, existing documentation, spatial information, and AI reasoning capabilities within an interactive environment.

The integration of ground-based measurements and satellite interferometry with the Cesium-based 3D environment and gaussian splatting visualization provides spatial access to heterogeneous monitoring information. The experimental examples also illustrate how specialists can guide AI-supported analyses through natural-language requests and user-defined constraints. In particular, the clustering example demonstrates an exploratory interaction between computational analysis and previous expert knowledge, rather than an independent quantitative validation of the resulting classification.

A central aspect of the proposed approach is the complementarity of AI capabilities and human expertise. The system can facilitate access to accumulated knowledge and support analytical tasks, but specialists remain responsible for formulating meaningful questions, evaluating the results, and interpreting their engineering significance. At the same time, greater access to information does not necessarily lead to better decisions. The potential benefits of these systems must be weighed against the risks of incorrect AI-generated interpretations, information overload, and the additional cognitive effort required from users. Understanding how such tools can be integrated into established monitoring practices remains an important direction for future research. As discussed in Section 4, human–AI interaction should facilitate rather than complicate the work of specialists, avoiding unnecessary effort and frustration. When effectively designed, such interaction can provide meaningful benefits by supporting expert interpretation.

Future developments may include more autonomous analysis of newly acquired observations. These capabilities would require further validation before being adopted for operational decision support.

Author Contributions: Conceptualization, L.B., F.R. and M.P.; methodology, L.B., R.E., M.P.; software, L.B., R.E., M.P.; validation, L.B., F.R. and M.P.; formal analysis, L.B., M.P.; investigation, L.B., R.E., A.K., M.P. and F.R.; data curation, L.B., R.E., M.P. and F.R.; writing, L.B., R.E. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request, subject to verification of the intended use.

Acknowledgments: The authors would like to thank the staff of the Veneranda Fabbrica del Duomo di Milano for providing access to Milan Cathedral and for their support during data acquisition. Special thanks are extended to Francesco Canali, Director of the Restoration Sites, for his valuable support and assistance. During the preparation of this manuscript, the authors used OpenAI ChatGPT (GPT-5.6 Sol) for the generation and iterative refinement of schematic Figures 1, 2 and 5 based on

scientific concepts, content, and instructions provided by the authors. The generated graphics were subsequently reviewed and revised by the authors to ensure consistency with the methodological framework presented in the manuscript. Generative AI was also used as an integral component of the knowledge-grounded system developed and evaluated in this study, as described in Section 2. Figure 9 presents an output generated by this system. The authors have reviewed and edited all AI-generated content and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Attaran, M.; Çelik, B.G. Digital Twin: Benefits, use cases, challenges, and opportunities. *Decis. Anal. J.* **2023**, *6*, 100165. https://doi.org/10.1016/j.dajour.2023.100165.
2. Botín-Sanabria, D.M.; Mihăiţă, A.; Peimbert-García, R.E.; Ramírez-Moreno, M.A.; Ramírez-Mendoza, R.A.; de J. Lozoya-Santos, J. Digital Twin Technology Challenges and Applications: A Comprehensive Review. *Remote Sens.* **2022**, *14*, 1335. https://doi.org/10.3390/rs14061335.
3. Xia, G.; Ghrairi, Z.; Wuest, T.; Hribernik, K.; Heuermann, A.; Liu, F.; Liu, H.; Thoben, K.D. Towards Human Modeling for Human-Robot Collaboration and Digital Twins in Industrial Environments: Research Status, Prospects, and Challenges. *Robot. Comput.-Integr. Manuf.* **2025**, *95* 1–20. https://doi.org/10.1016/j.rcim.2025.103043.
4. Ahsan, M.M.; Liu, Y.; Raman, S.; Siddique, Z. Digital Twins in Additive Manufacturing: A systematic review. *Internet Things* **2025**, *33*, 101692. https://doi.org/10.1016/j.iot.2025.101692.
5. Ai, L.; Ziehl, P. Advances in digital twin technology in industry: A review of applications, challenges, and standardization. *J. Intell. Constr.* **2025**, *3*, 1–19. https://doi.org/10.26599/JIC.2025.9180083.
6. Singh, M.; Srivastava, R.; Fuenmayor, E.; Kuts, V.; Qiao, Y.; Murray, N.; Devine, D.M. Applications of Digital Twin across Industries: A Review. *Appl. Sci.* **2022**, *12*, 5727. https://doi.org/10.3390/app12115727.
7. Tao, F.; Zhang, H.; Liu, A.; Nee, A. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2405–2415. https://doi.org/10.1109/tii.2018.2873186.
8. Sharma, A.; Kosasih, E.E.; Zhang, J.; Brintrup, A.; Calinescu, A. Digital Twins: State of the art theory and practice, challenges, and open research questions. *J. Ind. Inf. Integr.* **2022**, *30*, 100383. https://doi.org/10.1016/j.jii.2022.100383.
9. Li, Q.; Ye, Q.; Zhang, N.; Zhang, W.; Hu, F. Digital-Twin-Enabled Industrial IoT: Vision, Framework, and Future Directions. *IEEE Wirel. Commun.* **2025**, *32*, 173–181. https://doi.org/10.1109/mcom.003.2400515.
10. Ma, S.; Flanigan, K.A.; Bergés, M. Towards a Universal Digital Twin Framework for Standardized Development. In Proceedings of the 2025 IEEE 5th International Conference on Digital Twins and Parallel Intelligence (DTPI), Atlanta, GA, USA, 22–24 April 2025; pp. 1–7. https://doi.org/10.1109/dtpi65196.2025.11088873.
11. Chen, C.; Fu, H.; Zheng, Y.; Tao, F.; Liu, Y. The advance of digital twin for predictive maintenance: The role and function of machine learning. *J. Manuf. Syst.* **2023**, *71*, 581–594. https://doi.org/10.1016/j.jmsy.2023.10.010.
12. Alshathri, S.; Hemdan, E.E.; El-Shafai, W.; Sayed, A. Digital Twin-Based Automated Fault Diagnosis in Industrial IoT Applications. *Comput. Mater. Contin.* **2023**, *75*, 183–196. https://doi.org/10.32604/cmc.2023.034048.
13. Soori, M.; Arezoo, B.; Dastres, R. Digital twin for smart manufacturing, A review. *Sustain. Manuf. Serv. Econ.* **2023**, *2*, 100017. https://doi.org/10.1016/j.smse.2023.100017.
14. Carbonara, G. An Italian contribution to architectural restoration. *Front. Archit. Res.* **2012**, *1*, 2–9. https://doi.org/https://doi.org/10.1016/j.foar.2012.02.007.
15. Dagnaw, G.A.; Capuano, R.; Muccini, H. Digital Twins for Cultural Heritage: A Systematic Analysis of the State of the Art. *ACM Comput. Surv.* **2026**, *58*, 1–35. https://doi.org/10.1145/3793541.
16. Colace, F.; Limongiello, M.; Lorusso, A.; Pellegrino, M.; Santaniello, D.; Santoriello, A. Digital twin for cultural heritage: A computational approach to predictive conservation. *Digit. Appl. Archaeol. Cult. Herit.* **2026**, *40*, e00519. https://doi.org/10.1016/j.daach.2026.e00519.
17. Barazzetti, L. Spatio-temporal analysis of georeferenced time-series applied to structural monitoring. *J. Civ. Struct. Health Monit.* **2023**, *14*, 163–188. https://doi.org/10.1007/s13349-023-00743-z.
18. Buglisi, S.; Luca, L.D.; Lauria, M.; Quattrocchi, A. Spatio-Temporal Data Integration for Knowledge-Driven Analysis of Historic Built Heritage Conservation State. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*; Copernicus Publications: Göttingen, Germany, 2025; pp. 187–194. https://doi.org/10.5194/isprs-archives-xlviii-m-9-2025-187-2025.
19. Lei, Y.J.; Bruno, N.; Roncella, R. A Data-driven Information Modelling Approach for Cultural Heritage. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*; Copernicus Publications: Göttingen, Germany, 2025; pp. 813–820. https://doi.org/10.5194/isprs-archives-xlviii-m-9-2025-813-2025.

20. Laohaviraphap, N.; Waroonkun, T. Integrating Artificial Intelligence and the Internet of Things in Cultural Heritage Preservation: A Systematic Review of Risk Management and Environmental Monitoring Strategies. *Buildings* **2024**, *14*, 3979. https://doi.org/10.3390/buildings14123979.
21. Geng, S.; Yan, S.; Chau, H.W.; Jamei, E.; Vrcelj, Z. From Data to Cultural Response: A Machine Learning–Driven Digital Twin Model for Smart Heritage Precincts in Urban Context. *J. Inf. Technol. Constr.* **2025**, *30*, 1314–1314. https://doi.org/10.36680/j.itcon.2025.053.
22. Nie, Y.; Wu, Z.; Xing, Z.; Luo, M. Digital Twin for Architectural Heritage: A Comprehensive Conceptual Framework Integrating Structural Health, Microclimate, and Energy Performance. *Sustainability* **2026**, *18*, 3080. https://doi.org/10.3390/su18063080.
23. Casillo, M.; Colace, F.; Gaeta, R.; Lorusso, A.; Santaniello, D.; Valentino, C. Revolutionizing cultural heritage preservation: An innovative IoT-based framework for protecting historical buildings. *Evol. Intel.* **2024**, *17*, 3815–3831. https://doi.org/10.1007/s12065-024-00959-y.
24. Felicetti, A.; Niccolucci, F. Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data. *Data* **2025**, *10*, 1. https://doi.org/10.3390/data10010001.
25. García-Macías, E.; Venanzi, I.; Ubertini, F. Metamodel-based pattern recognition approach for real-time identification of earthquake-induced damage in historic masonry structures. *Autom. Constr.* **2020**, *120*, 103389. https://doi.org/10.1016/j.autcon.2020.103389.
26. Zhang, B.; Ren, Y.; He, S.; Gao, Z.; Li, B.; Song, J. A review of methods and applications in structural health monitoring (SHM) for bridges. *Measurement* **2025**, *245*, 116575. https://doi.org/10.1016/j.measurement.2024.116575.
27. Sivasuriyan, A.; Vijayan, D.S.; Piętocha, A.; Górski, W.; Łukasz Wodzyński.; Koda, E. Advanced Sensing and Digital Monitoring Technologies for Structural Health Assessment of Civil Infrastructure. *Buildings* **2026**, *16*, 656. https://doi.org/10.3390/buildings16030656.
28. Ergen, V.; Zülfikar, A.C. Scalable IoT-Based Structural Health Monitoring System for Post-Earthquake Rapid Assessment. *Buildings* **2026**, *16*, 950. https://doi.org/10.3390/buildings16050950.
29. Yu, J.; Meng, X.; Yan, B.; Xu, B.; Fan, Q.; Xie, Y. Global Navigation Satellite System-based positioning technology for structural health monitoring: A review. *Struct. Control Health Monit.* **2019**, *27*, 1–14. https://doi.org/10.1002/stc.2467.
30. Nesterenko, S.; Radzinska, Y.; Yevdokimov, A.; Shterndok, E. Modern Aspects of Methodological Support of Geodetic Monitoring of Objects and Territories. *Munic. Econ. Cities* **2024**, *6*, 199–203. https://doi.org/10.33042/2522-1809-2024-6-187-199-203.
31. Xie, Y.; Zhang, S.; Meng, X.; Nguyen, D.T.; Ye, G.; Li, H. An Innovative Sensor Integrated with GNSS and Accelerometer for Bridge Health Monitoring. *Remote Sens.* **2024**, *16*, 607. https://doi.org/10.3390/rs16040607.
32. Themistocleous, K.; Danezis, C.; Gikas, V. Monitoring ground deformation of cultural heritage sites using SAR and geodetic techniques: The case study of Choirokoitia, Cyprus. *Appl. Geomat.* **2020**, *13*, 37–49. https://doi.org/10.1007/s12518-020-00329-0.
33. Carlo, F.D.; Miano, A.; Giannetti, I.; Mele, A.; Bonano, M.; Lanari, R.; Meda, A.; Prota, A. On the integration of multi-temporal synthetic aperture radar interferometry products and historical surveys data for buildings structural monitoring. *J. Civ. Struct. Health Monit.* **2021**, *11*, 1429–1447. https://doi.org/10.1007/s13349-021-00518-4.
34. Yang, K.; Xu, B.; Qiu, Y.; Sun, P.; Liang, J.; Zhang, X.; Wang, X. Spaceborne InSAR for Monitoring and Health Assessment of Historical Architectures: The Tianjin Case Study. Available online: https://assets-eu.researchsquare.com/files/rs-2722550/v1_covered.pdf (accessed on 27 September 2026).
35. Zhou, W.; Chen, F.; Guo, H. Differential Radar Interferometry for Structural and Ground Deformation Monitoring: A New Tool for the Conservation and Sustainability of Cultural Heritage Sites. *Sustainability* **2015**, *7*, 1712–1729. https://doi.org/10.3390/su7021712.
36. Ponzo, F.C.; Auletta, G.; Ielpo, P.; Ditommaso, R. DInSAR–SBAS satellite monitoring of infrastructures: How temperature affects the "Ponte della Musica" case study. *J. Civ. Struct. Health Monit.* **2024**, *14*, 745–761. https://doi.org/10.1007/s13349-023-00751-z.
37. Fu, L.; Kim, D.; An, J.B.; Lee, C.; Kim, D.; Seo, H. A multi-scale remote sensing for monitoring of heritage building through integration of satellite SAR, laser scanning, and drones. *KSCE J. Civ. Eng.* **2025**, *30*, 100441–100441. https://doi.org/10.1016/j.kscej.2025.100441.
38. Falcone, M.; Origlia, A.; Campi, M.; Martino, S.D. From Architectural Survey to Continuous Monitoring: Graph-Based Data Management for Cultural Heritage Conservation with Digital Twins. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*; Copernicus Publications: Göttingen, Germany, 2021; pp. 47–53. https://doi.org/10.5194/isprs-archives-xliii-b4-2021-47-2021.
39. Duan, S.; Tan, X.; Guo, P.; Guo, Y.; Bao, Y. The transformative roles of generative artificial intelligence in vision techniques for structural health monitoring: A state-of-the-art review. *Adv. Eng. Inform.* **2025**, *68*, 103719–103719. https://doi.org/10.1016/j.aei.2025.103719.

40. Barazzetti, L.; Canali, F.; Torre, S.D.; Gentile, C.; Previtali, M.; Roncoroni, F. Monitoring the Cathedral of Milan: An Archive with More Than 50 Years of Measurements. In *Computational Science and Its Applications—ICCSA 2022*; Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Torre, C.M., Rocha, J.G., Taniar, D., Apduhan, B.O., Tarantino, E., Ryu, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13375, pp. 575–590. https://doi.org/10.1007/978-3-031-10522-7_39.
41. Canali, F.; Coronelli, D.; Angjeliu, G. The worksites of the Duomo di Milano: Extending the life of a Cathedral. *Struct. Infrastruct. Eng.* **2025**, *21*, 1101–1116. https://doi.org/10.1080/15732479.2025.2453563.
42. Crosetto, M.; Monserrat, O.; Cuevas-González, M.; Devanthéry, N.; Crippa, B. Persistent Scatterer Interferometry: A review. *ISPRS J. Photogramm. Remote Sens.* **2015**, *115*, 78–89. https://doi.org/10.1016/j.isprsjprs.2015.10.011.
43. Barazzetti, L.; Previtali, M.; Roncoroni, F. Visualization and Processing of Structural Monitoring Data Using Space-Time Cubes. In *Proceedings of the Computational Science and Its Applications—ICCSA 2022*; Gervasi, O., Murgante, B., Hendrix, E.M.T., Taniar, D., Apduhan, B.O., Eds.; Springer: Cham, Switzerland, 2022; pp. 19–31.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.