

Article

Not peer-reviewed version

---

# Knowledge Graphs and Artificial Intelligence for the Implementation of Cognitive Heritage Digital Twins

---

[Achille Felicetti](#)\*, [Aida Himmiche](#), [Miriana Somenzi](#)

Posted Date: 21 July 2025

doi: 10.20944/preprints202507.1673.v1

Keywords: artificial intelligence; ontologies; knowledge graphs; digital twins; cultural heritage



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

## Article

# Knowledge Graphs and Artificial Intelligence for the Implementation of Cognitive Heritage Digital Twins

Achille Felicetti <sup>1\*</sup>, Aida Himmiche <sup>2</sup> and Miriana Somenzi <sup>2</sup>

<sup>1</sup> VAST LAB, PIN, Piazza dell'Università 1, 59100 Prato, Italy

<sup>2</sup> ARIADNE Research Infrastructure AISBL, Rue de Grand-Bigard 14, Berchen-sainte-Agathe, 1082 Bruxelles, Belgium

\* Correspondence: achille.felicetti@gmail.com

**Abstract:** This paper explores the integration of Artificial Intelligence and semantic technologies to support the creation of intelligent Heritage Digital Twins (HDT), digital constructs capable of representing, interpreting, and reasoning over cultural data. The study focuses on transforming the often fragmented and unstructured documentation produced in cultural heritage into coherent Knowledge Graphs aligned with the CIDOC CRM family of ontologies, particularly CRMhs and RHDTO. Two complementary AI-assisted workflows are proposed: one for extracting and formalising structured knowledge from heritage science reports, and another for enhancing AI models through the integration of curated ontological knowledge. The experiments demonstrate how this synergy facilitates both the retrieval and the reuse of complex information, while ensuring interpretability and semantic consistency. Beyond technical efficacy, the paper also addresses the ethical implications of AI use in cultural heritage, with particular attention to transparency, bias mitigation, and meaningful representation of diverse narratives. The results highlight the importance of a reflexive and ethically grounded deployment of AI, where knowledge extraction and machine learning are guided by structured ontologies and human oversight, to ensure conceptual rigour and respect for cultural complexity.

**Keywords:** artificial intelligence; ontologies; knowledge graphs; digital twins; cultural heritage

## 1. Introduction

It is said that Helen of Troy, the most beautiful woman in the world according to Greek mythology, never set foot on the shores of Asia Minor. In a dramatic and unsettling version of the myth, told by Euripides and echoed in later traditions, it was not the real Helen whom Paris carried off to Troy. Instead, a simulacrum, an *eidolon*, a perfect likeness devoid of substance, was sent by the gods to take her place. While the true Helen waited in silence in Egypt, perhaps unaware of the war unfolding in her name, it was her image that ignited desire, deceived kings, and plunged the world into its most storied conflict. The *eidolon* was not a lie, but a form with agency: it acted, provoked, transformed the course of history. Though lacking life of its own, it possessed real effectiveness.

In this figure of the operative double, of an image that does not merely represent but intervenes, we find a striking anticipation of the modern notion of digital twin. Such an entity is designed not just to mirror the original, but to accompany or replace it in action, to operate in its stead, often with greater reach and precision. A digital twin, to fulfil this role, must do more than resemble. Like Helen's phantom, it must be endowed with structured intelligence and a "grammar" that sustains its coherence. The story of Helen's *eidolon* speaks to an ancient intuition: that a likeness, when crafted with precision and animated with purpose, can operate in the world with consequences equal to, or even surpassing, those of its original. This mythic logic finds a contemporary counterpart in the digital age, where cultural entities are increasingly represented by what might be called *simulacra*, i.e. constructed doubles whose very name shares etymological roots with the verb *to simulate*. These digital twins are not inert copies, but dynamic, structured representations designed to engage with the world through processes of analysis, interaction, and prediction.

In the realm of cultural heritage, the digital twin has evolved far beyond the notion of a static visualisation or isolated model. It now constitutes a complex informational organism, integrating diverse layers of documentation: historical texts, scientific analyses, sensory data, and curatorial knowledge. To preserve coherence across this multiplicity of sources, ontologies provide the foundational structure to build the memory of the digital twin, by serving as its semantic architecture and allowing knowledge to be expressed formally, queried effectively, and shared across systems and institutions.

Yet, structure alone cannot suffice. Just as Helen's *eidolon* required a divine breath to become operative, digital twins must be endowed with intelligence in order to act meaningfully. Artificial Intelligence, when guided by ontological frameworks, enables digital twins to transcend passive representation. Through the semantic interpretation of textual and visual sources, the detection of patterns in environmental data, and the simulation of complex behaviours, AI seems ideal to become the interpretive engine of the twin. In this framework, natural language processing, entity recognition, semantic inference, and predictive modelling are no longer isolated functionalities, but become part of a coherent system of knowledge and memory production.

It is precisely in the synergy between ontologies and AI that the true potential of digital twins can unfold. Their mutual interaction constitutes the epistemological core of a new generation of digital twins in which ontologies ensure that AI models are not only efficient, but also transparent, interpretable, and anchored in shared conceptual frameworks. In return, AI can enrich semantic information by identifying latent connections, disambiguating meanings, and populating knowledge graphs with contextually valid information.

The result is a self-expanding semantic system, capable of both reasoning and remembering, a system that learns, explains, and operates in alignment with the epistemic and ethical values of cultural heritage research. Together, ontologies and AI can be combined to give life to digital agents capable of supporting curators, scholars, and institutions in preserving not only the physical integrity of cultural assets, but also the knowledge systems through which these assets are understood and valued.

This paper builds upon the conceptual foundations introduced in our previous work on the Reactive Heritage Digital Twin (RHDT) and its associated ontology (RHDTO), offering a more operational perspective. It presents the design and implementation of AI-assisted semantic pipelines for the extraction, structuring, and reasoning over cultural information. By combining ontologies with intelligent processing components, it explores how to construct digital twins that are not only informative, but also sentient in their logic and responsive in their behaviour.

More specifically, the paper examines two complementary uses of Artificial Intelligence within this framework, situated at opposite ends of the semantic pipeline. On one side, AI is deployed for the extraction of knowledge from unstructured textual sources, in particular, heritage science reports, and its subsequent alignment with ontological structures, resulting in the enrichment of the digital twin's semantic graph with scientific information. On the other hand, the populated graph itself becomes a resource for training AI systems, enabling them to perform tasks such as querying semantic information through natural language or assisting in the extraction of new content. This virtuous cycle of knowledge generation and reinforcement creates a recursive architecture in which semantic data and intelligent agents co-evolve, gradually enhancing the digital twin's cognitive capacity. The goal is to show how this technological convergence can support the development of truly intelligent Heritage Digital Twins: systems that do not merely reflect culture, but actively participate in its documentation, understanding, and preservation.

The paper is structured as follows: Section 2 outlines the historical evolution of digital twins, particularly those enhanced by AI, from their origins in engineering and aerospace to their gradual redefinition within the cultural heritage domain. It traces the shift from technical replication to semantic modelling, and introduces the Heritage Digital Twin as a cognitively enriched, conceptually expressive construct. Sections 3 and 4 explore the uses of new Artificial Intelligence technologies to extract knowledge from cultural heritage documentation and encode it according to the semantics of ontological models widely used in this domain, to build the knowledge graphs of Heritage Digital Twins. Sections 5 and 6

investigate how knowledge graphs can be used to enrich the semantic toolbox of Artificial Intelligences, transforming them into agents capable of reasoning on large and intricate amounts of data, and answering complex scientific questions formulated in natural language. Section 7 presents a case study where the entire pipeline is tested on a set of scientific reports from the ARTEMIS project documentation, to identify and distil information used to populate the knowledge graphs that serve as the engine of the Heritage Digital Twins and is used to enrich and reinforce AI components, enhancing their ability to support advanced functionalities such as natural language interfaces, intelligent querying, and adaptive interaction with users. Following the discussion in Section 8, which highlights the potential of these technologies while also highlighting the inherent risks in their application, particularly when not subject to rigorous human oversight, the paper moves towards its final section, offering some conclusive notes and future perspectives.

## 2. Digital Twins: From Conceptual Models to Operational Systems

The relevance of digital twins has steadily expanded across disciplines, reflecting their growing role as sophisticated digital counterparts of real-world entities and systems [1,2]. While often loosely applied to any virtual replica, the term also designates complex, structured information systems capable of interfacing with the physical world. Initially developed to enhance industrial performance, optimise processes, and anticipate failures, digital twins have since evolved to support various domains for their ability to improve planning decisions [3], especially by integrating multi-source data [4]. The origins of digital twins are often traced to NASA's simulation technologies in the 1960s, which allowed engineers to replicate space conditions virtually—an epistemological necessity in the face of inaccessibility [5]. Although the term "digital twin" formally emerged only in 2010 [6], the conceptual lineage is deeper and more complex, as is often the case with terms that retrospectively colonise prior practices.

The applicability of digital twins has expanded dramatically since then. In manufacturing, for instance, digital twins underpin Industry 4.0 and its successor paradigms, enabling predictive maintenance, lifecycle optimisation, and human-machine collaboration [7].

In the construction sector, digital twins facilitate accurate simulation of design scenarios before physical implementation, reducing costs and enabling early detection of critical flaws. The BIM (Building Information Modelling) methodology, widely adopted in this sector, also offers a foundation for digital twins, albeit often limited by rigid data schemas and unidirectional workflows [8].

The scientific community has also embraced digital twins to build virtual laboratories, due to their ability to consistently anticipate failures, optimise performance, and support lifecycle optimisation based on real-time data flow between physical and digital components.

Urban applications, perhaps the most ambitious in scale and implication, leverage digital twins to model smart cities, integrating traffic, land use, energy consumption, and citizen feedback into coherent, semantically grounded infrastructures. These local digital twins (LDTs), increasingly interoperable and modular, promise to reform urban governance, though challenges of standardisation, scalability, and ethical oversight persist [9].

Even in more ethically charged domains, such as medicine, the digital twin paradigm is reshaping epistemologies of care. Patient Medical Digital Twins (PMDTs) integrate biometric, behavioural, and clinical data to simulate disease progression and treatment outcomes, offering personalised, dynamic, and privacy-preserving tools for preventive medicine [10]. These digital constructs function as adaptive avatars, supporting both clinicians and patients in decision-making.

This evolution is most visible in large-scale applications. The European Commission's Destination Earth initiative envisions a planetary-scale digital twin to simulate climate dynamics, incorporating not only physical systems but also social and behavioural data [11]. This human-in-the-loop paradigm reflects a growing recognition that digital twins are not neutral mirrors but value-laden constructs. They encode assumptions about relevance, causality, and intervention, and thus require governance frameworks that ensure transparency, ethical integrity, and stakeholder accountability [12].



Many attempts to formalise the digital twins framework [13] positioned the digital twin within a broader systems engineering perspective as a tool to mitigate uncertainty and monitor emergent behaviour in different environments. At this stage, the digital twin remained a construct rooted in mechanical logic, a surrogate model operating through numerical fidelity, maintained through sensor and IoT integration, and governed by technical control loops. Its primary value resided in its accuracy and reactivity. As the model has evolved, the digital twin has gradually ceased to be merely a static reflection and has become an active computational agent. It has begun to incorporate not only geometric and physical information, but also operational, behavioural, and even cognitive dimensions. As this evolution unfolds, it becomes clear that the digital twin can be employed not only as a monolithic solution but as a conceptual framework as well, open to reinterpretation across disciplinary borders.

### *2.1. Digital Twins and Artificial Intelligence*

The incorporation of Artificial Intelligence into digital twin technologies is ushering in a significant conceptual and operational shift. No longer confined to functioning as elaborate digital counterparts of physical entities, digital twins are now evolving into systems endowed with interpretive, adaptive, and semantically structured intelligence, capable not just of replicating but also of understanding and responding to the complexities of their real-world counterparts. This transformation marks a notable epistemological threshold, recasting the digital twin as an active agent of interpretation rather than a passive conduit of representation. It is AI that propels this transition, furnishing digital twins with mechanisms of learning, inference, and autonomous decision-making [7]. As such, digital twins are no longer mere digital projections but are emerging as computational systems capable of analysis, simulation, and strategic recommendation [14]. Evidence of this transformation is manifesting across a broad range of domains. In manufacturing, for example, AI-enhanced digital twins support predictive maintenance, improve workflow design, and enable real-time responsiveness by interpreting sensory inputs and operational histories. Factories thus begin to resemble adaptive environments, governed not solely by pre-programmed rules but by evolving patterns of interaction and feedback [15].

The healthcare sector also offers another compelling domain of application, where AI-driven digital twins are enabling personalised diagnostics, simulating pathological progressions, and refining therapeutic strategies. Initiatives have begun to explore the modeling of human digital twins for behavioural therapy and rehabilitation, while further extensions include pharmaceutical research and remote surgery. In these contexts, AI becomes the central enabler of systems capable of insight and adaptability [16].

In urban systems, the convergence of AI and digital twins is transforming how cities are modelled, understood, and governed. These intelligent frameworks are being applied to simulate environmental and infrastructural conditions—ranging from carbon emissions and heat distribution to traffic and pedestrian behaviour. Beyond passive monitoring, they enable predictive modelling and real-time adjustment based on the analysis of diverse data streams such as aerial imagery, energy consumption, and transportation flows. As a result, municipalities are empowered to enact more nuanced and anticipatory interventions [17,18].

Importantly, the symbiosis between AI and digital twins is extending its reach beyond industrial or scientific contexts, entering realms such as cultural heritage with promising, often experimental, applications. One particularly innovative avenue involves the reconstruction of lost or undocumented heritage through AI-based image generation. Here, AI is employed to synthesise visualisations derived from oral testimonies and collective memory, complementing tangible-based HBIM methods [19].

In heritage-related scenarios more broadly, AI integration into digital twins is already proving beneficial for documentation, conservation, and public access. Automation is expediting data collection, facilitating large-scale digitization, and enhancing accessibility via automated transcription and translation tools. At the same time, AI is processing real-time sensor data to detect threats and recommend countermeasures, increasing the system's responsiveness to environmental risks or structural

degradation. In this way, conservation scenarios can be virtually simulated, allowing interventions to be evaluated without exposing heritage objects to physical risk [20].

Despite these advances, the implementation of AI within digital twin ecosystems remains uneven and often constrained. Many applications persist in isolated pilot stages, lacking the infrastructural robustness or institutional momentum to scale effectively [21]. The interdependence of software, hardware, and human expertise highlights the need for deployment strategies that go beyond technical engineering to embrace socio-technical coordination [22].

Also on a conceptual level, the integration of AI into the epistemic fabric of digital twins remains underdeveloped. While machine learning models excel in identifying correlations and predicting outcomes, they often fall short in contributing to the construction of coherent knowledge systems [23]. Nevertheless, progress continues steadily, and the expectation is that AI will increasingly endow digital twins with the capacity to fulfil their epistemic and operational potential. Yet this trajectory raises critical concerns about transparency and information quality. The internal logic of many AI systems remains opaque, making it difficult to trace how conclusions are reached—especially problematic in domains requiring interpretive subtlety, such as cultural heritage, urban planning, or medicine. Here, data in isolation is insufficient; meaningful engagement demands systems that can not only reason but also explain. An overreliance on autonomous inference without appropriate mechanisms for interpretability risks introducing errors, distortions, or even ethical blind spots. This situation prompts a foundational question: how far can we entrust epistemic authority to systems incapable of articulating their own reasoning? Especially now that decisions once reserved for human deliberation start to be automated by systems that remain inaccessible to scrutiny, ensuring intelligibility and accountability is thus not a secondary concern, but a prerequisite for ethical and sustainable implementation [12].

## 2.2. Ontologies: illuminating Artificial Intelligence

In response to these fundamental questions, emerging approaches are foregrounding the importance of structured semantic architectures as a prerequisite for meaningful AI deployment. Rather than relying solely on data quantity or statistical modelling, these strategies emphasise the need for carefully curated knowledge frameworks, defining concepts, relationships, and ontological commitments, as foundational to advanced reasoning and explanation [24]. Ontologies, in this regard, are serving a dual role: they are enabling more nuanced and context-sensitive AI operations, and also acting as safeguards against the risks of misinterpretation or hallucination [25]. Additionally, ontologies are employed in the development of explainable machine learning pipelines, addressing crucial aspects like feature categorisation and metadata description, which are fundamental for the transparency of the machine learning process [26]. By embedding formal logic and expert-curated structure within digital twin systems, ontologies help ensure that AI outputs remain anchored in coherent and verifiable knowledge domains. This allows intelligent systems not merely to function, but to explain, interrogate, and evolve. In doing so, they contribute to the transformation of digital twins into epistemic infrastructures: semantic, reflexive, and dynamically coupled with the evolving landscape of human understanding [27]. This role of ontologies as a foundation for trustworthy AI is a crucial aspect, since they are able to provide more structured and explicit knowledge that enables AI systems to better articulate their reasoning. This extends beyond mere performance, addressing ethical concerns such as bias mitigation and transparency enhancement.

## 2.3. Towards the Cognitive Heritage Digital Twins

The Heritage Digital Twin we are designing and developing [20,28–30], is shaped as a point of confluence, a space where these diverse technologies do not merely coexist but actively intertwine. It is precisely at this intersection that ontologies, Artificial Intelligence, and semantic infrastructures converge to generate a new kind of cultural intelligence. Early applications of digital twins in heritage contexts largely focused on high-resolution 3D models, often derived from laser scanning or photogrammetry. These visual representations, while technically sophisticated, were semantically shallow.

They remained inert unless complemented by rich contextual metadata and interpretive frameworks, offering a simulacrum in the visual sense but lacking interpretive depth.

Rather than reducing cultural heritage to its visual or geometric likeness, the paradigm of Heritage Digital Twin, as we understand it, embraces the full constellation of digital documentation that relates to real-world cultural entities, whether movable, like artworks and artefacts; immovable, like monuments or architectural sites; or even intangible, such as rituals, traditions, and other expressions of cultural memory. This paradigm shifts the focus away from purely 3D-centered representations, proposing instead an holistic digital ecosystem that brings together visual depictions, textual accounts, scientific analyses, conservation records, and historical interpretations. It is a vision grounded in the belief that meaning arises not from isolated forms but from the interplay of all the fragments through which heritage is remembered, transmitted, studied, and understood.

Cultural entities, in fact, are not governed solely by physics, but essentially by memory, and cannot be fully described through geometry or physics alone, since they carry meanings, contexts, and symbolisms that resist reduction to quantitative attributes. As a result, the replication of an heritage entity does not have to rest on form alone but requires a cognitive and semantic infrastructure and the ability to articulate not only what an object is, but what it means, represents, and has undergone across time. As such, the digital twin in this domain demands not only a new technical infrastructure, but a rethinking of its very ontological assumptions.

At the heart of any meaningful digital representation of cultural heritage lies a conceptual operation as much as a technical one: the deliberate choice to model reality through shared ontological commitments. Cultural entities are not inert objects, they are constructed through layers of description, interpretation and transmission. If their digital counterparts are to achieve more than a superficial resemblance, they must be capable of reflecting this diachronic and interpretive richness. Ontologies provide the necessary framework for this task: a rigorous grammar through which cultural knowledge may be structured, preserved, and subjected to critical inquiry. By expressing information as entities and relationships, ontologies enable a deeper representation of the cultural object's life history, its transformations, and the meanings it acquires across time. In this framework, Artificial Intelligence can also reveal its full potential as a catalytic instrument capable of navigating, extracting, and reassembling meaning from the vast, fragmented archives of human knowledge, and at the same time, ingest and digest this knowledge to enhance its capabilities and serve as a privileged tool for exploring and interrogating the information embedded in the digital twin.

The culmination of this trajectory is thus the development of a new construct that unites the expressive power of ontologies with the interpretive capabilities of Artificial Intelligence in order to represent, reason about, and interact with cultural entities. Far from being a mere digital copy of its shape, the Heritage Digital Twin becomes an evolving semantic organism, an animated simulacrum (like Helen's *eidolon*) which not only has its external resemblance but reproduces the cultural object in its entirety. In its most advanced form, the Heritage Digital Twin becomes an agent (i.e., a cognitive entity) capable of dialogue, inference, and response within a meaningful digital continuum.

### 3. In Pursuit of Knowledge

#### 3.1. Retrieving and Reusing Cultural Knowledge

The conceptual framework outlined in the previous chapters finds an operative exemplification in the work presented in this and the following sections, aimed at illustrating how the convergence between ontological models and Artificial Intelligence may enable the development of intelligent (i.e., Cognitive) Heritage Digital Twins, capable of representing, interpreting, and reasoning over cultural data with increased semantic depth. This fruitful interplay unfolds through two complementary strategies: on one hand, ontologies and AI engage with the epistemic "archaeology" of cultural documentation, retrieving, structuring, and semantically articulating the dispersed and often latent knowledge embedded in visual and textual heritage records. On the other, they explore how this extracted knowledge, once formalised within a coherent ontological framework, can be reintroduced

into AI systems to enhance their capacity for informed reasoning, contextual sensitivity, and traceable inference. This experimental work constitutes a concrete instantiation of the broader paradigm outlined so far: the symbiotic integration of Artificial Intelligence and ontological modelling. Far from exhausting the spectrum of possibilities opened by this convergence, these studies exemplify one of its most immediately impactful trajectories, namely, the retrieval, organisation, and strategic deployment of knowledge embedded in cultural heritage documentation.

### 3.2. Natural Language Processing of Textual Documentation

Together with the broader macro-area of Artificial Intelligence, the subfield of Natural Language Processing (from here on referred to as NLP) has become one of today's most trending topics, especially thanks to the emergence of Large Language Models (LLMs), which have quickly become central to our everyday lives.

NLP focuses on enabling machines to understand, interpret, and generate human language. As simple as this may sound from an agnostic point of view, handling human language is one of the most difficult tasks a machine can undertake. Unlike programming languages, which are based on mathematical rules and are explicitly designed to be machine-understandable, natural language is not as easily interpreted or represented in numerical form.

What makes human language special, and inherently complex, are its peculiar features:

- Language is inherently ambiguous at all levels: lexical, semantic, and syntactic. While humans typically overcome this obstacle through contextual knowledge, computers often struggle to correctly grasp meaning, as context is frequently unspoken or culturally assumed
- Language is heavily context-dependent
- Language is creative: people frequently invent new words and phrases, and this process is always ongoing and can be virtually infinite
- Language is often ironic and metaphorical, making it difficult for computers to detect when meaning deviates from the literal or usual one
- Humans regularly leave information out, assuming the listener will infer it; however, computers often struggle with anaphora resolution and implicit references

These characteristics matter because they're deeply intertwined with the ways people form, narrate, aggregate and transmit their stories, culture, and heritage.

The nature and functioning of natural language are still not fully understood, even from a biological point of view, so it is only logical that most language-related tasks cannot be solved with simple, rule-based algorithms.

Over the past decade, the field of NLP has undergone a major transformation with the advent of deep learning and, more recently, Large Language Models (LLMs). These models are based on the Transformer architecture, which introduced the so-called attention mechanism:

*"Given a set of vector values and a vector query, attention is a technique to compute a weighted sum of the values, dependent on the query."* [31]

In particular, LLMs use a specialised version called *self-attention*.

In traditional attention mechanisms, each query maps exactly to one key-value pair, while in self-attention each query matches each key to varying degrees and then the returned result is a sum of values weighted by the query-key match. This mechanism allows the model to assign importance to different tokens (which are often single words) within a sequence, enabling it to focus on the most relevant parts of the input when generating output.

In addition to this, LLMs are trained on vast, multilingual corpora and have demonstrated impressive capabilities across tasks such as machine translation, question answering, summarisation, and semantic search. Their ability to perform complex tasks such as transfer learning, a machine learning technique in which the knowledge learned from a task is re-used to boost performance on a related task, few-shot generalisation, and contextual language modeling has opened new frontiers in computational language understanding, particularly in complex or previously underexplored text domains.



One such domain is cultural heritage, where textual artifacts may include ancient manuscripts, epigraphic inscriptions, archival inventories, scanned PDFs of scientific analyses, and handwritten letters. These materials offer rich but highly challenging content for computational analysis, often exhibiting historical language variants, inconsistent orthography, multilingual passages, and physical degradation, all of which complicate standard NLP workflows.

Recent efforts by computer scientists have focused on adapting LLMs and developing domain-specific NLP pipelines to address these challenges. Techniques such as fine-tuning on specialised corpora, integrating OCR (Optical Character Recognition) and HTR (Handwritten Text Recognition) outputs, and combining textual data with metadata or image features are increasingly used to support tasks like:

- Named Entity Recognition
- Semantic Enrichment
- Machine-assisted transcription
- Cross-document linking

As a result, state-of-the-art NLP is becoming an essential component in the digital preservation, accessibility, and scholarly analysis of cultural heritage texts.

### 3.3. Computer Vision and AI for Image Analysis

Alongside advancements in Artificial Intelligence, the field of Computer Vision has emerged as a transformative force in the digital analysis of visual materials, particularly within the domain of cultural heritage.

Computer vision refers to the ability of machines to interpret and extract meaningful information from images and visual data. While this may seem straightforward from a human perspective, visual understanding by machines remains a profoundly complex task.

This complexity is amplified in cultural heritage contexts, where visual artifacts are often heterogeneous, historically layered, physically degraded, or embedded within ambiguous symbolic systems. Cultural heritage datasets span a broad spectrum of visual media: paintings, frescoes, archaeological artefacts, cartographic materials, manuscript illuminations, early photographs, and architectural drawings just to cite some. These materials often suffer from age-induced damage, non-standard digitization, inconsistent resolution, and visual noise, all of which pose significant challenges for conventional image processing pipelines.

Additionally, cultural objects are increasingly accompanied by scientific imaging data such as X-ray fluorescence (XRF), hyperspectral imaging, infrared reflectography, and 3D reconstructions, offering new, modern layers of analysis. These multispectral and volumetric representations are crucial for revealing aspects and details that the human eye could not previously notice, like underlying hidden drawings, pigment composition, restoration traces, and structural anomalies. For example, XRF imaging has been applied to study underdrawings in Van Gogh paintings, while 3D photogrammetry and laser scanning are widely used to generate digital replicas of sculptures and monuments.

In recent years, breakthroughs in Deep Learning, particularly through Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) have enabled new approaches to visual analysis in heritage research.

These models support various tasks, such as object detection, image segmentation, style classification, digital restoration, and optical character recognition (OCR). When fine-tuned on domain-specific datasets, they have been used to automate the damage mapping of frescoes [32], identify iconographic motifs in religious art [33], and recognise ancient scripts and inscriptions [34].

These tools support scalable documentation and interpretation of cultural artefacts, particularly when integrated with 3D models and scientific imaging outputs. Building on these capabilities, a major recent development is the integration of vision and language models (Vision Language Models), multimodal AI systems capable of jointly analyzing images and text.

Models such as CLIP [35] (Contrastive Language-Image Pre-training) and its open-source variant

OpenCLIP [36] are trained to align visual and textual embeddings within a shared semantic space. This enables cross-modal retrieval, where a textual query (e.g., “15th-century illuminated manuscript with marginalia”) can retrieve relevant images even if metadata is sparse or missing. Conversely, images can be used to generate descriptive textual outputs, facilitating cataloguing and content discovery.

In cultural heritage, these vision-language models are being actively explored for:

- Semantic image retrieval across museum collections and digitized archives using natural language queries
- Automated captioning and metadata enrichment for unlabeled or poorly documented objects
- Cross-modal linking of visual and textual materials, such as aligning illustrations with manuscript passages or inscriptions with OCR outputs
- Clustering of visual-textual artefacts to detect stylistic, thematic, or geographic patterns

To further improve relevance in heritage-specific contexts, these models are often fine-tuned on curated corpora, including museum metadata, OCR-processed texts, and 3D model annotations, and are increasingly paired with temporal, geospatial, or material metadata.

The integration of scientific imaging modalities (like XRF, IR, or CT-scan slices) into multimodal pipelines is also beginning to allow models not only to see the surface, but to reason about what lies beneath, opening possibilities for computational material analysis and even automated conservation diagnostics.

As a result, AI-driven image analysis, now powered by multimodal vision-language models, is becoming a cornerstone in the digital transformation of cultural heritage.

These tools enhance the documentation, accessibility, and interpretation of visual cultural assets and support advanced applications such as semantic search, narrative reconstruction, damage tracking, and virtual exhibition design. In doing so, they allow scholars and institutions to reimagine how cultural knowledge is represented, retrieved, and recombined across diverse formats, periods, and platforms.

### 3.4. Identification and Semantic Extraction of Relevant Entities and Relationships

Another critical challenge in the digital processing of cultural heritage texts lies in the identification and semantic extraction of entities and relationships, a foundational step for building structured knowledge, supporting semantic search, and enabling automated reasoning.

Traditionally, this task has been addressed through Named Entity Recognition (NER) pipelines, which label predefined categories such as “person”, “location”, “organisation”, or “date” in unstructured text. Over the last decade, state-of-the-art NER systems have been greatly enhanced by transformer-based models like BERT and RoBERTa, showing high accuracy on well-structured, contemporary texts. However, these systems often struggle when applied to historical documents, multilingual corpora, or domain-specific ontologies, where named entities are ambiguous, infrequent, incompletely annotated, or entangled in complex relational structures (even though there are some successfully trained version of some specific domains, such as ArchaeoBERT<sup>1</sup>).

In the context of cultural heritage, standard NER often proves insufficient. Texts, such as archival inventories, epigraphs, marginalia, or curatorial notes, frequently reference entities in indirect, partial, or obsolete forms. Furthermore, the relevant concepts often extend beyond standard categories to include historical events, material properties, artistic styles, ritual roles, or geopolitical shifts, many of which are not well-covered by off-the-shelf NER models.

These limitations pose significant barriers to aligning extracted information with structured ontologies or linking it to a Knowledge Base.

To address and overcome this, we have adopted a large language model (LLM)-based approach allowing for more flexible, context-aware entity and relation extraction without the need of having a rigid dependence on pre-trained NER schemas.

<sup>1</sup> <https://huggingface.co/alexbrandsen/ArchaeoBERT>

Instead of classifying entities into fixed types, the LLM is prompted to identify relevant semantic units and their relationships in natural language, guided by domain-specific ontology definitions. This approach enables us to extract and match non-standard or composite entities and to associate them with structured entities in our custom Cultural Heritage Knowledge Graph.

This prompt-based method also facilitates the extraction of complex relationships, including part-whole, temporal, provenance, scientific analysis relations, which are essential for modelling cultural heritage information in ontological formats.

By skipping traditional NER in favour of in-context reasoning through LLMs, we can integrate multiple tasks: entity disambiguation, coreference resolution, attribute assignment, and relation typing, all into a single unified pipeline.

Our experiments show that this strategy significantly improves recall and semantic accuracy in entity-linking tasks when compared to traditional NER-based pipelines, particularly when dealing with fragmented historical records or low-resource languages. Moreover, the use of an open-source, locally deployable model ensures data privacy, customisability, and scalability, aligning with the needs of cultural institutions and digital humanities projects that require transparent, interpretable, and adaptable AI tools. The transition to an ontological encoding will then make this information unambiguous and ready to be embedded in the knowledge graph, and to then be incorporated in the Heritage Digital Twin.

## 4. From Fragments to Knowledge

### 4.1. Ontological Modelling of Extracted Information

The process of encoding extracted data to ontological structures is never a mere act of correspondence; it is a conceptual translation that requires both technical rigour and hermeneutic sensitivity. Information drawn from textual descriptions, archival records, scientific datasets, or visual analysis carries within it layers of context, intention, and uncertainty. To encode such data into an ontological framework is to interpret it through a lens that seeks formal clarity without erasing epistemic nuance.

Artificial Intelligence, particularly in its recent semantic incarnations, plays an increasingly central role in this mapping process. AI systems can identify entities, classify relationships, and extract candidate assertions from unstructured sources. Yet the true work of semantic alignment remains interpretive, demanding expert judgement in choosing the appropriate classes, properties, and event structures through which to articulate extracted content. This is one of the cases in which the synergy between AI and ontologies reveals its full potential, the former accelerating the identification of candidate knowledge, the latter ensuring that such knowledge is meaningfully situated within a coherent and reusable conceptual space.

### 4.2. Ontologies for Building Knowledge Graphs Ecosystems

From the above considerations, the ontological framework emerges as its essential scaffolding, the conceptual architecture upon which the entire system is constructed. Guided by this foundational premise, we have developed the Reactive Heritage Digital Twin Ontology (RHDTO), an extension blossomed from the solid foundations of the CIDOC CRM, as a means to organise, interconnect, and semantically sustain the heterogeneous data that form the informational substrate of cultural digital twins.

Among the various ontological frameworks developed for the heritage domain, the CIDOC Conceptual Reference Model (CRM) [37] is an internationally recognised standard (ISO 21127:2023). More than a model, CIDOC CRM constitutes an ecosystem that includes domain-specific extensions, such as CRMsci for scientific observation [38] and CRMdig for digital provenance [39], that offer a nuanced vocabulary capable of accommodating the multifaceted nature of cultural data. The adoption of CIDOC CRM offers a new path forward, disclosing the possibility to describe not just the form of an object, but the historical events, actors, techniques, and materials that constitute its biography. It is this ontological fluency, this capacity to speak the shared language of heritage knowledge, that renders

the CIDOC CRM ecosystem a cornerstone of semantic continuity across the heritage information landscape.

In some previous works, we have articulated the structural components of the RHDTO, delineating its main dimensions: the *documentary* one, which captures the breadth of scientific, historical, and multimedia documentation [28,29]; the *reactive* one, which models the operational mechanisms for real-world interaction through sensors, deciders, and actuators [20]; and the *intelligent* one, in which Artificial Intelligence is integrated to interpret, enrich, and activate knowledge [40]. Notably, in [40] we have also illustrated how AI can serve as the cognitive engine of the digital twin, supporting analytical reasoning and decision-making within key components such as the Decider and the Actuator, thereby enhancing the system's ability to respond to real-world conditions in a meaningful and context-aware manner. Each of these facets contributes to the evolving ecology of digital twins.

To operationalise these dimensions, the RHDTO introduces a set of classes and properties designed to model the multifaceted nature of heritage entities and their digital counterparts with semantic precision. At the core lies the class *HC1 Heritage Entity*, a general abstraction encompassing both tangible (*HC3 Tangible Aspect*) and intangible (*HC4 Intangible Aspect*) dimensions of cultural heritage. The digital twin itself is represented by *HC2 Heritage Digital Twin*, conceived as a structured network of informational components, including digital reproductions, documentation, and the activities underpinning their creation and management.

One of the major strengths of the RHDTO ontology lies in its seamless integration within the broader CIDOC CRM constellation of specialised models, that allows it to function as a connective tissue between datasets that originate from heterogeneous contexts. This conceptual harmonisation permits the assembly of coherent knowledge graphs without sacrificing domain specificity. Another notable feature of the ontology is the introduction of a dedicated set of classes and properties, like the *HC15 AI Component* and the *HC16 Simulation and Prediction*, specifically designed to model AI components and their role in knowledge extraction and decision-making processes. These allows the modelling of active agents within the digital twin architecture, particularly in key processes such as data interpretation, scenario simulation, and real-time activation management.

RHDTO, CIDOC CRM and its extensions provide the necessary paraphernalia for such knowledge transformation and transcoding, turning textual evidence and visual data into computable knowledge while preserving their contextual richness. This is particularly relevant in domains like heritage science, where the use of these conceptual tools allows for a fine-grained articulation of experimental procedures and datasets, as will be illustrated in Section 6. This semantic articulation reshapes the cognitive architecture through which heritage information is organised, shared, and understood, transforming the ontological framework of the knowledge graph in a philosophical intermediary between language and logic, between human interpretation and machine understanding [41]. This way, the digital twin is endowed with a cognitive dimension. It no longer merely stores or displays data but interprets and internalises it within an ontological order that mirrors, albeit partially and provisionally, the complex web of human understanding.

## 5. From Knowledge to Intelligence

### 5.1. Using Knowledge Graphs to Build Knowledge-Enriched AI Agents

Once knowledge crystallises in the form of semantic graphs, it can be used to nurture and enhance AIs. The complexity of cultural heritage data demands the use of intelligent systems that can reason over both explicit content and the nuanced context behind it. Knowledge graphs, for this purpose, have emerged as a powerful tool to build knowledge-enriched Artificial Intelligence systems. Information is structured preserving its rich semantics and remains machine-interpretable, which enables AI systems to achieve context-aware understanding and inference instead of limiting themselves to surface-level pattern recognition [42,43].

Cultural heritage tells a story, and as stories are composed of interrelated entities, they must be represented in such fashion. A knowledge base with its nodes and edges, perfectly captures this web



of elements, grounded in formal ontologies[43,44], and CIDOC CRM and its extensions play a crucial role in offering a shared conceptual schema that aligns historical data with contemporary semantic technologies [43,45,46].

In recent years, there has been a growing emphasis on reinforcing the reasoning capabilities of LLMs by integrating them with external knowledge structures, particularly knowledge graphs. While LLMs such as GPT, BERT, and Mistral [47–49] have demonstrated remarkable abilities in generating fluent and plausible language, they remain prone to hallucinations, generating incorrect or misleading content especially in high-stakes domains like healthcare or science [50].

These failures often stem from limitations in their training objective, typically maximizing the log-likelihood of the next token without robust grounding in factual correctness.

To address this, researchers are increasingly integrating LLMs with knowledge graphs, which encode structured, semantically rich, and traceable factual knowledge [51].

Unlike opaque neural representations, knowledge graphs enable models to reason over explicit entities and relations, align responses to verifiable facts, and offer interpretability via provenance metadata [52,53].

This integration allows models not only to retrieve precise domain-specific information but also to construct explainable inference chains, thus improving transparency and trust.

Moreover, the inclusion of knowledge graphs supports granular and contextually relevant knowledge retrieval, outperforming naive augmentation with random or noisy data [54]. By aligning entities and relationships from LLM outputs with ontologies and knowledge bases, systems can avoid shallow text pattern recognition and instead achieve conceptual generalisation and symbolic reasoning. This approach also opens the door to continuous learning pipelines, where AI outputs are validated against, or enriched by, structured data and subsequently reintroduced into the model in the form of feedback loops. These loops allow AI systems to iteratively refine their knowledge representations, adjust for emerging information, and correct previously observed errors. Combined with recent advances in multimodal and temporal input integration—e.g., linking structured data with images or spatial metadata [55]—the use of knowledge graphs is becoming central to the development of knowledge-enriched, context-aware AI systems that are both accurate and adaptable. As structured data continues to permeate all domains of life, designing robust mechanisms to represent and integrate this data into LLMs remains a crucial and active research frontier.

### *5.2. Implementing Feedback Loops for Continuous AI Models Improvement*

It is worth noticing at this point, that for any system that deals with rich cultural knowledge, human expertise remains crucial, even as AI models become better at processing, interpreting, and retrieving information from both structured and unstructured data. Human input is not just optional for AI-assisted querying of cultural heritage knowledge graphs; it is vital for the system's reliability, interpretability, and contextual relevance [56,57].

A feedback loop is a cyclical process in which the predictions or outputs of an AI model are evaluated by automated systems, human experts, or both. The results are then used to update the training data or refine the model over time. Feedback can occur at various levels in knowledge graph-powered semantic pipelines, from retraining models with additional data to improving inference rules and fixing mislinked entities and classification boundaries [56,57].

In practice, feedback loops in heritage AI systems could involve curators, researchers, or subject matter experts reviewing recommendations from the AI, such as suggested connections in a digital exhibit, suggested classifications, or inferred relationships between objects. These users can record their choices as valuable feedback and can accept, reject, or modify the AI-generated recommendations [42,56]. This feedback is turned into structured input that can enhance the underlying knowledge graph or fine-tune the model while also being used to evaluate performance metrics.

Creating user interfaces and workflows that allow for smooth user interaction is key to implementing feedback loops effectively. For example, users may see ranked recommendations for linking recently digitized content to existing entities in the knowledge graph within a digital heritage twin platform.

The system translates human insights into actionable learning signals when users edit or reject the AI's suggestions. This information is captured as labeled training data. The backend of the semantic pipeline can include modules for continuous or active learning, which allows the AI model to be retrained on new datasets that consider user input. Active learning techniques can help prioritise the most impactful or unclear cases for human review, making expert time more effective [56,57]. Furthermore, changes to the knowledge graph, such as newly verified links or updated entity types, can disseminate throughout the pipeline, improving future predictions by providing clearer contextual grounding.

Error analysis is an important part of feedback loops. Developers can identify model blind spots or weaknesses in the training data by investigating systematic errors, such as frequent incorrect classifications or concepts that are not linked. Besides model retraining, these insights can guide improvements to user instructions, ontology design, and data curation practices. From the perspective of knowledge engineering, feedback mechanisms also contribute to the development of the knowledge graph. When users challenge inferred facts or offer alternative interpretations, they correct the AI and modify the semantic framework it uses. This ultimately leads to a more accurate and representative knowledge base that aligns with the evolving goals and understanding of the heritage community [56].

## 6. Natural Language Access and Prompt Interfaces

As the scale, scope, and complexity of data increases, the need to simplify its querying methods does so equally. Past a time where only select individuals could access certain information thanks to their expertise in the topic and research skills, information is now at everyone's reach, a half-thought and a click away. Not only archeologists, scientists, or other specialists may wish to browse the knowledge base, so in the context of Heritage Digital Twins, traditional querying methods would rather prove obsolete in the face of approaches utilizing natural language as a medium [58].

If a user expresses their inquiry in their own language, the translation layer—natural language understanding (NLU)—bridges the gap between human intention and structured semantic content [59]. The query gains a certain level of interpretability, relaxing constraints and allowing for fuzzy reasoning during retrieval across both structural and semantic levels [60]. Not only does this present an advantage in query guarantee (retrieving at least one result), but also provides users enough feedback to adjust and shape their queries and best attract the answers they seek within the intricate knowledge graph.

At the core of this capability is a combination of tasks including semantic parsing, entity linking, intent classification, and relationship inference, to decompose and interpret the natural language queries [61]. Thanks to recent advances in prompt-based large language models, using separately trained components for each of these tasks is no longer a requirement, as LLMs offer an integrated, zero or few-shot approach to parsing natural language [62]. In this research work, a Mistral 7B LLM was utilised for this task. In the context of cultural heritage, where user queries are often imprecise, explanatory, and laden with historical, geographical, or culturally specific terms, properly instructed LLMs can produce a variety of structured outputs [42]. The LLM prompt construction includes the general context of the query to focus the interpretation, indications on the what and how of information extraction, and any additional enrichment LLMs should apply to its output. As such, the latter can be in the form of data triples—the building blocks of a graph structure [60].

Intuitively, these triples would be converted into SPARQL fragments to filter the knowledge graph data by mentioned entities and layer those results according to the specific query relationships. The SPARQL query, then, may only be constructed after correct assignment of ontology labels and properties to the processed query elements—a task which would either require the involvement of an expert for labeling and validation or the implementation of an additional layer of entity and relationship classification using an ontology-trained AI model [56]. However, the usage of multiple hierarchical ontologies, one more complex than the other, renders this task more complicated than not, and while the latter

option is more viable, the rigidity of SPARQL querying remains a risk; in the case of misclassification, the query will yield no results although they might exist [63]. To address this, we adopt a hybrid query execution that bypasses strict SPARQL-based querying and instead leverages graph neural networks and knowledge embeddings, allowing the system to execute natural language queries in a semantically flexible and structurally tolerant way, specifically via Composition-based Graph Convolutional Networks (CompGCN) [61].

Following the NLU stage, the nodes represent candidate entities, types, or conceptual elements (e.g., “cathedral”, “bronze”, “Charles II”), and edges represent inferred relations or semantic dependencies (e.g., “located\_in”, “issued\_by”, “made\_of”). The constructed query graph represents the user’s intent directly, without prematurely enforcing alignment to any specific ontology schema, while capturing its semantic and relational structure. The execution of the query, therefore, is done through the embeddings created using CompGCN for both the knowledge graph and the query graph in the same latent embedding space [60,61]. During inference, candidate subgraphs in the knowledge graph are scored against the query graph, and the top-ranked matches are returned as results. This replaces symbolic matching with vector-space similarity scoring, enabling multi-hop reasoning across event chains and indirect relationships for flexible and scalable retrieval, even in the event of incomplete data, missing links, or conceptual variation [42].

### 6.1. Similarity Computation in the Context of the HDT

For Heritage Digital Twins, similarity extends beyond the simple attribute matching, but rather enters the world of semantic, structural, and contextual equivalence. It is rare—if at all possible—for artefacts, historical and archeological remains, or intangible heritage to be identical. Instead they show overlap in historical context or provenance, form, function or symbolism; all types of data that are hosted and structured in the knowledge graph according to the RHDTO [64]. To identify and evaluate this contextual equivalence, there is a need for a system that can exploit the heterogeneous, relational and temporally dynamic data available.

Our system uses vector embeddings to compute similarity between cultural entities, taking in consideration temporal alignment, relational proximity and typological affinities. These computations involve a combination of two measures between the embedded graph structures: the cosine similarity which works by calculating the cosine angle between two vector representations (though it is generally used for high dimensions dense vector comparisons), and the contrastive distance learning, which improves the clustering of semantically similar subgraphs and disperses the unrelated ones [65,66]. For tangible heritage, similarity can be computed at the following levels for example:

- Pieces that belong to same object
- Different objects that were excavated at the same site but from different locations
- Different objects that were crafted in the same workshop
- Different objects made of material originating from the same place; inferred from the data of scientific analysis reports
- Different objects made by the same artist

This demonstrates the great advantage of similarity computation, since most of these results would not be retrieved by a basic querying system alone. In fact, any of these objectives requires reliance on visual input [67] such as images or 3D models, but using intelligent AI models, it may be possible in most cases to identify such similarities just based on metadata stored within the knowledge graph [68].

### 6.2. Design and Implementation of Prompt-Based Interfaces

Heritage Digital Twins are powerful archives, and along with the extensive effort to improve the method of how they are queried on a logical aspect, we must also interface these improvements to the users. The approach we chose aims to allow natural language as a querying medium, and as such, we built a prompt-based user interface instead of the traditional, rigid hierarchy of filters or drop-down UIs. The concept is that of a chat environment for users to interact with the knowledge graph in a

conversational manner. A key design consideration has been quality feedback and transparency. The model response must be synthesized and clear, allowing the user to understand how their prompt was parsed and interpreted. This is most valuable when the user's query is ambiguous or lacking specificity. For example, a user query saying: "Find me Greek figures from the classical period" would yield the response: "Finding Greek historical personalities from the classical period. Would you like me to expand my search to mythological entities or sculptures and statues?"

This leads us to another important element; iterative refinement. Following the model's reasoning during the query would let the user, therefore, correct, or refine the system's interpretation of their intent. This requires the maintenance of a persistent conversation context that the system can always refer back to during the validity of dynamic session state; stored in an in-memory session store with a unique identifier passed with each message, after each of which, the prompt context and history are updated and re-evaluated during the LLM convocations. The implementation of this feature further guarantees accessibility across expertise levels, as even unstructured queries can be evolved into their most effective form through a co-constructed exploration [69]. A well-rounded response is also often appreciated by non-experts if it provides additional information that might entice further interests in the topic at hand. For this, the LLM is also tasked to include adjacent context, or broader data summaries, informed by the knowledge graph into its response. This can be synthesised from the subgraphs scoring slightly lower in similarity against the query graph, and should be clearly specified as such [69].

### 6.3. Presenting Complex Query Results in User-Comprehensible Formats

When dealing with large knowledge graphs such as the HDT's, it is crucial to avoid clutter, filter through data, and present only the salient information initially. Progressive disclosure balances usability with transparency for the sake of both casual exploration and scholarly validation. Our current implementation of the UI presents the system's response in a narrative format describing in plain text its findings, structured, listed, and categorised as needed by the LLM for better digestion of the information. Text-based responses are suitable for any type of queries, from precise search to exploratory inquiries. They can be concise and direct, but can also weave together complex temporal and geographic information, benefiting from explanation and nuance to paint a holistic picture for the user [70]. Not to suggest that text-based narration is the only format supported in chat-UIs; on the contrary, an important advantage is that they are extensible. In the context of HDTs, more structured or even visual elements would render the response more impactful, and thus the interface can be upgraded to include 3D model displays, images and illustrations, geographic and temporal maps, tables, diagrams and raw graph structures, depending on the user's needs [70,71]. Ultimately, presenting complex query results isn't just a matter of UI feature design but a mediation between semantic accuracy and human cognition.

## 7. Case Study: Enriching Digital Twins with Heritage Science Data

As a testbed for exploring the integration and synergy between AI techniques and ontological modelling, we have selected some reports generated by analysis activities in the heritage science field, part of the documentation of the ARTEMIS initiative [72].

### 7.1. Application of AI-Driven Semantic Pipelines

A modular AI-based pipeline was designed to extract structured semantic knowledge from unstructured scientific texts belonging to the cultural heritage domain, with a focus on modern preservation and restoration techniques such as, but not limited to, X-Ray Fluorescence (XRF) and 3D modelling techniques.

The system is particularly tailored to support the enrichment and population of an ontology-driven Knowledge Graph designed to be an efficient base for the construction of Heritage Digital Twins, but it could also be possible to use it for other scenarios.

The defined methodology combines several state-of-the-art open-source technologies, including Nat-



ural Language Processing (NLP), Large Language Models (LLMs), semantic similarity search, and ontology-based reasoning.

The pipeline transforms raw textual content—taken, as stated in the previous section, from documentation of the ARTEMIS initiative—into machine-interpretable knowledge graphs in the form of RDF triples, fully aligned with standard cultural heritage ontologies. Images from PDFs were also extracted and kept separately for further experiments.

Text and images extraction has been performed with PyMuPDF<sup>2</sup>, a Python library able to handle even complex textual layout. The following phase involved some text preprocessing, to reduce noise and generally improve the handling of textual data by machines. SpaCy<sup>3</sup> library has been used to perform:

- Tokenization
- Sentence Segmentation
- Syntactic Parsing

To enable efficient processing by Transformer-based LLMs, the texts were then segmented into smaller, semantically coherent units called "chunks", each of which contained approximately 300 tokens. This size has been chosen to balance the model's context window constraints with our need to have meaningful textual context.

This segmentation allows the model to focus on localised contexts, ensuring that the extracted knowledge is specific and grounded in the source material.

Following segmentation each chunk was passed to an instance of the Mistral 7B Instruct<sup>4</sup> model. The choice of using Mistral 7B Instruct has been made both thanks to its performances and the fact that the model is open-source, ensuring full transparency and reproducibility of the model's decision-making process. Moreover, by leveraging the capabilities of instruction-tuned language models, we were able to perform zero-shot term extraction, without the need for extensive supervised training data to make it work.

The prompting strategy used explicitly asked the model to extract, from each chunk of text, relevant entities and concepts taken from the domain ontologies, with a focus on modern scientific analysis techniques, materials, instruments, and processes. These prompts were designed to elicit concise, list-form outputs containing only the terms of interest, without additional explanations or context, in order to have a list of potential candidate terms.

Once all candidate terms were extracted, they were normalised, including digit and punctuation removal and whitespace normalisation, to ensure terminological consistency. In addition to this, in order to avoid semantic redundancy, a semantic similarity strategy based on SentenceTransformers has been set up.

All terms were encoded using all-MiniLM-L6-v2<sup>5</sup>, a pretrained language model that creates dense vector embeddings of sentences. After the encoding part a cosine similarity matrix was computed, which allowed the clustering of terms that had a similarity score above a chosen threshold (in this case the value was 0.85). This process goes a step deeper than the usual string-based deduplication, allowing also to collapse variant spellings and synonyms into unified conceptual terms.

A following interactive validation step was also implemented, allowing domain experts to manually select or discard terms prior to finalisation. This hybrid human-in-the-loop design is crucial to ensure both automation and accuracy while trying to avoid hallucinations and still have some control over the process.

A key innovation of the developed pipeline lies in the retrieval of semantic context for each validated term: while processes like Term Extraction or Named Entity Recognition successfully provide a list of relevant concepts, their meaning is often best understood in relation to the text from which they originate, and this is especially true in our case where the data needs to be mapped following domain

<sup>2</sup> <https://pymupdf.readthedocs.io/en/latest/>

<sup>3</sup> <https://spacy.io/>

<sup>4</sup> <https://mistral.ai/news/announcing-mistral-7b>

<sup>5</sup> <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

ontologies rules. To this end, we used FAISS<sup>6</sup> to identify the most contextually relevant text chunks for each ontology term.

FAISS is an open-source library for efficient similarity search and clustering of dense vectors, optimised for high-dimensional data, such as embeddings from neural networks. The way it works is by returning *k nearest neighbors of a query vector*, typically with their indices and distances/similarity scores.

To use FAISS effectively both of the refined list of terms and the original text chunks must be re-encoded using the same SentenceTransformer model. These embeddings are then used to construct a FAISS index that allows for efficient retrieval using dot product similarity.

For each ontology term embedding the index is queried to retrieve the *top-k most semantically similar* text chunks, which enables the recovery of the local textual environment in which each concept appears, a crucial step for accurate knowledge representation in the next stage.

With the contextual chunks retrieved, Mistral 7B is used once again, this time to perform RDF triple generation. The model is prompted with the retrieved text and instructed to extract concise semantic triples that align, both in structure and semantics, with existing cultural heritage ontologies.

The final outputs of this AI assisted pipeline include:

- A validated list of ontology-relevant terms, stored in JSON format
- A collection of RDF triples, structured and domain ontology-aligned
- A mapping between each term and its original context, ensuring interpretability and traceability

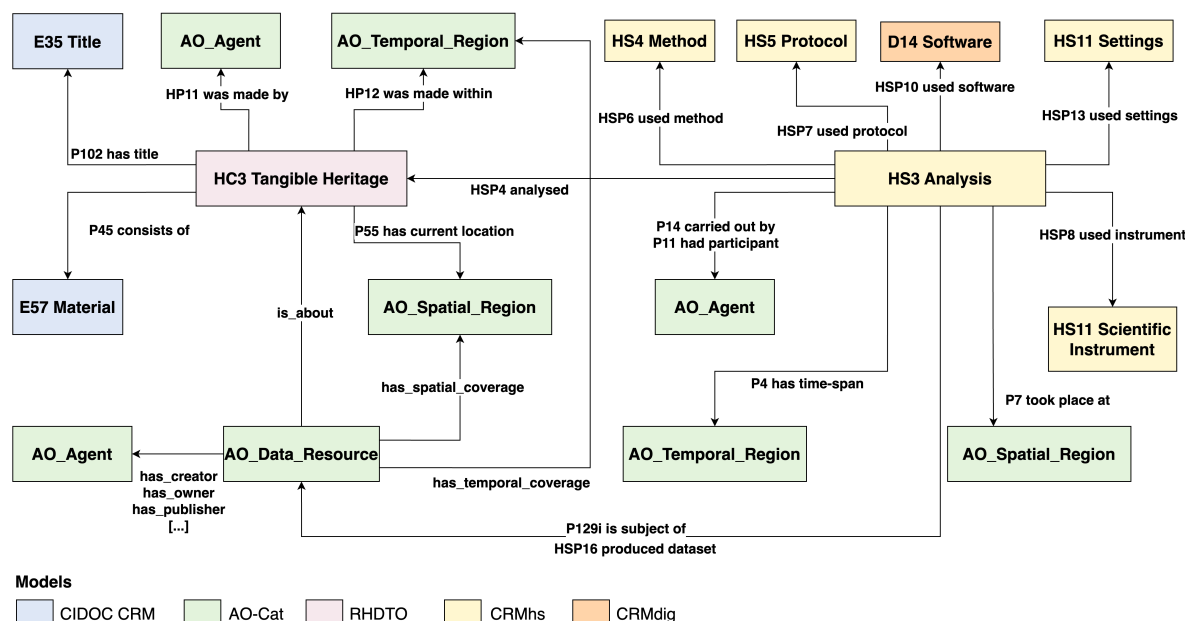
The underlying conceptual model chosen for this effort is CRMhs [73], an ontological extension of CIDOC CRM developed to represent the documentation and operational structure of scientific analyses carried out on cultural heritage objects. As an extension of the CIDOC CRM, CRMhs integrates seamlessly with RHDTO and gives it the ability to formalise the domain of heritage science by introducing specific classes and properties that describe scientific activities (such as *HS3 Analysis*), instruments (e.g., *HS11 Scientific Instrument*), datasets, sampling procedures, and their relationships to both the physical object under investigation and the broader research context. It supports the representation of tangible and intangible entities involved in the scientific process, such as samples, areas of interest, methods employed, and results obtained. The model also accounts for the configuration of complex devices, their software components, and even the parameters used during analysis. Moreover, since the ARTEMIS ontological architecture is intentionally modular, reflecting the diversity and complexity of the domains it seeks to represent, the project also employs the AO-Cat ontology [74] developed within the ARIADNE initiative [75] to describe datasets and their specific characteristics, including provenance, access conditions, and content typologies. Our encoding experiments, also using AO-Cat to encode these entities, have fully demonstrated how these models are able to interact perfectly with each other, allowing the construction of semantic graphs that, despite being internally heterogeneous, constantly maintain a solid semantic coherence.

A general overview of the ontological module designed by ARTEMIS to model scientific data is shown in Figure 1.

By anchoring AI outputs to a domain-specific ontology, the system ensures that extracted data are both semantically coherent and interoperable. This ontological framework thus provides the structure for an AI-assisted information extraction workflow aimed at transforming heritage science reports into semantically enriched data. Using the categories defined by this semantic ecosystem as guidance, natural language processing tools are applied to identify and annotate references to analytical techniques, measurements, instruments, and related entities. The extracted knowledge is then converted into structured triples, for instance, in Turtle format, encoded using the grammar provided by the ontological model employed, and ingested in the knowledge graph.

The following examples show some results of the pipeline tested on a paper by Mazzinghi et al. [76] about the use of Macro X-ray Fluorescence Scanning on a Raffaello's Portrait of Leo X. From the

<sup>6</sup> <https://faiss.ai/index.html>



**Figure 1.** Classes and properties used by ARTEMIS to model heritage science information.

starting point of the original PDF we were able to successfully and automatically extract the json file containing the triples, encode it in Turtle format and ingest it into the ARTEMIS knowledge base to visualise the graph. A fragment of the Turtle encoding (below) shows the ontological representation of Raffaello's painting as a cultural object by means of the *HC3* class of the RHDT Ontology.

```
artemis:LeoX_Portrait_Study_Object a rhdto:HC3_Tangible_Entity ;
  rdfs:label "Portrait of Leo X" ;
  owl:sameAs <http://www.wikidata.org/entity/Q5597> ;
  crm:P102_has_title "Portrait of Leo X with Cardinals" ;
  rhdto:HP11_was_made_by artemis:Artist_Raffaello ;
  crm:P45_consists_of "Oil on wood" ;
  rhdto:HP12_was_made_within artemis:LeoX_Portrait_Period ;
  crm:P55_has_current_location artemis:LeoX_Portrait_Place .
```

The following Turtle fragment, instead, shows the encoding of the analysis event conducted by scholars of the Italian Institute of Nuclear Physics (INFN) on the same cultural object, and the way in which CRMhs properties are used to specify the instrumentation and methodology employed, and the location and dates on which the analyses were performed.

```
artemis:Analysis_MA_XRF_LeoX a crmhs:HS3_Analysis ;
  rdfs:label "Macro X-ray fluorescence analysis" ;
  crm:P2_has_type "MA-XRF" ;
  crmhs:HSP1_has_activity_title "MA-XRF analysis on Leo X Portrait" ;
  crm:P3_has_note "Macro X-ray fluorescence analysis" ;
  crmhs:HSP4_analysed artemis:LeoX_Portrait_Study_Object ;
  crmhs:HSP6_used_method artemis:XRF_Method ;
  crmhs:HSP8_used_instrument artemis:Device_INFN_CHNET_Scanner ;
  crmhs:HSP9_used_component artemis:Component_XRay_Tube_Moxtek ,
    artemis:Component_SDD_Amptek ,
    artemis:Component_Telemeter_Keyence ;
  crmhs:HSP10_used_software artemis:INFN-CHnet_Software ;
  crmhs:HSP13_used_settings artemis:Settings_001 ;
```

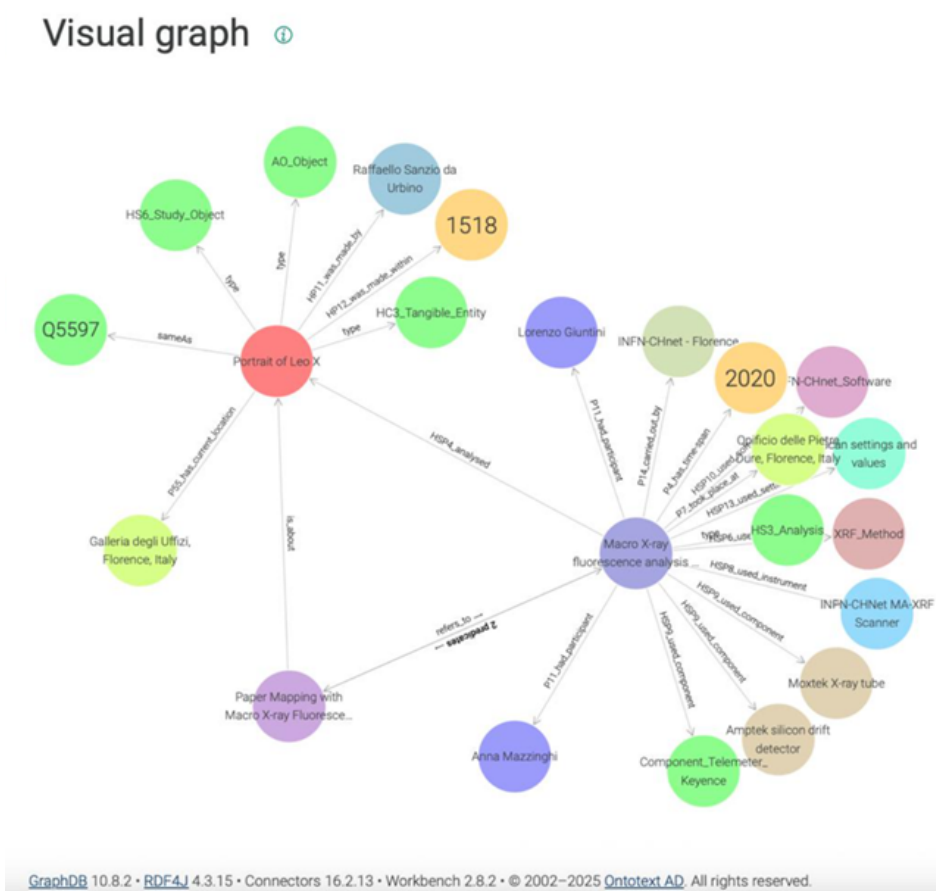
```

crmhs:HSP16_produced_dataset  artemis:XRF_LeoX_Resulting_Paper ;
crm:P14_carried_out_by  artemis:Group_INFN ;
crm:P11_had_participant  artemis:Person_Lorenzo_Giuntini ;
crm:P4_has_time-span  artemis:Analysis_Time_Frame ;
crm:P7_took_place_at  artemis:Analysis_Place .

```

These outputs can be ingested into Heritage Digital Twin Knowledge Graphs, enabling advanced features such as semantic querying, dynamic visualisation, and automated knowledge integration from heterogeneous scientific sources, and the subsequent feeding of AIs for advanced data analysis and interactions.

A visual representation of a section of the Digital Twin Knowledge Graph enriched with semantic context through our AI assisted process is shown in the Figure 2 below.



**Figure 2.** Graph visualisation of a scientific analysis of a cultural heritage asset directly obtained from the PDF through our AI assisted pipeline

### 7.2. Populating and Querying the ARTEMIS Knowledge Graph

Once the HDT knowledge graph has been built and the cultural heritage data is semantically structured, the next step is implementing the querying system which allows to access it. This process involves a few steps beginning with an additional preprocessing stage. This time, we are not preprocessing textual documents, but rather preparing the stored data for artificial intelligence model training. The ontology-based graph structure resulting from the preceding work maintains the information in a parametrised way, linking all attributes through property edges, including the node types (e.g., by means of the *P2 has type* property of CIDOC CRM). It is good practice for data archival, consistency and reusability, however, it poses particular challenges for an AI model, as then the nodes don't encapsulate enough meaningful information to distinguish them from others. To address this, the knowledge graph is restructured so that each node in the graph corresponds to an instance of a

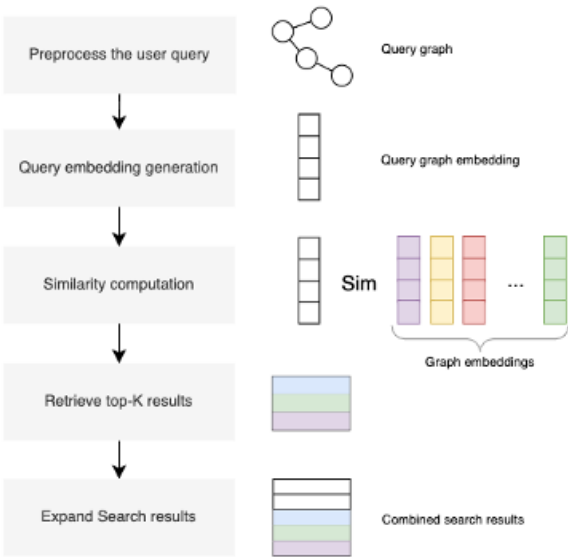


RHDTO class. Therefore, we embed the type of each entity in its node as a static feature, along with any human-readable label, title, or description it is linked to; for all of which simple embeddings were created using SentenceBERT. An example of the resulting node in JSON format is the following:

```
{ "id": "obj_0001",
  "label": "Mona Lisa",
  "type": "E22_Man-Made_Object",
  "features": {
    "label_embedding": [0.23, 0.78, ...],
    "type_embedding": [0.12, -0.09, ...] }
```

The next step of the process is to feed the new knowledge graph to an AI model and begin training. Though, grounding the knowledge graph in the rich RHTDO makes it too complex for standard neural networks to capture its nuanced relational structure of data. The model we have selected, CompGCN, is a state-of-the-art method for learning multi-relational graph representations, because unlike traditional graph networks that see edges as but a structural link between two nodes, CompGCN explicitly models both nodes and relations jointly. For cultural heritage data where each connection has a semantic meaning that depends on the nature of the relationship, it is critical to be able to make a difference between—for example, whether a person *created* an object, *participated in* an event, or *curated* an exhibition [77]. CompGCN looks at each of these relations as its own learnable element and trains to learn their differences. It also accounts for the directionality of relationships; in the knowledge graph, many relationships are inherently directional and semantically loaded as there is an asymmetry in reading them (e.g., “was influenced by” and “influenced”). The network learns this by identifying the roles of the relations and the flow of information between nodes [77].

Implementation-wise, the graph is transformed into triple format, and each relation is modelled explicitly with its inverse counterpart. CompGCN then uses a message-passing mechanism that jointly embeds nodes and relations to spread information throughout the graph during training. Node embeddings are updated at each graph convolution layer according to the unique semantic composition of the connecting relation and the characteristics of their neighbors. The model uses composition operators on the embedding vectors to integrate relation semantics into the neighborhood aggregation, which allows the network to model logic and semantics of the paths traversed additionally to their proximity [77]. After passing two or three layers of graph convolution, the node embeddings encode rich enough structural and semantic information to perform tasks such as similarity computation: once the user query is transformed into a graph structure with the LLM, it is passed through the same CompGCN encoder to generate a comparable embedding in the same latent space as that of the knowledge graph. Cosine similarity is then computed between this query embedding and candidate entity or subgraph embeddings from the knowledge graph, to find semantically relevant matches even in the absence of exact alignment (the candidates are scored in an interval of 0 to 1; 1 signifying high similarity). Figure 3 visualizes this process into a simplified diagram.



**Figure 3.** Diagram of the querying process using CompGCN embeddings and similarity computation

A successful query later, we obtain a sorted list of high-scoring entity or subgraph candidates for the given user prompt, and now the task is to present it to the user in a human-readable and layout. First the LLM interprets the result data searching for ways to highlight important information, categorize the entries, or provide contextual summaries where better suited, then returns it to the user in narrative format as seen in Figure 4.

What would you like to know?

Find information about coins minted in Famagusta

Send

💡 HDT's Response:

Famagusta, Cyprus.

### Coins of the Lusignan Kingdom of Cyprus (1192-1489 AD)

During the Crusades, Famagusta was part of the Lusignan Kingdom of Cyprus. The Lusignans issued gold coins known as Gros and silver coins known as Deniers. Below are some examples:

#### Gold Gros (Cyprus, 1376-1401 AD)

The obverse of the gold gros features a bust of Saint Christopher holding the Christ child on his shoulders. The reverse shows the Lusignan royal arms with three fleurs-de-lis and two lions rampant.

#### Silver Denier (Cyprus, 1376-1401 AD)

The obverse of the silver denier depicts a bust of Saint Catherine of Alexandria with a palm branch and a wheel in her hands. The reverse displays the Lusignan royal arms as on the gold gros.

### Coins of the Venetian Republic (1489-1571 AD)

After the fall of Famagusta to the Ottomans in 1571, it became part of the Venetian Republic. The Venetians issued coins with various designs for different regions under their control. Below are some examples specific to Famagusta:

#### Silver Grosso (Famagosta, 1492-1503 AD)

**Figure 4.** Snapshot of the functional chat-UI demonstrating a user prompt and the query results in narrative format

## 8. Discussion: Taming the Wild Beast

The hybrid querying approach described above, clearly outperforms strict SPARQL translation in several important dimensions. First, it offers greater flexibility when dealing with user queries that are incomplete, underspecified, or expressed in natural language, such as "old buildings with mythological carvings", where formal ontology mapping would struggle or fail entirely [78]. The system provides meaningful results even in the absence of precise predicates or labels by interpreting semantic intent and mapping it onto the knowledge graph structure instead of depending on exact term matching. Secondly, explainability is supported by the architecture. The query graph can be visualised, annotated, and even interactively refined by creating an intermediate graph representation for every query. Because it makes clear which entities and relations are being taken into consideration [79], this transparency aids users in understanding how their prompts are interpreted. Third, the method allows for gradual improvement. The system can mimic a human-like dialogue by asking clarifying follow-up questions, such as whether "Hellenistic" refers to a time period or an artistic style, in order to gradually refine the search instead of returning a null result because of strict schema constraints [42]. Moreover, the system empowers prompt engineering and context enrichment. A prompt like "Find ancient settlements near rivers active during early agriculture" can be interpreted by the LLM and enriched with relevant temporal ontologies and spatial reasoning constructs, then structured into a graph-aware query. This graph is then embedded by CompGCN, which jointly encodes nodes and relations to support matching based on structure and context [77].

### 8.1. The Indispensable Role of Human Expertise in AI-Assisted Processes

Despite the growing capabilities of AI models to process, interpret, and retrieve information from both structured and unstructured data, human expertise remains a key part of any system that engages with rich cultural knowledge [56,57]. In AI-assisted querying over cultural heritage knowledge graphs, human input is not just an addition; it is essential for the reliability, interpretability, and contextual relevance of the whole system [56]. AI models, including prompt-based large language models (LLMs) and graph neural networks like CompGCN, are fundamentally statistical and generalizing systems. Their ability to infer structure, relevance, and semantic similarity is limited by the quality and coverage of their training data and the ontological framing in the knowledge graph. Cultural heritage data is inherently varied, deeply contextual, and often only partially curated. This field includes ambiguities, historical gaps, and interpretive uncertainties that cannot fully be resolved by algorithms alone [56]. Human expertise is vital for curating representative datasets, designing prompt templates, aligning graph structures with domain ontologies like CIDOC CRM, and interpreting model outputs with historical context in mind [56,57]. Moreover, domain experts are crucial for assessing and improving the outputs of AI models. For example, interpreting a user query like "religious sites with serpent imagery" may involve understanding cultural motifs, symbolic meanings, and changes in iconography over time that are not clearly included in the ontology. In these situations, an expert is needed to confirm the system's graph-based matches, identify semantically inappropriate connections, and help refine the query or results [56]. This ongoing feedback not only enhances the user experience but also supports the long-term improvement and strength of the AI model, especially when the feedback is included in training processes [57]. Additionally, explainability, which is vital for building trust in AI systems, relies on human interpretation. While neural architectures can generate relevance scores or attention weights, only experts can turn those signals into valid domain interpretations, such as explaining why a particular Hellenistic inscription appeared in response to a vaguely formulated mythological query [80]. Their knowledge helps bridge the gap between the model's behavior and human cultural understanding. Ultimately, the relationship between human expertise and machine intelligence creates a hybrid form of intelligence that is particularly effective in cultural heritage contexts. In this scenario, machines can enhance human reach and recall across extensive data collections, while humans provide interpretation, critique, and contextual relevance. This partnership ensures that AI tools do not simply

replicate patterns but also contribute meaningfully to the advancement of knowledge, honoring both academic rigor and cultural sensitivity [81,82].

## 8.2. Ethical Implications and the Pursuit of Transparent AI in Cultural Heritage

As artificial intelligence systems play a larger role in accessing and interpreting cultural heritage data, important ethical issues need to be addressed. We must ensure these technologies uphold values like transparency, accountability, and cultural sensitivity. Unlike traditional datasets, cultural heritage knowledge bases include narratives, identities, and power dynamics that reflect human histories and ways of knowing. Therefore, using AI in this area poses risks that go beyond technical bias, leading to issues such as historiographic distortion, cultural erasure, and unfair knowledge representation [83]. One major ethical challenge comes from the lack of clarity in current AI models, especially large language models and graph-based neural systems like CompGCN. While these systems can generate or retrieve rich context responses, they operate like "black boxes." This makes it hard for users and curators to trace where a specific output comes from or to understand how a query match is made [84,85]. In cultural heritage applications, this lack of transparency is particularly concerning. Users may unknowingly receive results influenced by hidden biases in the data, such as Eurocentric narratives, underrepresentation of indigenous knowledge systems, or overly simplistic event interpretations outlined in systems like CIDOC CRM [83,86]. To reduce these risks, AI interfaces in this field should focus on transparency. This means developing interpretable structures, like graph-based representations used in query transformations, that users can visualise and annotate. Keeping a clear record of how data transforms from natural language input to query graph, and then to retrieved subgraphs, allows scholars and curators to review the model's behavior and ensure accuracy [84]. Additionally, creating feedback systems that let users question or adjust AI outputs is crucial for maintaining human oversight and encouraging discussion [84]. Another key ethical issue is supporting diverse perspectives. Cultural heritage is not just a fixed set of facts; it is a dynamic field full of interpretation, often featuring contested meanings and various community viewpoints. AI models should be taught and assessed with awareness of these complexities. This involves using inclusive training materials, enabling multilingual and multicultural querying options, and making sure minority narratives are not oversimplified or ignored by dominant patterns [87,88]. Transparency in how models are developed, including clear documentation of data sources, design choices, and evaluation methods, helps prevent accidental cultural oversimplification [83,89]. Finally, using AI in this area must align with legal and ethical standards concerning data ownership and management. Cultural artifacts, especially those obtained through colonial means or held in sacred regard, require careful governance when they are digitized and processed by algorithms. Transparent and explainable AI should not only reveal its inner workings but also honor the rights and values of the communities whose heritage it engages with [83,90].

## 9. Conclusions and Future Work

The experiments described in this paper offer promising results, particularly in their capacity to simplify human tasks and provide tools that accelerate the extraction, organisation, and use of cultural knowledge. What emerges is not simply a more efficient pathway from text to data, but a recursive system in which knowledge begets cognition, and cognition in turn refines the very tools through which knowledge is acquired. This methodological loop, anchored in ontological rigour and empowered by machine learning, reconfigures both the means and the ends of heritage data processing, transforming interpretation into interaction.

Yet, despite these encouraging outcomes, they represent only an early stages of a much longer journey towards the realisation of a fully-developed Cognitive Digital Twin. In future developments, we plan to refine and extend the AI-assisted semantic pipeline presented in this work by enhancing the accuracy of entity disambiguation, incorporating multimodal inputs such as images and tabular data, and introducing feedback loops that allow curators and domain experts to validate and iteratively improve the extracted knowledge while also enhancing the models being used.



Some promising zero-shot classification testing has already been conducted on images, using open-source Vision Language Models like OPENCLIP to automate the extraction of analysis types and semantic meaning to fuel Digital Twins and Knowledge Bases, and we plan to further investigate this line of research and fine-tune the AI models to boost recognition of domain-specific figure types.

We also aim to streamline the integration between natural language queries and graph-based responses, enabling more fluid and transparent interactions with the underlying knowledge base. These enhancements will not only increase the robustness of the extraction process but will further consolidate the cognitive capacities of Heritage Digital Twin, aligning their growth with the evolving needs of cultural heritage research. This entails not only refining the underlying semantic infrastructures, but also designing more intuitive, transparent, and context-aware interfaces that allow scholars and practitioners to explore, interrogate, and contribute to the knowledge base with greater ease. By empowering both the machine and the human interlocutor, we envision a more dialogic relationship with cultural data, one in which digital twins are no longer static repositories, but dynamic environments for discovery, reasoning, and collaborative interpretation.

Particular attention will also be devoted to the applications of the AI–ontology tandem in the development of decisional components of Heritage Digital Twins, enabling them to react meaningfully to sensor-derived inputs and to activate appropriate, context-aware responses. Beyond decision-making and reactive behaviour, further functionalities will be explored, such as predictive modelling, simulation of possible conservation scenarios, and projection of future states of cultural assets under varying environmental or human-induced conditions. These developments will draw on the inferential and connective power of semantic graphs, which provide the epistemic infrastructure necessary to support such complex reasoning processes. This approach echoes the insight of Heraclitus, who warned that "much learning does not teach understanding"<sup>7</sup>. It is a powerful reminder, now as then, that true knowledge arises not from the mere accumulation of data, but from its structuring, contextualisation, and meaningful integration. Our work constantly strives to answer this call: to move beyond erudition towards insight; beyond cataloguing towards comprehension.

Naturally, all future enhancements to the pipeline and to the cognitive architecture of the Heritage Digital Twin will be guided by rigorous ethical considerations and by a persistent commitment to human oversight. As artificial intelligence systems become more sophisticated in navigating and reconstructing cultural knowledge, their design must remain transparent, verifiable, and accountable. As we narrated at the beginning of this paper, the gods fashioned a simulacrum, identical to the real Helen in form, voice, and grace, to deceive mortals and at the same time to preserve and keep her safe from the violence of men. Digital twins too are conceived to preserve and protect, but certainly not to deceive, since by definition their *raison d'être* lies not in illusion, but in historical and epistemic fidelity. Yet in pursuing verisimilitude and cognitive depth, we risk being misled by the very intelligences we have created. Artificial systems, however powerful, often yield responses that are incomplete, imprecise, or altogether fabricated, i.e., hallucinations mistaken for knowledge. The risk, then, is that we ourselves are deceived by the AI meant to assist us, subtly shifting from curators of memory to unwitting custodians of fictions. The future of Heritage Digital Twins thus resides at the intersection of culture, computation and semantics [91], transforming them into dynamic epistemological machines capable of simulating worlds, diagnosing faults, predicting futures, and structuring meaning. From this vision emerges a new notion of the Cognitive Digital Twin: a system capable of transcending passive simulation to engage in active reasoning, hypothesis generation, and adaptive control [92].

**Author Contributions:** All authors have equally contributed to the present paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper received no external funding.

**Institutional Review Board Statement:** Not applicable.

---

<sup>7</sup> Fragment B 40.

**Data Availability Statement:** Data is unavailable due to privacy restrictions.

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

## References

1. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology* **2020**, *29*, 36–52. Publisher: Elsevier BV, <https://doi.org/10.1016/j.cirpj.2020.02.002>.
2. Semeraro, C.; Lezoche, M.; Panetto, H.; Dassisti, M. Digital twin paradigm: A systematic literature review. *Computers in Industry* **2021**, *130*, 103469. Publisher: Elsevier BV, <https://doi.org/10.1016/j.compind.2021.103469>.
3. Schrotter, G.; Hürzeler, C. The Digital Twin of the City of Zurich for Urban Planning. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* **2020**, *88*, 99–112. Publisher: Springer Science and Business Media LLC, <https://doi.org/10.1007/s41064-020-00092-2>.
4. Caprari, G.; Castelli, G.; Montuori, M.; Camardelli, M.; Malvezzi, R. Digital Twin for Urban Planning in the Green Deal Era: A State of the Art and Future Perspectives. *Sustainability* **2022**, *14*, 6263. Publisher: MDPI AG, <https://doi.org/10.3390/su14106263>.
5. Grieves, M.; Vickers, J., Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*; Kahlen, F.J.; Flumerfelt, S.; Alves, A., Eds.; Springer International Publishing: Cham, 2017; pp. 85–113. [https://doi.org/10.1007/978-3-319-38756-7\\_4](https://doi.org/10.1007/978-3-319-38756-7_4).
6. Li, L.; Aslam, S.; Wileman, A.; Perinpanayagam, S. Digital Twin in Aerospace Industry: A Gentle Introduction. *IEEE Access* **2022**, *10*, 9543–9562. <https://doi.org/10.1109/ACCESS.2021.3136458>.
7. Alfaro-Viquez, D.; Zamora-Hernandez, M.; Fernandez-Vega, M.; Garcia-Rodriguez, J.; Azorin-Lopez, J. A Comprehensive Review of AI-Based Digital Twin Applications in Manufacturing: Integration Across Operator, Product, and Process Dimensions. *Electronics* **2025**, *14*, 646. <https://doi.org/10.3390/electronics14040646>.
8. Lovell, L.J.; Davies, R.J.; Hunt, D.V.L. The Application of Historic Building Information Modelling (HBIM) to Cultural Heritage: A Review. *Heritage* **2023**, *6*, 6691–6717. <https://doi.org/10.3390/heritage6100350>.
9. Raes, L.; Ruston McAleer, S.; Croket, I.; Kogut, P.; Brynskov, M.; Lefever, S., Eds. *Decide Better: Open and Interoperable Local Digital Twins*; Springer Nature Switzerland: Cham, 2025. <https://doi.org/10.1007/978-3-031-81451-8>.
10. Papazoglou, M.P.; Krämer, B.J.; Raheem, M.; Elgammal, A. Medical Digital Twins for Personalized Chronic Care, 2025. <https://doi.org/10.21203/rs.3.rs-5826330/v1>.
11. European Commission, Destination Earth. Available online: <https://destination-earth.eu>. (Accessed on 15 July 2025).
12. Nikolinakos, N.T. *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies - the AI Act*, 1st ed ed.; Number v.53 in Law, Governance and Technology Series, Springer International Publishing AG: Cham, 2023.
13. Grieves, M. Digital Twin: Manufacturing Excellence through Virtual Factory Replication **2014**.
14. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* **2020**, *8*, 108952–108971. Publisher: Institute of Electrical and Electronics Engineers (IEEE), <https://doi.org/10.1109/access.2020.2998358>.
15. Mayer, A.; Greif, L.; Häußermann, T.M.; Otto, S.; Kastner, K.; El Bobbou, S.; Chardonnet, J.R.; Reichwald, J.; Fleischer, J.; Ovtcharova, J. Digital Twins, Extended Reality, and Artificial Intelligence in Manufacturing Reconfiguration: A Systematic Literature Review. *Sustainability* **2025**, *17*, 2318. Publisher: MDPI AG, <https://doi.org/10.3390/su17052318>.
16. Roopa, M.S.; Venugopal, K.R. Digital Twins for Cyber-Physical Healthcare Systems: Architecture, Requirements, Systematic Analysis, and Future Prospects. *IEEE Access* **2025**, *13*, 44963–44996. Publisher: Institute of Electrical and Electronics Engineers (IEEE), <https://doi.org/10.1109/access.2025.3547991>.
17. Charalabidis, Y.; Kontos, G.; Zitianellis, D., Local Digital Twins for Cities and Regions: The Way Forward. In *Decide Better: Open and Interoperable Local Digital Twins*; Raes, L.; Ruston McAleer, S.; Croket, I.; Kogut, P.; Brynskov, M.; Lefever, S., Eds.; Springer Nature Switzerland: Cham, 2025; pp. 233–259. [https://doi.org/10.1007/978-3-031-81451-8\\_9](https://doi.org/10.1007/978-3-031-81451-8_9).

18. Tampère, C.; Ortmann, P.; Jedlička, K.; Lohman, W.; Janssen, S., Future Ready Local Digital Twins and the Use of Predictive Simulations: The Case of Traffic and Traffic Impact Modelling. In *Decide Better: Open and Interoperable Local Digital Twins*; Raes, L.; Ruston McAleer, S.; Croket, I.; Kogut, P.; Brynskov, M.; Lefever, S., Eds.; Springer Nature Switzerland: Cham, 2025; pp. 203–230. [https://doi.org/10.1007/978-3-031-81451-8\\_8](https://doi.org/10.1007/978-3-031-81451-8_8).
19. Arsalan, H.; Heesom, D.; Moore, N. From Heritage Building Information Modelling Towards an ‘Echo-Based’ Heritage Digital Twin. *Heritage* **2025**, *8*, 33. Publisher: MDPI AG, <https://doi.org/10.3390/heritage8010033>.
20. Niccolucci, F.; Felicetti, A. Digital Twin Sensors in Cultural Heritage Ontology Applications. *Sensors* **2024**, *24*, 3978. Publisher: MDPI AG, <https://doi.org/10.3390/s24123978>.
21. Singh, M.; Fuenmayor, E.; Hinchy, E.; Qiao, Y.; Murray, N.; Devine, D. Digital Twin: Origin to Future. *Applied System Innovation* **2021**, *4*, 36. Publisher: MDPI AG, <https://doi.org/10.3390/asi4020036>.
22. Leipnitz, M.T.; Petry, R.H.; Rodrigues, F.H.; De Souza Silva, H.R.; Correia, J.B.; Becker, K.; Wickboldt, J.A.; Carbonera, J.L.; Netto, J.C.; Abel, M. Architecting Digital Twins: From Concept to Reality. *Procedia Computer Science* **2025**, *256*, 530–537. Publisher: Elsevier BV, <https://doi.org/10.1016/j.procs.2025.02.149>.
23. Rathore, M.M.; Shah, S.A.; Shukla, D.; Bentafat, E.; Bakiras, S. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities. *IEEE Access* **2021**, *9*, 32030–32052. Publisher: Institute of Electrical and Electronics Engineers (IEEE), <https://doi.org/10.1109/access.2021.3060863>.
24. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction* **2020**, *114*, 103179. Publisher: Elsevier BV, <https://doi.org/10.1016/j.autcon.2020.103179>.
25. Karabulut, E.; Pileggi, S.F.; Groth, P.; Degeler, V. Ontologies in digital twins: A systematic literature review. *Future Generation Computer Systems* **2024**, *153*, 442–456. Publisher: Elsevier BV, <https://doi.org/10.1016/j.future.2023.12.013>.
26. XMLPO: An Ontology for Explainable Machine Learning Pipeline. In *Frontiers in Artificial Intelligence and Applications*; IOS Press, 2024. ISSN: 0922-6389, 1879-8314, <https://doi.org/10.3233/faia241306>.
27. Semantic Data-Modeling and Long-Term Interpretability of Cultural Heritage Data—Three Case Studies. In *Digital Cultural Heritage*; Springer International Publishing: Cham, 2020; pp. 239–253. [https://doi.org/10.1007/978-3-030-15200-0\\_16](https://doi.org/10.1007/978-3-030-15200-0_16).
28. Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. Publisher: MDPI AG, <https://doi.org/10.3390/data7080105>.
29. Niccolucci, F.; Markhoff, B.; Theodoridou, M.; Felicetti, A.; Hermon, S. The Heritage Digital Twin: a bicycle made for two. The integration of digital methodologies into cultural heritage research. *Open Research Europe* **2023**, *3*, 64. Publisher: F1000 Research Ltd, <https://doi.org/10.12688/openreseurope.15496.1>.
30. Felicetti, A.; Meghini, C.; Richards, J.; Theodoridou, M. The AO-Cat Ontology. Technical report, Zenodo, 2025. Version Number: 1.2, <https://doi.org/10.5281/ZENODO.7818374>.
31. Alammari, J.; Grootendorst, M. *Hands-on large language models: language understanding and generation*, first edition ed.; O'Reilly Media, Inc: Sebastopol, CA, 2024.
32. Wu, X.; Yu, Y.; Li, Y. A damage detection network for ancient murals via multi-scale boundary and region feature fusion. *npj Heritage Science* **2025**, *13*. Publisher: Springer Science and Business Media LLC, <https://doi.org/10.1038/s40494-025-01546-9>.
33. Santini, C.; Posthumus, E.; Tan, M.A.; Bruns, O.; Tietz, T.; Sack, H. Multimodal Search on Iconclass using Vision-Language Pre-Trained Models, 2023. Version Number: 1, <https://doi.org/10.48550/ARXIV.2306.16529>.
34. Detecting and Deciphering Damaged Medieval Armenian Inscriptions Using YOLO and Vision Transformers. In *Lecture Notes in Computer Science*; Springer Nature Switzerland: Cham, 2024; pp. 22–36. ISSN: 0302-9743, 1611-3349, [https://doi.org/10.1007/978-3-031-70642-4\\_2](https://doi.org/10.1007/978-3-031-70642-4_2).
35. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision, 2021. Version Number: 1, <https://doi.org/10.48550/ARXIV.2103.00020>.
36. Cherti, M.; Beaumont, R.; Wightman, R.; Wortsman, M.; Ilharco, G.; Gordon, C.; Schuhmann, C.; Schmidt, L.; Jitsev, J. Reproducible Scaling Laws for Contrastive Language-Image Learning. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023; pp. 2818–2829. <https://doi.org/10.1109/cvpr52729.2023.00276>.
37. CIDOC CRM. Available online: <https://cidoc-crm.org>. (Accessed on 10 July 2025).

38. CRMsci - Scientific Observation Model. Available online: <https://cidoc-crm.org/crmsci>. (Accessed on 11 July 2025).
39. CRMdig - Model for Provenance Metadata. Available online: <https://cidoc-crm.org/crmdig>. (Accessed on 14 July 2025).
40. Felicetti, A.; Niccolucci, F. Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data. *Data* **2024**, *10*, 1. Publisher: MDPI AG, <https://doi.org/10.3390/data10010001>.
41. Fensel, D.; Şimşek, U.; Angele, K.; Huaman, E.; Kärle, E.; Panasiuk, O.; Toma, I.; Umbrich, J.; Wahler, A., How to Build a Knowledge Graph. In *Knowledge Graphs: Methodology, Tools and Selected Use Cases*; Springer International Publishing: Cham, 2020; pp. 11–68. [https://doi.org/10.1007/978-3-030-37439-6\\_2](https://doi.org/10.1007/978-3-030-37439-6_2).
42. Maree, M. Quantifying Relational Exploration in Cultural Heritage Knowledge Graphs with LLMs: A Neuro-Symbolic Approach for Enhanced Knowledge Discovery. *Data* **2025**, *10*, 52.
43. Bruns, O. The Persistence of Temporality: Representation of Time in Cultural Heritage Knowledge Graphs. *ISWC 2023* **2023**.
44. Faralli, S.; Lenzi, A.; Velardi, P. A Large Interlinked Knowledge Graph of the Italian Cultural Heritage. In *Proceedings of the LREC, 2022*, pp. 6280–6289.
45. Carriero, V.A.; Gangemi, A.; Mancinelli, M.L.; et al. ArCo: the Italian Cultural Heritage Knowledge Graph. *arXiv preprint arXiv:1905.02840* **2019**.
46. Barzaghi, S.; Moretti, A.; Heibi, I.; Peroni, S. CHAD-KG: A Knowledge Graph for Representing Cultural Heritage Objects and Digitisation Paradata. *arXiv preprint arXiv:2505.13276* **2025**.
47. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018. Version Number: 2, <https://doi.org/10.48550/ARXIV.1810.04805>.
48. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners, 2020. Version Number: 4, <https://doi.org/10.48550/ARXIV.2005.14165>.
49. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and Efficient Foundation Language Models, 2023. Version Number: 1, <https://doi.org/10.48550/ARXIV.2302.13971>.
50. Gekhman, Z.; Yona, G.; Aharoni, R.; Eyal, M.; Feder, A.; Reichart, R.; Herzig, J. Does Fine-Tuning LLMs on New Knowledge Encourage Hallucinations?, 2024. Version Number: 3, <https://doi.org/10.48550/ARXIV.2405.05904>.
51. Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; Wu, X. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering* **2024**, *36*, 3580–3599. Publisher: Institute of Electrical and Electronics Engineers (IEEE), <https://doi.org/10.1109/tkde.2024.3352100>.
52. Agrawal, A.; Suzgun, M.; Mackey, L.; Kalai, A.T. Do Language Models Know When They're Hallucinating References?, 2023. Version Number: 3, <https://doi.org/10.48550/ARXIV.2305.18248>.
53. Sui, Y.; He, Y.; Ding, Z.; Hooi, B. Can Knowledge Graphs Make Large Language Models More Trustworthy? An Empirical Study Over Open-ended Question Answering, 2025. arXiv:2410.08085 [cs], <https://doi.org/10.48550/arXiv.2410.08085>.
54. Sun, J.; Du, Z.; Chen, Y. Knowledge Graph Tuning: Real-time Large Language Model Personalization based on Human Feedback, 2024. arXiv:2405.19686 [cs], <https://doi.org/10.48550/arXiv.2405.19686>.
55. Zhu, Y.; Wang, X.; Chen, J.; Qiao, S.; Ou, Y.; Yao, Y.; Deng, S.; Chen, H.; Zhang, N. LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities, 2024. arXiv:2305.13168 [cs], <https://doi.org/10.48550/arXiv.2305.13168>.
56. Zhang, B.; Meronyo Peñuela, A.; Simperl, E. Towards Explainable Automatic Knowledge Graph Construction with Human-in-the-loop. *Semantic Web Journal* **2023**.
57. Manzoor, E.; Tong, J.; Vijayaraghavan, S.; Li, R. Expanding Knowledge Graphs with Humans in the Loop. *arXiv preprint arXiv:2212.05189* **2023**.
58. Meroño-Peñuela, A.; Sack, H. Move Cultural Heritage Knowledge Graphs in Everyone's Pocket. *Semantic Web Journal* **2020**.
59. Yuan, H.; Li, Y.; Wang, B.; et al. Knowledge Graph-Based Intelligent Question Answering System for Ancient Chinese Costume Heritage. *npj Heritage Science* **2025**, *13*, 198.
60. Huang, Y.; Yu, S.; Chu, J.; Fan, H.; Du, B. Using Knowledge Graphs and Deep Learning Algorithms to Enhance Digital Cultural Heritage Management. *Heritage Science* **2023**, *11*, 204.
61. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems* **2021**.



62. Majumder, e.a. Injecting Knowledge Graphs into Large Language Models. *arXiv preprint arXiv:2505.07554* **2025**.
63. Mahmud, F.; Zhang, W. SparqLLM: Natural Language Querying of Knowledge Graphs via Retrieval-Augmented Generation. In Proceedings of the arXiv preprint arXiv:2502.01298, 2025.
64. Wang, J.; Cui, Z.; Wang, B.; Pan, S.; Gao, J.; Yin, B.; Gao, W. IME: Integrating Multi-curvature Shared and Specific Embedding for Temporal Knowledge Graph Completion. *ArXiv* **2024**.
65. Xu, D.; Cheng, W.; Luo, D.; Chen, H.; Zhang, X. InfoGCL: Information-Aware Graph Contrastive Learning. *ArXiv* **2021**.
66. Liang, K.; Liu, Y.; Zhou, S.; Tu, W.; Wen, Y.; Yang, X.; Dong, X.; Liu, X. Knowledge Graph Contrastive Learning Based on Relation-Symmetrical Structure. *ArXiv* **2022**.
67. Turki, T.; et al. Semantic-Based Image Retrieval: Enhancing Visual Search with Knowledge Graph Structures. *Saudi Journal of Engineering and Technology* **2025**.
68. Castellano, G.; Sansaro, G.; Vessio, G. Integrating Contextual Knowledge to Visual Features for Fine Art Classification. *arXiv preprint arXiv:2105.15028* **2021**.
69. Füll, A.; Nissen, V. Design of Explanation User Interfaces for Interactive Machine Learning. *Requirements Engineering* **2025**.
70. Barzaghi, S.; Colitti, S.; Moretti, A.; Renda, G. From Metadata to Storytelling: A Framework For 3D Cultural Heritage Visualization on RDF Data. *arXiv preprint arXiv:2505.14328* **2025**.
71. Romphf, J.; Neuman-Donihue, E.; Heyworth, G.; Zhu, Y. Resurrect3D: An Open and Customizable Platform for Visualizing and Analyzing Cultural Heritage Artifacts. *arXiv preprint arXiv:2106.09509* **2021**.
72. ARTEMIS Applying Reactive Twins to Enhance Monument Information Systems. Available online: <https://www.artemis-twin.eu>. (Accessed on 12 July 2025).
73. Castelli, L.; Felicetti, A.; Proietti, F. Heritage Science and Cultural Heritage: standards and tools for establishing cross-domain data interoperability. *International Journal on Digital Libraries* **2021**, 22, 279–287. Publisher: Springer Science and Business Media LLC, <https://doi.org/10.1007/s00799-019-00275-2>.
74. The AO-Cat Ontology. Available online: <https://zenodo.org/records/7818375>. (Accessed on 14 July 2025).
75. ARIADNE Research Infrastructure. Available online: <https://www.ariadne-research-infrastructure.eu>. (Accessed on 12 July 2025).
76. Mazzinghi, A.; Ruberto, C.; Giuntini, L.; Mandò, P.A.; Taccetti, F.; Castelli, L. Mapping with Macro X-ray Fluorescence Scanning of Raffaello's Portrait of Leo X. *Heritage* **2022**, 5, 3993–4005. Publisher: MDPI AG, <https://doi.org/10.3390/heritage5040205>.
77. Vashishth, S.; Sanyal, S.; Nitin, V.; Talukdar, P. Composition-based Multi-Relational Graph Convolutional Networks. *Proceedings of ICLR* **2020**. Published as conference paper in 2019.
78. Emonet, V.; Bolleman, J.; Duvaud, S.; de Farias, T.; Sima, A.C. LLM-based SPARQL Query Generation from Natural Language over Federated Knowledge Graphs. *arXiv preprint arXiv:2410.06062* **2025**.
79. Mahmud, F.; Zhang, W. SparqLLM: Augmented Knowledge Graph Querying leveraging LLMs. *arXiv preprint arXiv:2502.01298* **2025**.
80. Celino, I. Who is this Explanation for? Human Intelligence and Knowledge Graphs for eXplainable AI. *arXiv preprint arXiv:2005.13275* **2020**.
81. Kamar, E. Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence. In Proceedings of the Proceedings of AAAI Spring Symposium: Technical Report, 2016.
82. Te'eni, D.; Yahav, I.; Zagalsky, A.; Schwartz, D.G.; Silverman, G. Reciprocal Human–Machine Learning: A Theory and an Instantiation for the Case of Message Classification. *Management Science* **2023**.
83. Foka, A.; Griffin, G.; Ortiz Pablo, D.; Rajkowska, P.; Badri, S. Tracing the bias loop: AI, cultural heritage and bias-mitigating in practice. *AI & Society* **2025**.
84. Tóth, G.M.; Albrecht, R.; Pruski, C. Explainable AI, LLM, and digitized archival cultural heritage: a case study of the Grand Ducal Archive of the Medici. *AI & Society* **2025**.
85. Peters, U.; Carman, M. Cultural Bias in Explainable AI Research: A Systematic Analysis. *Journal of Artificial Intelligence Research* **2024**.
86. Mitchell, M. AI Is Spreading Old Stereotypes to New Languages and Cultures. *Wired* **2025**.
87. Upmann, P. What Role Does Cultural Context Play in Defining Ethical Standards for AI? *AI Global* **2024**.
88. Prabhakaran, V.; Qadri, R.; Hutchinson, B. Cultural Incongruencies in Artificial Intelligence. *ArXiv preprint arXiv:2211.13069* **2022**.
89. Inuwa-Dutse, I. FATE in AI: Towards Algorithmic Inclusivity and Accessibility. *ArXiv preprint arXiv:2301.01590* **2023**.



90. Kukutai, T.; Taylor, J. *Indigenous Data Sovereignty: Towards an Agenda*; ANU Press, 2020.
91. Tooth, J.M.; Tuptuk, N.; Watson, J.D.M. A Systematic Survey of the Gemini Principles for Digital Twin Ontologies, 2024. Version Number: 1, <https://doi.org/10.48550/ARXIV.2404.10754>.
92. D'Amico, R.D.; Sarkar, A.; Karray, M.H.; Addepalli, S.; Erkoyuncu, J.A. Knowledge transfer in Digital Twins: The methodology to develop Cognitive Digital Twins. *CIRP Journal of Manufacturing Science and Technology* **2024**, *52*, 366–385. Publisher: Elsevier BV, <https://doi.org/10.1016/j.cirpj.2024.06.007>.

**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.