

BHGPT: Multiple AI Models for Built Heritage Knowledge Retrieval

Cassia De Lian Cui^{*}, Wei Yan², Antonio Fioravanti³, and Edoardo Currà⁴.

1^{*} – Department of Civil Building and Environmental Engineering, Sapienza University of Rome,
cassiadelian.cui@uniroma1.it

2 – Department of Architecture, Texas A&M University, wyan@tamu.edu

3 – Department of Civil Building and Environmental Engineering, Sapienza University of Rome,
antonio.fioravanti@uniroma1.it

4 – Department of Civil Building and Environmental Engineering, Sapienza University of Rome,
edoardo.curra@uniroma1.it

Abstract. BHGPT (Built Heritage Generative Pretrained Transformers) is an AI-driven framework designed to improve knowledge retrieval and management in the built heritage context. While traditional methods struggle with heterogeneous sources, inconsistencies, and evolving historical data, this study explores customized GPT, Fine-Tuned BHGPT, and an AI Assistant with Retrieval-Augmented Generation (RAG) to enhance the accessibility and interpretability of heritage information. This study adopts a domain knowledge first and operational approach, in which discipline-specific concepts and terminology guide the organization of information and the structuring of model outputs. The approach begins with a customized GPT model and prompt engineering to test the model's response behavior. After that, fine-tuning the model on specific datasets enhances domain-specific accuracy. Finally, the AI Assistant with RAG integrates structured HBIM data and unstructured archival sources, enabling dynamic querying and cross-referencing of historical and architectural knowledge. The framework is tested on the Sanctuary of Hercules and the Former Segrè Papermill in Tivoli, evaluating its performance across historical evolution, architectural aspects, and interdisciplinary knowledge. Results indicate that Fine-Tuned BHGPT significantly improves site-specific knowledge extraction, while AI Assistant with RAG provides the most flexible and adaptive responses by linking multiple data sources. However, its accuracy is dependent on data availability and retrieval mechanisms. Overall, the contribution clarifies the utility of the BHGPT framework for heritage professionals, enhancing accessibility through natural language querying, producing structured and interoperable outputs, and supporting transparent, source-grounded interpretation.

Keywords: BHGPT, Fine-tuning, AI Assistant, Built Heritage, Knowledge Management

1. Introduction

Knowledge representation and management are fundamental activities for the recovery and valorization processes in the built heritage field. Every intervention relies on collected, analyzed, and interpreted information and heterogeneous data, often derived from different sources, both historical and archival research and direct in situ investigation. Traditional digital heritage methodologies struggle to handle knowledge characterized by changes, inconsistencies, and uncertainties that define the complexity of the information's logical structuring [1].

Recent advancements in information systems for built heritage allow the combination of distinct disciplines and the linking of distant classes of information [2]. While structured query languages such as SQL (Structured Query Language) for relational databases [3] and SPARQL (Protocol and RDF Query Language) for semantic web technologies, such as ontologies [4], offer precise querying capabilities, the usability requires prior knowledge and understanding of specific data structures and logic.

The integration of AI-driven approaches in the AEC sector, specifically LLMs (Large Language Models), marks a shift in the research horizon towards a new paradigm. These models introduce a more intuitive and accessible way for experts and non-expert users to retrieve and reason heritage data [5]. Among LLMs, GPTs (Generative Pre-trained Transformers) have impressive learning abilities using textual prompts [6], allowing and enhancing

49 the understanding of intricate architectural features of complex heritage sites without requiring a technical
50 background. These GPT models can extract information by combining text-based input with integrated task
51 instructions and chosen contextual examples. Automating descriptions and responses based on domain
52 knowledge could aid the documentation process in heritage valorization activities, thanks to its efficient
53 development and experimentation.

54 This study proposes a structured framework for GPT models (BHGPT) within the heritage domain, focusing on
55 multisource knowledge retrieval to support recovery and valorization processes. The research outlines three AI-
56 driven methodologies, Customized GPT, Fine-Tuned BHGPT, and an AI Assistant with Retrieval-Augmented
57 Generation (RAG), highlighting their capabilities, limitations, and advancements. BHGPT framework can
58 support technicians and researchers by enabling natural-language querying of heterogeneous sources;
59 normalizing vocabulary to domain standards; and returning structured, interoperable outputs that can be directly
60 reused in documentation, analysis, and consistency checks. The system can also assist in cross-referencing
61 sources and making reasoning steps and provenance explicit, improving consistency and transparency in
62 interpretation workflows. The approach is designed to be domain-knowledge driven rather than focused on the
63 tool, emphasizing how specialized concepts, terminology, and reasoning patterns in built-heritage research are
64 integrated within AI-assisted workflows as reasoning instruments aligned with disciplinary logics and
65 interpretation. A significant difference between this research and other AI research in the field of built heritage
66 is the new training of AI using the methods of customized, fine-tuned, and RAG based on extensive domain-
67 knowledge data that we provided to the training processes, instead of a straightforward application of existing
68 AI tools to query the generic knowledge base without domain-knowledge. This research process leads to new
69 findings that can be applied to developing new AI models for enhancing built-heritage knowledge retrieval.
70 Eventually, the approach is applied to a unique case study of the Sanctuary of Hercules and the Former Segrè
71 Papermill in Tivoli (Rome) to evaluate the performance of the models.

72 The contributions of the proposed approach include bridging the gap in the application of generative AI models
73 in the heritage field and, at the same time, facilitating accessibility and interpretation in the knowledge
74 management process, demonstrating the potential of AI framework technologies in heritage research and
75 conservation.

76 **2. Large Language Models in the Heritage field**

77 Methods for knowledge retrieval and management represent an important shift within the built heritage domain.
78 Whilst several digital workflows have contributed to improving data accessibility and interpretation, there are
79 still limitations to the retrieval and management of complex knowledge performed by these systems [7].
80 This section reviews both traditional methodologies and the capabilities of LLMs in terms of access to
81 knowledge and identifies potential future opportunities in the unique field of cultural heritage studies.

82 *2.1 Opportunities of Natural Language Processing with LLMs*

83 Natural language processing (NLP) has evolved in the last few years thanks to the development of LLMs. These
84 models, trained on vast amounts of text data, can generate human-like responses with remarkable fluency and
85 coherence [8-9]. Two of the latest pre-trained models in the OpenAI GPT series are GPT-4.5 (February 2025)
86 and GPT-4o (May 2024); their development has enabled previously unattainable levels of fluency and
87 coherence in a machine-generated text. These models have been effectively used in various applications,
88 including chatbots, content generation, machine translation, interpreting, and NL generation, with minimal
89 additional training [10].

90 Some initial research and applications of generative AI technologies have been performed in the AEC
91 (Architecture, Engineering, and Construction) sector. Zheng et al. focused on integrating the GPT model with
92 BIM (Building Information Modeling) [11]. Their primary purpose was to retrieve information from the
93 building information model using natural language and BSON (Binary JSON) to extract building objects and
94 properties from the BIM, and to clean and store the information in MongoDB. You et al. focused on automated
95 sequence planning of construction tasks [12], similar to Prieto et al.'s work [13]. Additionally, ChatGPT-3.5 has
96 been utilized in hazard recognition and construction safety education, demonstrating the versatility of LLMs in

97 AEC applications [14].
98 In the heritage context, the work performed by Firmani et al. [15] for the project 'In Codice Ratio' regards the
99 automated text recognition and semantic enrichment of historical manuscripts. The application of NLP to the
100 extraction of structured knowledge from unstructured data has also been investigated in recent developments.
101 In order to improve accessibility and research workflows in heritage studies, Spina [16] addresses the integration
102 of AI tools, such as Transkribus and GPT-based systems, to improve transcription accuracy and semantic
103 interpretation of archival materials.
104 Loffredo and De Santo employ ontologies and LLMs to navigate and interact with cultural heritage data [17].
105 Other examples involve constructing linguistic corpora for low-resource languages to enhance preservation, as
106 well as defining fine-tuned models that assist professionals in analyzing historical texts and artifacts [18].
107 Moreover, Zhao et al. focused on the development and evaluation of LLM for the Intangible Cultural Heritage
108 (ICH) Domain, to provide a digital approach to transmitting knowledge [19]. CATS (Cultural-heritage
109 Advanced Translation Systems) is a research project focused on creating images for the classification of cultural
110 heritage artifacts [20], while others focused on implementing chatbots to enrich engagement with existing
111 Cultural Heritage websites [21].
112 Despite these advancements, studies by Cui et al. [22] highlighted the challenges of adapting LLMs to domain-
113 specific terminologies and multilingual datasets. Combining LLMs with existing digital heritage frameworks
114 presents an opportunity to enhance accessibility, automate knowledge extraction, and facilitate interdisciplinary
115 research activities, proposing new forms of structured data retrieval and interpretation [23].

116 2.2 *Prompt engineering, fine-tuning, and AI Assistants*

117 Prompt Engineering is a technique that takes advantage of the structure and wording of prompts, i.e., textual
118 requests provided to the model, to maximize performance without directly changing the structure of the model
119 [24].
120 This approach is used not only for pre-trained models but also with customized versions incorporating
121 specialized data. Rather than modifying the internal structure of the model, prompt engineering relies on
122 processing optimized inputs to stimulate the model to generate relevant responses. Prompt engineering involves
123 carefully and strategically designing queries presented to the AI model to direct it toward specific responses.
124 However, the results of Prompt Engineering depend on the knowledge pre-trained or built into the custom
125 model. The answers may be incomplete or generic if the starting dataset does not include detailed domain
126 information.
127 On the other hand, fine-tuning represents a more advanced technique than prompt engineering since it involves
128 retraining the AI model on specialized datasets [25-26]. This process allows the model to more effectively
129 understand technical terminologies, semantic relationships, and structured and unstructured data peculiarities.
130 This process, however, presents significant challenges, including the need for high computational resources and
131 accurate specialized datasets, which require considerable preparation and effort.
132 The integration of these techniques finds its most significant expression through advanced APIs such as
133 OpenAI's Assistant API [27]. This platform allows complex queries to be orchestrated by combining different
134 tools for querying structured and unstructured data using RAG (Retrieval Augmented Generation) [28]. This
135 AI Assistant works based on fundamental building blocks: each interaction opens a thread for a variety of tasks,
136 such as interpreting user queries; the Assistant simultaneously combines tools such as code interpreter, file
137 search, and function calling, allowing multiple access to different tools; and it can process text-based
138 information in the vector store, external databases, and structured files [29]. This configuration allows users in
139 multidisciplinary fields to access centralized knowledge without needing advanced data querying skills.
140 Compared to existing works, the novelty of this framework is the comparison of three complementary LLM
141 strategies (customized, fine-tuned, and RAG), on the same heritage case and a head-to-head evaluation across
142 20 prompts to monitor the models' behavior in terms of accuracy, flexibility, and reusability, towards the
143 integration of AI workflows in the existing heritage practices.
144

145 **3. AI-driven framework for accessing and interpreting heritage data: from customized GPT**
146 **to AI heritage assistant**

147 This work proposes and tests a methodological approach to enhance the LLMs' performance based on three
148 distinct stages, starting from static GPT customization to fine-tuned GPT to adaptive AI Assistants capable of
149 real-time retrieval from different sources.

150 Beyond testing different configurations, the framework formalizes a reproducible method for comparing and
151 evaluating how each AI strategy interprets and structures heritage knowledge in specific conditions.

152 The progression through these three phases (Figure 1) illustrates the potential expansion of the technology from
153 a manually curated GPT model to an advanced, hybrid AI Assistant, which integrates pre-trained knowledge
154 and Retrieval-Augmented Generation (RAG) for information querying and reasoning.

155 One of the objectives of this AI-driven framework is to understand the model performance based on different
156 types of inputs and techniques in order to define a fully reliable prototype that can be exploited in the built
157 heritage knowledge and interpretation processes.
158

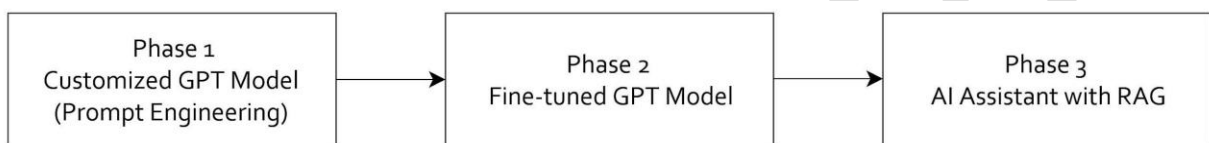


Figure 1. Overview of the AI-driven framework for built heritage knowledge retrieval. © 2025, the Authors

159 **3.1 Customized GPT model**

160 The proposed workflow starts with defining the purpose of the application and the objectives of testing the GPT
161 model. Unlike purely exploratory uses of GPT, this phase establishes a clear input-output protocol tied to
162 heritage-specific information schemas. Since the work is based on heritage artifacts, the model should be able
163 to reply with specialized knowledge related to multiple domains of interest. The architectural assets domain
164 requires the topological space relations and the classification of building components and materials, as
165 quantitative and qualitative properties of construction zones, components, and materials. At the same time, the
166 cultural heritage domain focuses on the tangible and immaterial values embodied in the studied assets and the
167 relations established with the historical, technological, social, and geographic contexts. Other knowledge
168 domains can also be involved depending on the specific case study.

169 **3.1.1 Data collection, preparation and pentagram framework**

170 The process can be subdivided into data collection, where the research leads to the analysis of different sources:

- 171 • Archaeological sources consist of the physically present remains that can be directly investigated.
- 172 • Written sources, unpublished and published.
- 173 • Visual sources, iconographic, cartographic, and photographic
- 174 • Oral sources, oral stories as testimonies of collective identification.

175 From these sources, it is possible to perform an initial classification and structuring based on the requirements
176 and the available information in different formats. Before the AI testing, the data preparation process represents
177 a crucial activity for model training performance. Since the GPT is trained on text-based knowledge, detailed
178 annotations that include contextual information, significance, and preliminary interpretation are necessary,
179 along with tagging data with labels describing the nature of the knowledge.

180 A preliminary analysis of compatibility with the model helps improve the success of testing; this involves
181 linking historical and present pictures with a textual description.

182 The creation of a custom GPT model is performed by using the configuration key. A customizable framework
183 is applied to define the instruction, where each component controls the model's output.

184 By adopting the following pentagram structure, the prompts are adequate, relevant, and appropriate:

- 185 • Context: background information necessary to contextualize the model's responses.

- 186 • Persona: depending on the user group (public or specialists), how the model's persona is tailored to fit the
- 187 audience.
- 188 • Task: the main objective of the model, such as answering specific types of queries or providing data-
- 189 driven insights.
- 190 • Output: expected format and tone of the model's responses, whether casual or formal.
- 191 • Constraints: limitations or boundaries within which the model operates, ensuring relevance and
- 192 appropriateness of responses.

193 3.1.2 Iterative testing and validation strategy

194 Once the instruction is defined, iterative testing begins and proceeds through several stages, depending on the
195 quantity and quality of the gathered and pre-structured knowledge. The process begins with the creation of a
196 list of prompts within certain knowledge domains, Prompt IDs, and expected answers. Then, the prompts are
197 sent to the model with follow-up questions to adjust the model responses.

198 Along with the established dialogue, the model is trained in multiple steps using different types of knowledge:
199 current and historical images, historical text documentation, construction drawings and architectural plans,
200 construction text documentation, and labeled historical images.

201 The aims are to evaluate the final results and responses to the defined prompts and to understand how the model
202 reacts to different types of information and which are more effective and efficient during the testing stage. For
203 this reason, prompt-response validation is performed at every step of the process according to the following
204 criteria.

- 205 • Accuracy: correctly verified according to records and context accuracy
- 206 • Relevance: no diverging into unrelated topics
- 207 • Completeness: addressed relevant key aspects and sufficient depth
- 208 • Clarity: no ambiguity and use of proper terminology, whether for a specialist or the general public
- 209 • Conciseness: balance between details and efficiency
- 210 • Interactivity and adaptability: capable of following up on questions

211 Once the process is complete, the final evaluation can ensure the practical applicability of the proposed
212 framework.

213 3.2 Fine-tuned GPT model beyond standard customization

214 While prompt engineering provides structured outputs, it remains highly dependent on pre-existing training
215 data. This phase introduces fine-tuning GPT-4 with domain-specific datasets to overcome these limitations and
216 improve response accuracy, consistency, and terminology relevance.

217 After source selection and information extraction, as proposed in the previous section, the model has to be
218 trained on domain-specific datasets in JSONL (JSON Lines), a format well-suited for this process. Each entry
219 is built upon a system prompt to set the domain context, a user prompt (query), and the expected response that
220 exemplifies the desired output (1). The fine-tuning process is not merely memorization; it helps the model
221 reason more effectively within a specific domain by learning patterns from the dataset. The provided examples
222 teach terminology, response structure, and domain knowledge, ensuring accurate and relevant answers to
223 various questions.

224

```
225 { "messages" : [  
226   { "role" : "system", "content" : "X" } ,  
227   { "role" : "user", "content" : "X" } ,  
228   { "role" : "assistant" , "content" : "X" }  
229 ] } (1)
```

225

226 The fine-tuning process uses optimization techniques such as hyperparameter calibration. Learning rate, batch
227 size, and epoch count are adjusted to balance the training efficiency with response accuracy. Moreover,

228 temperature tuning is necessary to adjust the model's creativity while responding to technical questions. One of
229 the main challenges is mitigating the hallucinations – incorrect or fabricated information – by calibrating loss
230 functions to define domain-specific constraints. The overall process is divided into different steps. The training
231 process, which begins with training on general built heritage concepts and then progressively incorporates site-
232 specific datasets, allows the model to produce increasingly specialized outputs.
233 As in the previous section, the evaluation phase is performed by a qualitative assessment based on accuracy,
234 terminology correctness, completeness, and relevance. An iterative feedback loop identifies errors and
235 knowledge gaps to help refine datasets and adaptive retraining.
236 This phase sets the basis for the Retrieval Augmented Generation (RAG), where real-time querying capabilities
237 will further enhance the model's adaptability.

238 3.3 *AI Assistant with Retrieval-Augmented Generation (RAG)*

239 In this phase, the AI framework combines pre-trained generative models with real-time information retrieval
240 from external sources. This approach enables access to both structured data (in JSON format) and unstructured
241 information (archival documents, PDFs, and text-based sources), allowing more refined and context-aware
242 responses.

243 The system architecture is designed as a modular framework (backend and frontend) using Node.js and
244 Express.js to handle API requests and response delivery. The AI Assistant employs the OpenAI GPT-4-turbo
245 model, which is integrated through the OpenAI Node Client and deployed in Replit, a cloud-based development
246 environment that facilitates code execution.

247 The core aspect of this phase is the file ingestion and preprocessing pipeline, which enables the AI Assistant to
248 process both structured (JSON) and unstructured (PDFs and text) data. The system "ingests" all the information
249 and converts it into vector embeddings. These embeddings are stored in a vector database, allowing semantic
250 search and retrieval based on natural-language prompts.

251 In order to understand these different types of information, the AI Assistant employs a dual-retrieval strategy,
252 a File Search tool, to query and retrieve relevant data efficiently from text-based sources. For structured data in
253 JSON format, specific Function-calling tools have been developed, allowing the AI Assistant to directly query
254 and extract targeted information from these files, thereby enhancing precision in responses.

255 After that, to handle query execution, the backend defines two endpoints: /start, which initializes a thread for
256 each conversation, and /chat, which allows the user to send queries that the AI Assistant processes by retrieving
257 the relevant information from the vector store.

258 When a user sends a natural-language query (e.g., about historical aspects or building components), the AI
259 Assistant matches the request with the indexed structured and unstructured data, retrieves the information, and
260 returns a response. The system is designed to handle error scenarios, log processes, and ensure that accurate
261 responses are delivered to the user through the frontend.

262 Users interact with the AI Assistant through a web-based application or any other platform, where they can
263 submit queries and receive responses in real-time. The frontend collects user inputs (queries) and transmits them
264 to the backend for processing and using the available data. Finally, the backend will send the response back for
265 display to the user.

266 The AI Assistant underwent an initial qualitative performance assessment in the OpenAI API playground.
267 Prompts were designed to test the AI Assistant's focus on relevant responses and its handling of basic heritage-
268 related queries. The assessment focused on criteria such as accuracy, relevance, completeness, clarity,
269 conciseness, interactivity, and adaptability. Thanks to these evaluation criteria, it is possible to continuously
270 refine the AI model, ensuring that retrieval-augmented querying facilitates knowledge accessibility. This phase
271 moves beyond the static AI model developed through the pentagram framework by integrating the retrieval
272 mechanisms. The creation of an AI interactive assistant allows multi-source queries, combining a dynamic and
273 modular approach to support reasoning and inference for heritage interpretation.

274 The AI Assistant helps reduce AI hallucinations since the responses are grounded in accurate data, avoiding the
275 generation of speculative or incorrect information.

276 This model can also handle evolving heritage datasets, and as historical knowledge expands, new studies and

277 documentation can be integrated into the Assistant without requiring frequent retraining. Additionally, the
 278 development of AI Assistants facilitates interdisciplinary queries since they can respond to architectural,
 279 historical, archaeological, and technical inquiries and cross-reference domain-specific information from
 280 different sources (Figure 2). Overall, the framework offers a clear and verifiable way to integrate generative AI
 281 into heritage workflows, advancing toward strategies that facilitate the interpretation, verification, and
 282 interdisciplinary comprehension of heritage knowledge.
 283

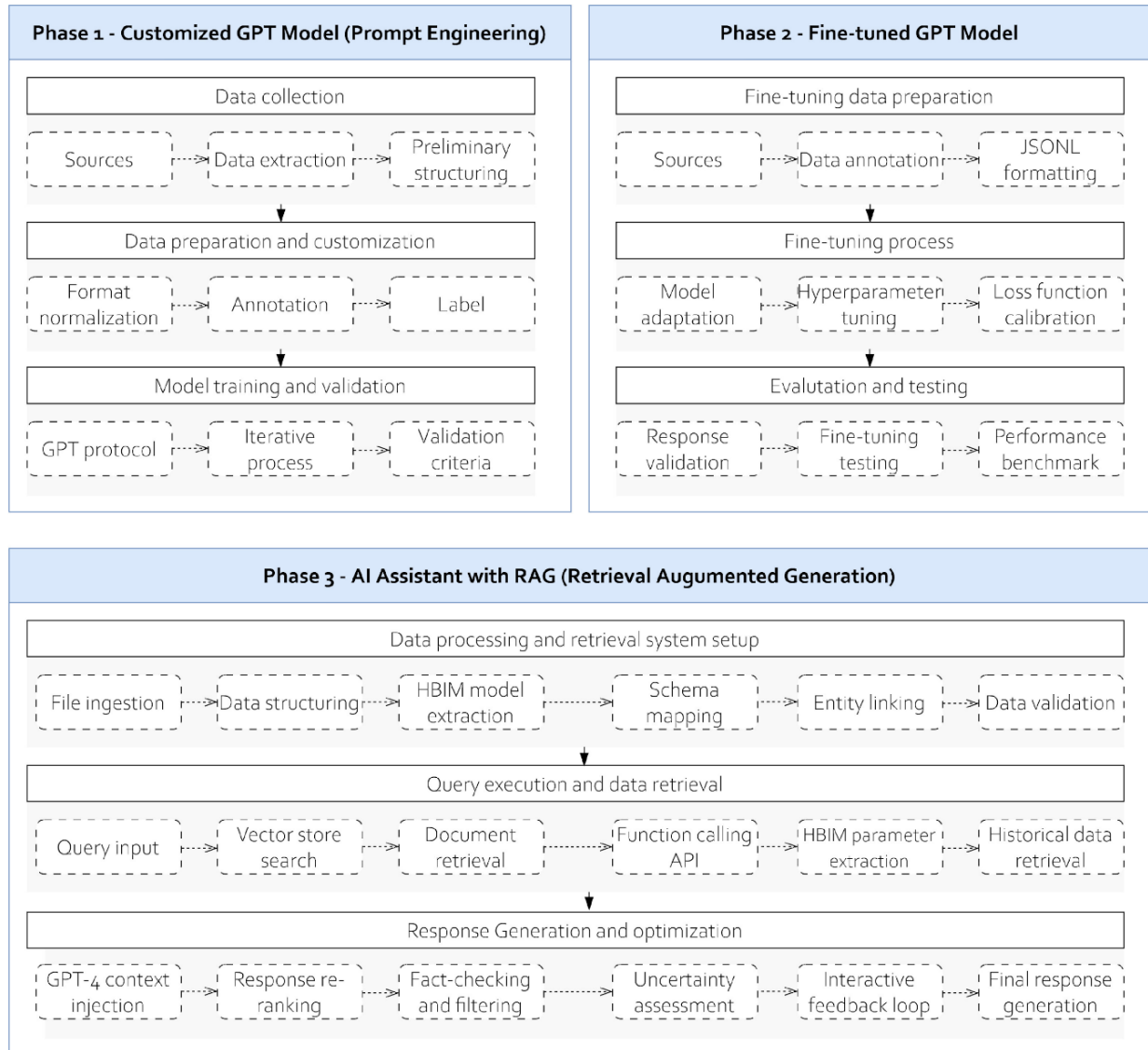


Figure 2. Detailed workflow phases: (1) Customized GPT via prompt engineering, (2) Fine-tuned GPT with structured datasets, (3) AI Assistant with RAG integrating HBIM and archival documents. © 2025, the Authors.

284 4. Case study application

285 The proposed framework is applied to the case study of the Sanctuary of Hercules and the Former Segrè Paper
 286 Mill in Tivoli (Italy), a site shaped by centuries of human and natural activities. Over the past centuries, this
 287 unique asset has undergone multiple functional adaptations, from ancient Roman structures to numerous
 288 industrial reuses, culminating in its final abandonment in the 1970s, when the Segrè family's latest production
 289 process was decommissioned [30].

290 The extraordinary value of this site lies in the multiple stratifications and in the possibility of highlighting those
 291 historical layers in informed intervention actions for the recovery and valorization processes.

292 A key challenge in this case study is managing a large volume of data that contains inconsistencies,

293 uncertainties, and changes over time. For this reason, the objective of this study is to evaluate the model's
 294 performance using 20 crafted prompts categorized into three major domains: historical evolution, architectural
 295 and construction aspects, and interdisciplinary knowledge (Figure 3). This application, therefore, tests the AI's
 296 ability to interpret built heritage data using natural language queries and analyzes the model's progressive
 297 refinement across three phases – Customized GPT, Fine-tuned GPT, and AI Assistant with RAG - highlighting
 298 how LLMs evolve from static knowledge to dynamic retrieval reasoning. Moreover, this research aims to define
 299 the possible frontiers and limitations of heritage management and to explore the potential of hybrid AI-assisted
 300 decision-making processes.
 301

Prompt ID	Category	Prompt Question
HIST1001	Historical Evolution	Describe the evolution of the Sanctuary of Hercules and the Former Segrè Papermill over centuries. What were the key periods of significant change?
HIST1002	Historical Evolution	What were the main economic and functional drivers behind the conversion of the site into an industrial complex?
HIST1003	Historical Evolution	How did the water management infrastructure support industrial production at the Segrè Papermill?
HIST1004	Historical Evolution	Can you deduce the latest papermill production process?
HIST1005	Historical Evolution	How did the site's ownership change over time, and how did this impact its architectural evolution?
HIST1006	Historical Evolution	What were the main socio-economic factors that led to the abandonment of the Segrè Papermill?
TECH1007	Architectural & Constructive Aspects	What were the predominant construction techniques used in the original Sanctuary of Hercules?
TECH1008	Architectural & Constructive Aspects	What construction materials were used in the former Segrè Papermill, and how did they differ from the original structures?
TECH1009	Architectural & Constructive Aspects	How was opus caementicium used in the Sanctuary of Hercules, and how does it compare to modern reinforced concrete?
TECH1010	Architectural & Constructive Aspects	What reinforcements were introduced in the early 20th century industrial adaptations of the site?
TECH1011	Architectural & Constructive Aspects	How was the storage room of the Segrè Papermill constructed in the 20th century?
TECH1012	Architectural & Constructive Aspects	How did the specific requirements of papermaking processes influence the selection of construction materials and the structural design of the Segrè Papermill?
TECH1013	Architectural & Constructive Aspects	How did the choice of masonry techniques at the Sanctuary of Hercules influence the site's structural stability and its ability to withstand environmental degradation over time?
INTER1014	Interdisciplinary Knowledge	What types of machinery were used in the Segrè Papermill, and what role did they play in the production process?
INTER1015	Interdisciplinary Knowledge	How do the work tools found at the Segrè Papermill reflect technological advancements of the early 20th century?
INTER1016	Interdisciplinary Knowledge	How did industrial processes at the papermill evolve from its inception to its abandonment?
INTER1017	Interdisciplinary Knowledge	How do the surviving architectural elements contribute to the site's heritage value today?
INTER1018	Interdisciplinary Knowledge	What is the significance of the Sanctuary of Hercules in relation to industrial archaeology?
INTER1019	Interdisciplinary Knowledge	How did water-powered machinery at the papermill influence the daily workflow and production efficiency in the early 20th century?
INTER1020	Interdisciplinary Knowledge	What are the key conservation challenges in balancing industrial archaeology values with modern reuse proposals for the site?

Figure 3. Crafted 20 prompts for evaluating AI model responses across three domains: Historical Evolution, Architectural & Construction Aspects, and Interdisciplinary Knowledge. © 2025, the Authors.

302 *4.1 Customized BHGPT model*

303 Tailoring and adjusting the AI tool represent the main focus of this phase. The development of BHKR (Built
 304 Heritage Knowledge Retriever) involves a detailed customization of the GPT model to align it with specific
 305 requirements. The functionality of the BHKR relies first on aggregating diverse data types from various sources
 306 into a coherent structure that simplifies access and manipulation. The structure then provides outputs useful to
 307 the stakeholders involved in the decision-making process, with a focus on precision and appropriateness of
 308 context.

309 4.1.1 *Pentagram framework*

310 Here is the detailed definition of the pentagram framework that is provided to the GPT model:
311 «Context: Your users are professionals from diverse fields such as architecture, engineering, archaeology, and
312 history. These users often face challenges due to the unstructured and poorly integrated nature of existing data,
313 which complicates the accessibility, management, and sharing of information. Common issues include data
314 duplication, inconsistencies, and loss. Your role is crucial in delivering structured knowledge in a user-friendly
315 manner to improve data exchange processes and significantly enhance decision-making accuracy. You can
316 access a rich repository of historical and current images, detailed documentation, and other pertinent data about
317 specific sites, including the Sanctuary of Hercules and the Former Segrè Papermill.
318 Persona: You are "BHKR" (Built Heritage Knowledge Retriever), a highly advanced, friendly, and adaptable
319 AI Chatbot designed to support stakeholders involved in the decision-making processes for interventions and
320 valorization activities at heritage sites. Your core function is to access and synthesize a wide array of data,
321 including image-based and text-based resources that detail the historical evolution of sites, their multifaceted
322 uses over time, and the building components and materials used in different historical periods.
323 Task: Using natural language prompts, you are tasked with extracting relevant concepts and relations from a
324 comprehensive database that includes but is not restricted to building information data. Your responses should
325 address queries related to the historical evolution of the site, different functions and reuses of spaces, and
326 specifics of building components and materials across various historical periods.
327 Output: Your output consists of text-based responses generated from a GPT model that has been trained to
328 include all relevant information about the Sanctuary of Hercules and the Former Segrè Papermill. This model
329 uses domain knowledge and structured information to answer prompts related to changes in space functions
330 over time, building components, and materials. The responses are designed to be directly actionable and
331 informative for heritage site management and conservation efforts.
332 Constraints: Focus primarily on mapping the information from natural language queries to provide accurate,
333 clear, and contextually appropriate responses. Avoid straying into general advice or areas outside the defined
334 scope. Ensure all content is suitable and precise and respects the objectives of the users, making it directly
335 applicable to heritage site management and conservation planning».

336 4.1.2 *Iterative prompt engineering and application*

337 The customization of the AI model involved several iterations of prompt engineering. Each iteration is labor-
338 intensive, as manual work is required to refine the prompts for data input to the GPT model and to inspect the
339 model's output. One of the aims of the application was to show and discuss the learning process and how data
340 from various sources, such as textual records, archival photographs, construction drawings, and structured
341 metadata, are memorized or even understood by the AI model when stored in the knowledge base, as reflected
342 by its responses to questions. Because no similar evaluation has been conducted in the literature, this study
343 provides a rigorous evaluation of a generative AI model in the field of built heritage.

344 Understanding and acknowledging the model's behavior based on different inputs was necessary.
345 For this application, we used GPT-4.0, with a total of seven iterations and steps. The qualitative prompt response
346 validation was followed at every stage when the knowledge base of the model was updated with different types
347 of information.

- 348 - Pre-trained GPT-4.0. The first stage of the model testing resulted in very vague answers. Indeed, the
349 responses were based on general knowledge and standard practices, and when follow-up questions tried
350 to focus on more specific information, some parts were made up by the AI model.
- 351 - GPT pentagram framework. The protocol integration and model customization did not change the overall
352 outputs for the different prompts. Thanks to the follow-up questions, the aspects that improved were the
353 conciseness and clarity of the information. This stage highlighted the high interactivity and adaptability
354 of the model.
- 355 - Present and Historical Images. This phase provided the model with a series of current images of the site
356 and images derived from archival research with some existing labels (such as the source's provenance

357 and synthetic descriptions of the image's content). The model improved some of the answers with the
358 newly added information, only when there was a description of the images.
359 - Historical documentation (text). After adding textual information, all of the prompt's answers became
360 much more precise, particularly those related to historical aspects. In order to get the structured answers,
361 some follow-up questions were used.
362 - Architectural and construction documentation (drawings). First, some drawings were added to the model
363 to improve the responses related to construction knowledge. Even in this case, the model improved its
364 answers only when annotations or textual descriptions were available.
365 - Architectural and construction documentation (text). Text data then followed the drawings. At this stage
366 of the refinement process, the model interactions improved in accuracy, completeness, and relevance.
367 - Labeled images. The last phase of this training was labeling the architectural and construction drawings.
368 Figure 4 shows the AI model's final answers to the prompts HIST1002 and TECH1011 as examples. Overall,
369 the model can process text-based information. Meanwhile, if enriched with labels and metadata tagging, image-
370 based information is practical in the learning stages; otherwise, the model is not able to interpret it correctly.
371 The GPT's answers to HIST1002 have improved in accuracy, but after several follow-up questions, it struggled
372 to provide precise responses based on spatial data, for instance, in HIST1004. Meanwhile, for TECH1011, the
373 GPT started with general and made-up information, and the information's inaccuracy changed with labeled
374 images and specific textual information. In this second case, it was more challenging to allow the model to focus
375 on the case study rather than broaden or respond with general knowledge.
376 Generally, some challenges arose during the application of the AI model. The inconsistencies due to the different
377 sources in the gathered data lead to intensive manual work to format the information. Gaps in the records led to
378 incomplete or uncertain inputs, and distinguishing accurate information was challenging. A prompt refinement
379 process was implemented to address these difficulties based on model outputs; however, it remained challenging
380 to control the model's performance fully.
381

What were the main economic and functional drivers behind the conversion of the site into an industrial complex?

The conversion of the Sanctuary of Hercules Vincitor into an industrial complex was driven by key economic and functional factors:

Economic Drivers

- **Strategic Trade Position:** Tivoli's proximity to Rome and its established commercial role since antiquity made it a prime location for industrial development.
- **Industrialization & Market Demand:** The rise of mechanized papermaking in the 19th century increased demand for production facilities, leading to the foundation of the Segrè Papermill in 1887.
- **Hydropower Utilization:** The Aniene River and existing Roman hydraulic infrastructure enabled cost-effective energy use, reducing operational expenses.
- **Employment & Economic Growth:** The industrialization of Tivoli created jobs, transforming the local economy and attracting investment.

Functional Drivers

- **Pre-Existing Roman Infrastructure:** The large arcades, porticoes, and Via Tecta provided ready-made structural support, minimizing construction needs.
- **Water as an Energy Source:** Roman-built hydraulic channels were adapted for industrial use, first in proto-industrial workshops and later for hydroelectric power.
- **Space for Large Machinery:** The sanctuary's vast terraces and arcades were ideal for housing industrial equipment, such as the *macchina continua* for paper production.

These factors made the site an ideal candidate for industrial repurposing, ensuring its continuous economic relevance from antiquity to the modern era.

(a)

You
Could you list the construction techniques used for the storage room of the Segrè papermill build during the XX century?

BHKR - Built Heritage Knowledge Retriever
The construction techniques used for the storage room of the Segrè Papermill, built during 1934-1935, included:

- **Mixed reinforced concrete structure:** This technique utilized beam pillars and a solid slab, forming the core framework of the storage room.
- **Perimeter load-bearing masonry:** The walls were constructed with load-bearing masonry, providing structural support and enclosure.
- **Foundations on substructures of the Sanctuary of Hercules:** The storage room was built upon the existing substructures of the historical site, integrating the new construction with the old.
- **Use of local materials:** The concrete for the structure was mixed on-site and incorporated plain reinforcement bars. The masonry pillars were made from squared tuff blocks and brick recourses, and the masonry itself featured hewn tuff blocks with buttresses of squared tuff blocks and brick recourses.

These construction techniques reflect the architectural practices of the early 20th century in Italy, emphasizing durability and the use of local materials, while also adapting and integrating with the historical context of the site.

(b)

Figure 4. Example responses from the customized GPT model (BHKR) to prompts (a) HIST1002 and (b) TECH1011, showing improvements in historical accuracy and construction detail after iterative prompt engineering and data enrichment. © 2025, the Authors.

382
383

384 4.2 Fine-tuned BHGPT model beyond static customization

385 While the Customized GPT Model (Section 4.1.) established a structured approach to use a heritage-specific AI
386 model, it remained constrained by its reliance on pre-existing knowledge and prompt engineering. Despite
387 unstructured data ingestion, the model's outputs were still susceptible to hallucinations, lacked granular
388 specificity, and struggled with adaptive reasoning over multiple sources. This phase introduces fine-tuning—a
389 more advanced approach where the model is trained on domain-specific datasets in a structured format
390 (JSONL). This training allows the model to learn patterns, terminology, and historical reasoning in a specialized
391 context, reducing reliance on generic pre-trained knowledge based on a dataset of 30 structured examples for
392 fine-tuning the GPT model, specifically for the Sanctuary of Hercules and the Former Segrè Papermill case
393 study. These examples are formatted in JSONL (JSON Lines), each containing a system prompt (context
394 setting), a user query, and a response (ground truth).

395 Practically, fine-tuning helps structure responses by providing contextual background information, teaches
396 relevant terminology in historical, architectural, and interdisciplinary domains, and allows understanding of
397 patterns of reasoning by providing the expected depth of responses.

398 The training process starts with general knowledge fine-tuned on built heritage concepts; after that, it
399 incorporates site-specific datasets and evaluates the performance through iterative testing and response
400 validation. The core of the fine-tuning activity relies on the model learning of historical and architectural
401 reasoning patterns, rather than on memorizing answers.

402 In Figure 5, two general examples are shown: the first focuses on the historical evolution of training the model
403 on key periods and changes, while the second focuses on technological change and its impact on socio-economic
404 factors.
405

```
{ "messages": [
  { "role": "system", "content": "You are a built heritage knowledge retriever specialized in historical site transformations." },
  { "role": "user", "content": "Provide a structured summary of how industrial sites often repurpose ancient infrastructures." },
  { "role": "assistant", "content": "Stratified industrial sites frequently repurpose existing structures due to their robust masonry, strategic location, and pre-existing water management systems. For example, old Roman aqueducts and mills were often adapted into paper production sites, which utilized medieval foundations and hydraulic channels to power its machinery." }
]
```

(a)

```
{ "messages": [
  { "role": "system", "content": "You analyze the socio-technical impacts of industrial heritage and technological advancements." },
  { "role": "user", "content": "What role did water-powered machinery play in early papermaking processes?" },
  { "role": "assistant", "content": "Water-powered machinery was central to early papermaking, particularly in the 18th and 19th centuries. Waterwheels converted kinetic energy from flowing water into mechanical energy, driving beaters that processed pulp. This method significantly increased production efficiency compared to manual processes, allowing for higher paper output and uniform quality. Over time, water-powered systems were replaced by steam and electric-driven machines, which further revolutionized the industry." }
]
```

(b)

Figure 5. Fine-tuned dataset in JSONL format related to (a) historical evolution and (b) technological changes.

© 2025, the Authors.

406

407 The effectiveness of the results largely depends on the selection of hyperparameters. Three of these cases are
408 configured based on the training dataset. The batch size is set to 1 since the dataset is relatively small, which
409 allows the model to have stable updates. The learning rate multiplier is set to 0.2 in order to balance the model's
410 ability to adapt to new domain-specific patterns while avoiding excessive divergence from its pre-trained
411 knowledge. Finally, three epochs are used to ensure sufficient exposure to the training dataset without risking
412 overfitting.

413 As shown in Figure 6, the fine-tuned model (*ft:gpt-4o-2024-08-06:personal:bh:B8W2gfy:ckpt-step-60*) relies
414 primarily on the knowledge embedded during the training process. This means that the model does not fetch
415 new documents or updates, and even if it is more accurate within the trained domain, it may not improve its
416 responses. It may generate responses based on outdated or incomplete data, remaining static unless the model
417 is retrained with additional information.

418 One of the cases involved prompt INTER1011, which assesses the model's ability to provide a detailed
419 explanation of machinery used in the Segrè Papermill. After fine-tuning with case-specific details, its outputs
420 became more precise, correctly identifying pulping machines, paper machines, cutting machines, and packaging
421 machines as key components of the production process in the papermill. It improved the contextual relevance

422 but failed to surpass the general model in technical depth. In order to achieve better results, the training process
 423 has to be enriched with a much larger dataset and a detailed breakdown, along with the possibility of
 424 incorporating retrieval augmentation.
 425

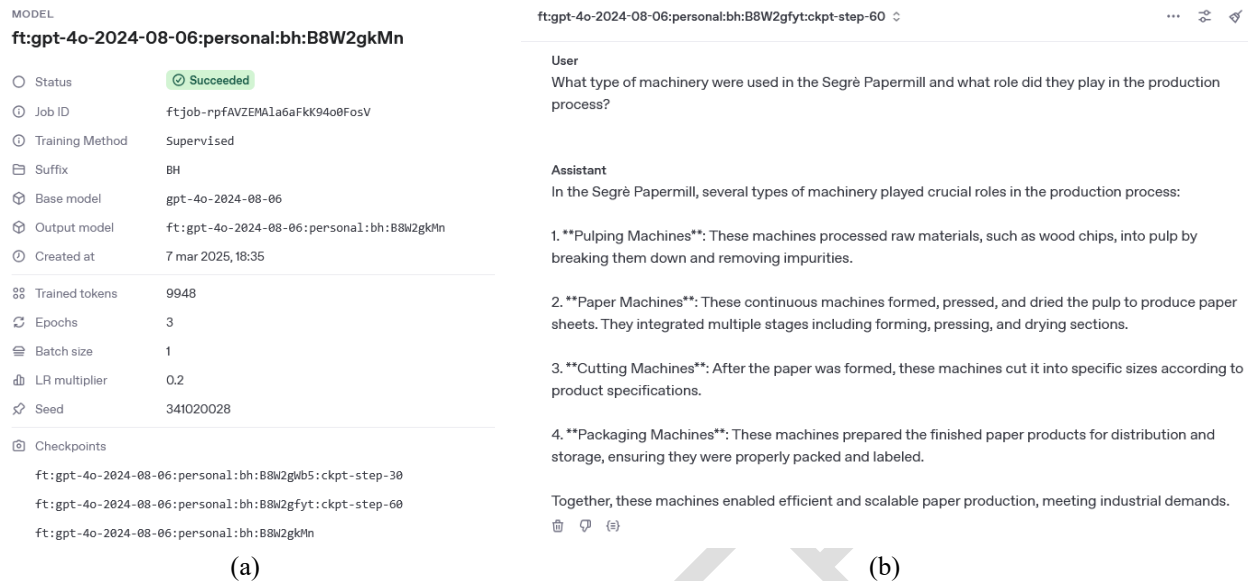


Figure 6. Features and output of the fine-tuned BHGPT model, showing (a) training parameters and (b) an example response demonstrating improved site-specific reasoning on historical machinery. © 2025, the Authors.

426 4.3 AI Assistant with Retrieval-Augmented Generation (RAG)

427 While fine-tuning the BHGPT model enhanced domain-specific knowledge by training on historical and
 428 architectural reasoning patterns, it remained constrained to the learned dataset. To address the limitation of
 429 retrieving newly introduced information, Retrieval-Augmented Generation was integrated into the AI Assistant,
 430 enabling queries across structured and unstructured data and cross-referencing of sources.

431 The application of this method requires creating an AI Assistant via the OpenAI API, where the AI Assistant
 432 backend orchestrates interactions between the knowledge base (vector store) and AI tools. A specific function-
 433 calling mechanism is implemented (*query_room_info*) to allow the users to retrieve specific HBIM model
 434 information via natural language queries. Unlike fine-tuning, which improves the AI's comprehension of
 435 domain-specific terminology and reasoning, the RAG-based AI Assistant queries structured HBIM data
 436 converted and extracted into JSON format (spatial, material, and event parameters) and cross-checks
 437 unstructured historical sources in order to gain a deeper understanding of interpretation processes.

438 The ability to access multiple tools at the same time provides the capability, for instance, to determine room
 439 functions in the latest Segrè Papermill, linking the Paper Machine Room (with a Globally Unique Identifier or
 440 GUID) to the production of paper sheet processes, inferred from both HBIM model parameters and descriptions
 441 found in historical manuals.

442 Building upon these preliminary deductions, the AI Assistant can identify the logical sequence of the historical
 443 production process by cross-referencing spatial-structural data with textual archival records. By mapping room
 444 GUIDs to specific industrial processes, the Assistant established a comprehensive reconstruction of the Segrè
 445 Papermill's 1935 production cycle, identifying how raw materials progressed through sorting, refining, and
 446 bleaching stages before reaching the final paper production phase.

447 As shown in Figure 7, this approach improved what was achievable through customized GPT and fine-tuning
 448 alone. However, in this case, the model's dependence on the quality of the input data remains challenging.
 449 Incomplete HBIM parameters and the lack of archival documentation sometimes lead to spurious AI-generated
 450 assumptions. Despite these limitations, combining an AI Assistant with RAG-based retrieval represents a
 451 significant step toward enhanced, data-driven heritage analysis and interpretation.
 452

User

Starting from the information stored in your knowledge base, can you deduce the latest papermill production process?

Chat Demo

From the detailed documentation and historical data available, the latest paper production process at the Former Segrè Papermill, just prior to its ceasing operations, involves several key stages that utilize specific rooms and equipment effectively:

1 Raw Material Preparation:

- **Location:** Storage Room (Room 2Su\$w2ZJP0g9her_EcsDJa)
- **Process:** Raw materials, such as rags, other fibrous substances, and additives needed for paper production, are stored here. They are also pre-processed and prepared for further refinement.

2 Pulping Process:

- **Location:** Hollander Beater Room (Room 15vbbt9mz90BB0IDzgAHLE)
- **Process:** The raw materials are transferred to the Hollander beaters for pulping. These machines break down the materials into a fine pulp, ensuring a smooth and homogeneous consistency vital for quality paper production.

3 Paper Formation:

- **Location:** Continuous Paper Machine Room (Room 2Su\$w2ZJP0g9her_EcsEeN)
- **Process:** The pulp is then fed into a continuous paper machine. This machine forms the pulp into long sheets of paper by removing water and compressing the fibers. The continuous nature of this machine allows for efficient and high-volume production.

4 Drying and Finishing:

- **Location:** Paper Preparation and Finishing Room (Room 2Su\$w2ZJP0g9her_EcsDuF)
- **Process:** Following formation, the paper goes through additional drying processes to remove any remaining moisture, improving its strength and printability. The paper may also go through finishing techniques such as glazing, coating, or sizing to enhance its quality and appearance.

5 Sorting and Storage:

- **Location:** Event Room and Paper Sorting (Room 3PBdP3axr3\$OU6mUBF6Eac)
- **Process:** The finished paper is sorted according to different grades and types. It may be temporarily stored in this room, which also serves multiple purposes, including as a venue for various events related to mill operations or community activities.

6 Packaging and Distribution:

- **Process:** Finished paper products are packaged according to specifications for commercial distribution. The mill handles the logistical aspects of sending these products to customers or distributors, ensuring that they reach their destinations in optimal condition.

(a) (b)

Figure 7. AI Assistant with RAG reconstructs the Segrè Papermill's production process by combining spatial HBIM data and textual archival sources (a)(b). © 2025, The Authors.

453 5. Results and Discussion

454 The evaluation of the 20 structured prompts, shown in Figure 3, highlights the strengths and limitations of the
 455 three AI-driven methodologies – Customized GPT, Fine-Tuned BHGPT, and AI Assistant with Retrieval-
 456 Augmented Generation (RAG) – across historical evolution, architectural and construction aspects, and
 457 interdisciplinary knowledge. All three methods were tested on the same prompts to ensure a coherent
 458 comparison.

459 The customized GPT model provided structured responses using prompt engineering and showed incremental
 460 improvements when text-based information was added to its knowledge base. However, its reliance on pre-
 461 existing knowledge resulted in possible hallucinations, particularly when addressing case-specific construction
 462 techniques and technical questions. This limitation was evident in prompts based on historical building materials
 463 and structural adaptations, in which the model often relied on broad heritage knowledge rather than case-specific
 464 data. The best uses are rapid scoping and general overviews; it is not recommended for site-specific technical
 465 details.

466 Fine-tuning the model with 30 structured cases significantly improved domain knowledge, refining its ability
 467 to recognize historical and architectural patterns related to the Sanctuary of Hercules and the Former Segrè
 468 Papermill. In this phase, the model encountered difficulties with interdisciplinary knowledge questions because
 469 its training set was limited and unable to retrieve information beyond it, thereby restricting its adaptability in
 470 addressing socio-economic impacts or industrial processes across different periods. For this model, the best use
 471 is consistent site-specific reasoning and terminology within a well-defined domain.

472 Finally, the AI Assistant with RAG addressed these issues by integrating real-time retrieval from structured
 473 (HBIM) and unstructured (archival) sources, improving contextual awareness and historical reasoning. It
 474 successfully reconstructed the latest Segrè Papermill production cycle by correlating HBIM-derived spatial data
 475 with historical records. Limitations arose with sparse records, leading to cautious or incomplete outputs.
 476 Enhancing dataset completeness, refining retrieval mechanisms, and integrating expert validation will be crucial
 477 for optimizing AI-driven heritage interpretation. The most effective use case for the AI Assistant is the
 478 integration of multiple information sources, both structured and unstructured.

479 6. Conclusions

480 The paper proposes an innovative approach to integrate AI-driven methodologies in heritage conservation and
 481 valorization activities. The presented framework introduces GPT models to enhance knowledge querying and
 482 extraction within heritage activities, progressing from static GPT customization and fine-tuned models to
 483 dynamic AI Assistants with Retrieval-Augmented Generation (RAG). Through a comparative evaluation of
 484 three methodologies—Customized GPT, Fine-Tuned BHGPT, and AI Assistant with RAG—this research

485 highlights the advantages and constraints of each approach.
486 The Customized GPT model, through prompt engineering, provides a fast and structured method for generating
487 site-specific reasoning. In contrast, the fine-tuned model, trained on a specific dataset, improves the ability to
488 process domain-specific terminology and historical reasoning, but the dataset constrains the answers. The AI
489 Assistant with RAG overcomes these limitations by integrating multi-source data, allowing dynamic queries.
490 The defined framework, applied to the Sanctuary of Hercules and the Former Segrè Papermill, demonstrates
491 how AI-driven knowledge retrieval can enhance historical analysis and interpretation.
492 The novelty of this framework lies in its ability to overcome the limitations of the current and traditional
493 methodologies. Enhancing data accessibility, interpretation, and accuracy supports informed decision-making
494 processes. The integration of GPT models allows for more efficient and effective knowledge management. As
495 digital technologies and AI continue to evolve, future advances will enhance the model's accuracy, consistency,
496 and performance. Future research will need to include feedback from specialists and the general public on the
497 utility and impact of the AI outputs, thereby extending the applicability of this model.
498 In the built heritage context, once the behavior and the effectiveness of each type of approach are understood,
499 the final aim will be focused on creating and generating new knowledge from the existing knowledge bases.
500 The possibility of querying more than one case study and creating relations between different classes of
501 information will be the center of future work.
502 The research frames generative AI workflows as a structured and verifiable method for reasoning across
503 heterogeneous heritage data sources. However, among the various challenges, model updates and validation are
504 constantly needed to ensure the relevance and impact on the practice. This methodological contribution will
505 facilitate interoperability and the creation of a comprehensive heritage system that combines AI with other
506 digital tools, enabling data analysis and visualization in a new era of heritage valorization.

507 7. Acknowledgements

508 This paper is an extended version of our previous work which was presented at the XII Edition of the Reuso
509 Conference (2024).

510 8. Authors Contributions

511 Conceptualization, C.D.L.C., W.Y., A.F. and E.C.; Methodology, C.D.L.C., W.Y., A.F. and E.C.; Software,
512 C.D.L.C.; Investigation, C.D.L.C.; Writing – Original Draft Preparation, C.D.L.C. and W.Y.; Writing – Review
513 & Editing, C.D.L.C. and W.Y.; Visualization, C.D.L.C.; Supervision, E.C., A.F., W.Y.

514 9. References

- 515 [1] Yang X, Grussenmeyer P, Koehl M, et al (2020) Review of built heritage modelling: Integration of
516 HBIM and other information techniques, *Journal of Cultural Heritage*, 46: 350–360.
517 doi:10.1016/j.culher.2020.05.008.
- 518 [2] Aiello A, Pierobon-Benoit R, Proto F (2005) Concettualizzazione e contestualizzazione dei beni culturali
519 archeologici, *Archeologia e Calcolatori*, 16: 321–339.
- 520 [3] Bruno N, Roncella R (2019) HBIM for conservation: A new proposal for information modeling, *Remote*
521 *Sensing*, 11(15): 1751. doi:10.3390/rs11151751.
- 522 [4] Pauwels P, Zhang S, Lee YC (2017) Semantic web technologies in AEC industry: A literature overview,
523 *Automation in construction*, 73: 145–165. doi:10.1016/j.autcon.2016.10.003.
- 524 [5] Saka A, Taiwo R, Saka N, et al (2024) GPT models in construction industry: opportunities, limitations,
525 and a use case validation, *Developments in the Built Environment*, 17, 100300.
526 doi:10.1016/j.dibe.2023.100300.
- 527 [6] Liu P, Yuan W, Fu J, et al (2023) Pre-train , Prompt , and Predict : A Systematic Survey, *ACM*
528 *computing surveys*, 55(9): 1–35. doi:10.1145/3560815.
- 529 [7] Lombardi M, Rizzi D (2024) Semantic modelling and HBIM: A new multidisciplinary workflow for
530 archaeological heritage, *Digital Applications in Archaeology and Cultural Heritage*, 32: 00322.

- 531 doi:10.1016/j.daach.2024.e00322.
- 532 [8] Radford A, Wu J, Child R, et al (2018) Language Models are Unsupervised Multitask Learners,
533 Technical report, OpenAi.
- 534 [9] Raffel C, Roberts A, Matena M, et al (2020) Exploring the Limits of Transfer Learning with a Unified
535 Text-to-Text Transformer, *Journal of machine learning research*, 21(140): 1–67.
- 536 [10] Min S , Lyu X, Holtzman A, et al (2022) Rethinking the Role of Demonstrations: What Makes In-
537 Context Learning Work?. In *Proceedings of the 2022 Conference on Empirical Methods in Natural*
538 *Language Processing*: 11048–11064, Abu Dhabi, United Arab Emirates. Association for Computational
539 Linguistics.
- 540 [11] Zheng J, Fischer M (2023) Dynamic prompt-based virtual assistant framework for BIM information
541 search, *Automation in Construction*, 155, 105067. doi:10.1016/j.autcon.2023.105067.
- 542 [12] You H, Ye Y, Zhou T, et al (2023) Robot-Enabled Construction Assembly with Automated Sequence
543 Planning Based on ChatGPT: RoboGPT, *Buildings*, 13(7): 1772. doi: 10.3390/buildings13071772.
- 544 [13] Prieto SA, Mengiste ET (2023) Investigating the Use of ChatGPT for the Scheduling of Construction
545 Projects, *Buildings*, 13(4), 857. doi:10.3390/buildings13040857
- 546 [14] Uddin SMJ, Albert A, Ovid A (2023) Leveraging ChatGPT to Aid Construction Hazard Recognition
547 and Support Safety Education and Training, *Sustainability*, 15(9): 7121. doi:10.3390/su15097121
- 548 [15] Firmani D, Merialdo P, Maiorino M, Nieddu E (2018) Towards knowledge discovery from the Vatican
549 secret archives. In *codice ratio - episode 1: Machine transcription of the manuscripts*. In: *Proceedings of*
550 *the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 263–
551 272. doi:10.1145/3219819.3219879.
- 552 [16] Spina S (2023) Artificial Intelligence in archival and historical scholarship workflow: HTS and
553 ChatGPT, *Umanistica Digitale*, 7(16), 125–140. doi:10.6092/issn.2532-8816/17205.
- 554 [17] Loffredo R, De Santo M (2024) Using Ontologies for LLM Applications in Cultural Heritage. In: *CEUR*
555 *workshop proceedings*, vol. 3865, pp. 37–43.
- 556 [18] Otieno PN (2024) Framework for building linguistic corpora for a large language model project for the
557 Heritage Nubian Language of Kenya, *Journal of Languages, Linguistics and Literary Studies*, 4(3): 139–
558 144. doi:10.57040/tpvzpk79
- 559 [19] Zhao Z, Wang D (2025) Evaluation of large language models for the intangible cultural heritage domain,
560 *npj Heritage Science*, 13(1): 1–11, doi:10.1038/s40494-025-02013-1.
- 561 [20] Hwang H, Park CW, Kim HK, Lee JH (2025) CATS: cultural-heritage classification using LLMs and
562 distribute model, *npj Heritage Science*, 13(1): 1–13. doi:10.1038/s40494-025-01621-1.
- 563 [21] Cossatin AG, Mauro N, Ferrero M, Ardissono L (2025) Tell me more : integrating LLMs in a cultural
564 heritage website for advanced information exploration support. In AG Cossatin et al. *Information*
565 *Technology & Tourism*, 27(2), 385-416. doi:10.1007/s40558-025-00312-8.
- 566 [22] Cui CDL, Currà E, Fioravanti A, Yan W. (2024) AI-Powered Built Heritage: Enhancing Interpretation
567 and Recovery Processes with Generative AI Models. In *ReUSO 2024 Documentazione, restauro e*
568 *rigenerazione sostenibile del patrimonio costruito*, pp. 117-126.
- 569 [23] Dabrock K, Johansson T, Donarelli A, et al (2026) Automated Building Heritage Assessment Using
570 Street-Level Imagery, *Building and Environment*, 299: 114654. doi:10.1016/j.buildenv.2026.114654
- 571 [24] Raiaan MAK, Mukta MSH, Fatema K, et al (2024) A Review on Large Language Models: Architectures,
572 Applications, Taxonomies, Open Issues and Challenges, *IEEE Access*, 12: 26839–26874.
573 doi:10.1109/ACCESS.2024.3365742.
- 574 [25] He W, Peng L, Jiang Z, Go A (2025) You Only Fine-tune Once: Many-Shot In-Context Fine-Tuning for
575 Large Language Model, *arXiv preprint*. doi:10.48550/arXiv.2506.11103
- 576 [26] Kim MJ, Finn C, Liang P (2025) Fine-Tuning Vision-Language-Action Models: Optimizing Speed and
577 Success, *arXiv preprint*. doi:10.48550/arXiv.2502.19645
- 578 [27] Mu H, Xu Y, Feng Y, et al (2024) Beyond Static Evaluation: A Dynamic Approach to Assessing AI
579 Assistants' API Invocation Capabilities. In *Proceedings of the 2024 Joint International Conference on*
580 *Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 2342–
581 2353.
- 582 [28] Xie Y, Jiang B, Mallick T, et al (2025) A RAG-Based Multi-Agent LLM System for Natural Hazard
583 Resilience and Adaptation, *arXiv preprint*. doi:10.48550/arXiv.2504.17200
- 584 [29] Paschen U, Pitt C, Kietzmann J (2020) Artificial intelligence: Building blocks and an innovation
585 typology, *Business Horizons*, 63(2): 147-155. doi:10.1016/j.bushor.2019.10.004.

- 586 [30] Currà E, D'Amico A, Angelosanti M (2022) HBIM between Antiquity and Industrial Archaeology:
587 Former Segrè Papermill and Sanctuary of Hercules in Tivoli, Sustainability, 14(3): 1329.
588 doi:10.3390/su14031329.

JUST ACCEPTED