Article

Not peer-reviewed version

GraphCoBots: A Knowledge Graph–Based Distributed and Collaborative Multi-Chatbot System for Museums

Savvas Varitimiadis * , Aristotelis Skamagkis , Georgios Tsakiris , Ioanna Gkika , Konstantinos Kotis *

Posted Date: 3 June 2026

doi: 10.20944/preprints202606.0203.v1

Keywords: museum chatbot; conversational AI; knowledge graphs; distributed AI; collaborative AI; generative AI; multi-agent systems

Article

GraphCoBots: A Knowledge Graph–Based Distributed and Collaborative Multi-Chatbot System for Museums

Savvas Varitimiadis [1,2,*], Aristotelis Skamagkis [2], Georgios Tsakiris [2], Ioanna Gkika[2] and Konstantinos Kotis [1,*]

[1] Intelligent Systems Lab, Department of Cultural Technology and Communications, University of the Aegean, 81100 Mytilene, Greece
[2] Aegean Solutions S.A., Dodekanisou 21, 56429, Thessaloniki, Greece
* Correspondence: svaritimiadis@aegean.gr (S.V.); kotis@aegean.gr (K.K.); Tel.: +30-6972458550 (S.V.); +30-6974822712 (K.K.)

Abstract

In our previous work, a novel approach emphasizing a Knowledge **Graph**-based, distributed, and **Co**llaborative conversational Artificial Intelligence (AI) multi-chat**Bot** system (GraphCoBots) for museums was introduced, towards addressing limitations in scalability, knowledge reuse, and coordinated interaction across conversational services. The architecture was designed to deliver an efficient deployment solution where cultural knowledge can be distributed and shared among different chatbots that collaborate when needed. In this paper, we investigate the realization and effectiveness of this architecture through the implementation and evaluation of the GraphCoBots system. Three chatbot services were developed: (a) Nikos Kazantzakis Museum in Crete, (b) life and work of Nikos Kazantzakis and (c) selected artifacts of the museum. Two knowledge graphs (KGs) were designed and implemented, enabling structured knowledge representation, interoperability, and collaborative reasoning. The system was also designed to utilize external Linked Open Data (LOD) sources. An expert's evaluation was conducted to assess usability and interaction quality. The outcome indicated that users prefer receiving comprehensive information delivered in a human-like manner, without encountering technological barriers or complexity. Based on these findings, the researchers integrated Generative AI services, addressing limitations and improving system efficiency. Finally, a broad audience evaluation was conducted, leading to further improvements.

Keywords: museum chatbot; conversational AI; knowledge graphs; distributed AI; collaborative AI; generative AI; multi-agent systems

---

1. Introduction

Conversational AI technologies have increasingly been adopted by museums and cultural organizations that are aiming for better visitor engagement, accessibility, and personalized interpretation of cultural heritage. Museum chatbots are implemented to provide contextual information, support educational narratives, and enable interaction with collections and artifacts on-site and remotely [1]. More recently, Generative AI–based conversational models have offered new capabilities, due to their ability to produce (in a faster/easier manner) fluent, adaptive, and human-like interactions. However, the integration of such technologies into museum contexts raises practical problems and critical challenges related to data accuracy/privacy, content diversity, knowledge governance, interpretative authority, and ethical management of sensitive or incomplete cultural data [2]. In institutional environments where scholarly responsibility and public trust are essential, keeping curatorial control and transparency is crucial, calling for carefully designed architectural and methodological frameworks [3,4].

Despite recent technological advances, the majority of deployed museum chatbot solutions, whether conversational AI rule-based and intent-driven or powered by Generative AI, exhibit significant limitations. Traditional conversational AI chatbot solutions mostly rely on static intents,

 doi:10.20944/preprints202606.0203.v1

conversational routes and predefined responses. While these solutions offer knowledge control and reliability, they suffer from poor scalability, restricted knowledge, and low public acceptance due to reduced conversational quality [5]. Conversely, conversational systems relying primarily on Generative AI frequently operate without explicit knowledge structure or well-defined domain constraints. This increases the risk of hallucinated content and makes it difficult to guarantee factual accuracy, transparency, and accountability/explainability of generated responses [6]. Furthermore, most existing systems rarely support distributed collaboration among multiple specialized conversational agents, leading to inadequate handling of user questions that refer to institutional, biographical, and artifact-level domains. These limitations underline the need for hybrid architectures that combine structured knowledge management, collaborative conversational behavior, and supervised integration of both Conversational AI and Generative AI technologies in accordance with museum-specific requirements.

Our previous work [5] has introduced a KG-based, distributed, and collaborative multi-chatbot framework tailored for museums, aiming to provide a knowledge-driven chatbot solution that addresses the challenges presented. The proposed architecture supports the distribution of institutional knowledge across multiple specialized chatbots, each responsible for a distinct knowledge domain. This design allows museums to retain curatorial control over publicly disseminated information, while supporting dynamic collaboration when user queries cannot be answered by individual chatbots' knowledge. KGs were selected as the core mechanism for structured knowledge representation, due to their ability to be incrementally populated with new knowledge, reused across conversational contexts, and aligned with disparate and heterogeneous knowledge (interoperability requirements). In addition, the architecture was designed to facilitate integration with external LOD sources, enabling enrichment of institutional knowledge. The implemented system GraphCoBots, was designed based on this framework, with scalability, modularity, and coordinated conversational interaction as central principles, allowing museums to maintain control over their knowledge assets.

Building on this architecture, the present paper examines the actual implementation and effectiveness of the proposed framework [5] through a real-world deployment in the context of the Nikos Kazantzakis Museum in Crete [7]. Three collaborating chatbot services were implemented, addressing distinct yet complementary knowledge domains: (a) information related to Nikos Kazantzakis Museum in Crete, (b) knowledge about life and work of Nikos Kazantzakis and (c) knowledge about selected artifacts from the museum collections. To support these services, two KGs were designed and implemented, enabling structured knowledge representation, interoperability, and collaborative reasoning. The GraphCoBots system was designed to support collaboration among all chatbots, as well as access to external LOD sources for knowledge enrichment.[1] The GraphCoBots system is available online [8], while the source code is available on GitHub repositories .

In addition, the paper explores the integration of Generative AI services into the GraphCoBots framework, motivated by findings from expert usability evaluation and user feedback. The objective is to assess how Generative AI can enhance interaction quality and system efficiency when embedded within a KG–driven, collaborative framework that preserves curatorial oversight and knowledge reliability. The main contributions of this work are summarized as follows:

1. Implementation of a distributed and collaborative multi-chatbot system for museums, grounded in KG technology.
2. Design, development and use of KGs for supporting reusable, continuously updatable, and securely managed cultural heritage knowledge.
3. Empirical evaluation of the GraphCoBots system, including an expert-based usability assessment using the System Usability Scale (SUS) questionnaire, and a broad audience user study, offering insights into interaction quality, usability, and user expectations on a museum chatbot.

[1] Source code GraphCoBots Infobot: https://github.com/svaritimiadiss/GraphCoBots-Infobot. Source code GraphCoBots-NKMA-KGbot: https://github.com/svaritimiadiss/GraphCoBots-NKMA-KGbot. Source code GraphCoBots-NKLW-KGbot: https://github.com/svaritimiadiss/GraphCoBots-NKLW-KGbot. Source code GraphCoBots-KG: https://github.com/svaritimiadiss/GraphCoBots-KG.

4. The controlled integration of Generative AI services within a KG–based conversational architecture, addressing limitations of both traditional and purely generative chatbot approaches, while preserving curatorial oversight and cultural knowledge reliability.

This study seeks to address the following research questions:

1. Do KGs contribute to more robust, reusable, and continuously updatable knowledge management in a distributed multi-chatbot system for museums?
2. Can collaborative multi-chatbot architecture, concerning museums domain, be effectively supported through knowledge obtained from external Web resources?
3. Does the collaborative behavior of multiple chatbots enhance user experience in museums, and what limitations or drawbacks emerge?
4. Can Generative AI be safely and effectively integrated into a KG–driven multi-chatbot architecture while preserving curatorial control and factual reliability over museum artifacts and cultural data?

The rest of this paper is organized as follows. Section 2 reviews related work on conversational AI KG-based chatbots, and multi-chatbot systems. Section 3 presents the GraphCoBots architecture and its implementation. Section 4 describes the evaluations methodology, results and conclusions. Section 5 discusses the findings of the evaluations, especially the integration of Generative AI and the resulting system improvements. Section 6 concludes the paper and outlines directions for future research.

2. Related Work

2.1. Related Work on Conversational AI and KGs

In our previous article [5] a number of chatbot systems were presented. These systems were implemented by combining KGs and Natural Language Processing (NLP), to increase the quality of conversations and the accuracy of the provided knowledge. SynchroBot [9] and OntBot [10] use ontologies created with Protege [11], KGs, and NLP techniques to deliver intelligent, ontology-driven, human-like responses. Disbot, built with the Rasa Framework and linked to a KG, manage domain-specific tasks in disaster support[12]. Structured data and knowledge extraction models like BERT are implemented by other systems, such as PolarisX-bot [13] and Wit.ai-based e-commerce chatbot [14], to enhance their KGs and provide more robust knowledge to end users. Finally, IBM Watson chatbot uses OWL ontologies to support robust, domain-specific conversations [15], while KBot leverages SPARQL querying and multiple Linked Data sources to generate dynamic responses [16]. All of the chatbots systems presented serve as an example of how effectively KGs and NLP can be combined to produce conversational agents that are intelligent, flexible, and context-aware.

Over the past three years, numerous researchers have developed and presented novel conversational AI chatbots that leverage KGs, tailoring their designs to the specific requirements of their respective solutions. During the implementation of EU-funded project CEFAT4Cities, researchers developed a KG-based chatbot aimed at providing multilingual information concerning eGoverment services. The chatbot was developed using the Rasa framework and was connected to the language agnostic KG System, Coreon. The researchers focused on resolving the multilingualism challenge on a KG, by developing individual Natural Language Understanding (NLU) models per language, while keeping the entities referring to the KG language-agnostic. In addition, they focused on the quality of the data that are stored in the KG and defined a process for handling unknown or irrelevant intents by routing such queries to the KG instead of triggering the standard fallback[17] .

A research team from Romania developed a KG-based chatbot implemented using the Rasa open-source framework and GraphDB DBMS. Their work focused on improving how users could obtain accurate responses in cross-domain use cases. They introduced the microworld methodology, in which each domain was trained to provide more specialized and precise responses. Additionally, they explored the use of the KG to store user interaction, map them to KG nodes as properties, and enable dynamic learning and continuous updates[18]. In 2023 a group of researchers from Italy released AIDA–Bot 1 and later AIDA–Bot 2, a KG-based chatbot capable of providing responses related to research articles on various topics and answering open questions by summarizing information from relevant publications. The AIDA–Bot relies on the Academia/Industry DynAmics (AIDA) KG infrastructure and integrates multiple datasets, such as DBpedia, Microsoft Academic Graph (MAG) and others. The AIDA - Bot does not rely on a traditional chatbot framework and was developed in Python. It has been integrated across multiple platforms, including a web interface, Amazon Alexa, Telegram, and even a humanoid robot. A user study was also conducted, providing insightful results[19,20].

Another research team from Greece designed and implemented two KG-based chatbots tailored for providing citizens with information about public services. Both systems were developed using the Rasa open-source framework; one was connected to GRANK.ai, while was linked to the TypeDB KG database. Considering the complexity of public service information, the team chose to guide users either to formulate appropriate questions or to select from predefined options. The researchers also conducted a user evaluation to test and improve the chatbots [21,22]. Chen and his research team provide an overview of integrating a KG with a chatbot for prostate cancer (PCa) management and lifestyle interventions. The KG was implemented using Neo4J and was constructed from 300 scholarly articles, comprising 21,546 entities and 66,493 relationships that link lifestyle factors, medical outcomes, and related entities. The chatbot was developed with the Flask framework and supports 12 types of questions, assisting users in understanding PCa risks associated with specific lifestyle choices. While innovative, the system is limited by its reliance on a template-based approach and a relatively small dataset, highlighting areas for further development and refinement [23].

Finally, Bao also demonstrates the effective use of KGs in a medical chatbot by storing structured data to answer domain-specific queries with high precision. KGs support logical reasoning tasks by linking symptoms to diseases through RDF triples. To address limitations in handling casual language, the chatbot integrates KGs with a neural-based model, thereby enhancing conversational flexibility. This hybrid approach highlights the potential of combining KGs with NLP for scalable healthcare applications [24].

All the KG-based systems described above aim to deliver precise and reliable knowledge to users, support the development of chatbots capable of understanding natural language, and enable the seamless integration of new and potentially unlimited information. Although KG-based chatbots represent a promising and effective solution, they face significant challenges, particularly in the integration of NLU with knowledge extraction techniques.

2.2. Related Work on Collaborative Conversational AI

Distributed collaboration, involving humans, software, and computer agents is not a new concept. It has been proposed and Implemented to support complex systems such as GIS, internet infrastructure, gaming environments and crowd-sourcing platforms [25]. In conversational AI systems, distributed collaboration can be achieved by sharing knowledge across multiple chatbots, each managing a portion of a distributed KG or another knowledge source [26].

Chaves and Gerosa [27] concluded that multi-chatbot systems may confuse users; however, they can also provide richer interactions when supported by well-designed coordination mechanisms. Pinhanez et al. [28] found that single-bot systems are easier to use, whereas multi-bot systems offer deeper insights at the cost of increased complexity. Kantharaju and Pelachaud [29] developed a 3D multi-chatbot system that promotes healthy habits through dynamic turn-taking based on non-verbal cues and user engagement. These early studies highlight both the benefits and the design challenges of multi-chatbot collaboration. Following these observations, IBM's COBOTS attempted to address the routing problem, by introducing an orchestration bot that route user intent to specialized expert bots [30]. In parallel, established conversational AI platforms have also explored orchestration approaches. OpenDialog's 2021 manifesto promotes dynamic chatbot conversations involving multiple users and agents coordinated under a supervising chatbot [31]. Rasa also introduced multi-chatbot orchestration mechanisms to manage intent routing and context switching [32]. Kore.ai, in turn, offered a universal

 doi:10.20944/preprints202606.0203.v1

chatbot designed to direct queries across multiple bots [33], although its focus remained on distribution rather than deep collaboration.

From a user-centered perspective, a 2022 study conducted a comparative evaluation of single and multi-chatbot interfaces in m-commerce platforms. In the multi-chatbot condition, the system consisted of collaborative chatbots, each specializing in a different product category. The study examined how these different approaches influenced users' perceptions of chatbot competence, social presence, trust and purchase intention. The results showed that male participants were more favorable toward the single chatbot. When interacting with multi-chatbot interface, they reported satisfaction only when each chatbot was clearly labeled as an expert and demonstrated domain-specific expertise. Another notable finding was that both male and female participants expressed greater trust in male-presenting chatbots, reflecting prevailing societal perceptions that associate males with higher trustworthiness [34]. In the same year, A. Briel developed the JuggleChat framework [35], introducing a user interface designed to route user intents to three different chatbots focused on educational content related to Covid-19. The researcher used the Rasa framework for intent detection and routing to the appropriate chatbot. The system also incorporated an open source generative model for text summarization. The evaluation results indicated that users were largely indifferent to the underlying multi-chatbot architecture; instead, they were primarily concerned with the accuracy and usefulness of the responses provided.

In March 2022, a research team proposed a collaborative framework designed to unify leading conversational voice agents, including Siri and Alexa. These systems are referred as agents because they can perform specific tasks, such as placing orders in e-commerce platforms or playing music. Considering that these agents are closed-source proprietary systems, the researchers proposed not an intent routing mechanism but an answer-pairing procedure called MARS, in which the system selects the most appropriate response among multiple agent outputs. The method was tested by combining 19 commercial agents and achieved significantly higher accuracy than any individual agent [36]. In 2023, another study introduced CommunityBots, a multi-agent chatbot platform in which each chatbot specializes in a specific domain. The authors introduced a Conversation and Topic Management (CTM) mechanism to enable smooth topic and chatbot switching. An evaluation with 96 participants showed that CommunityBots led to higher engagement, improved response quality, and fewer conversational interruptions than a single-agent chatbot [37].

Distributed collaboration in chatbot systems remains an active area of research and presents numerous ongoing challenges. Key issues include intent routing, turn-taking, context switching, backend coordination, and maintaining user satisfaction—all of which complicate the design and development of multi-chatbot systems based on a distributed collaboration model. The studies discussed in this article propose various solutions to address these challenges. However, a critical question arises: are these solutions generalizable and adaptable across diverse chatbot ecosystems, or are they narrowly tailored to the specific problems targeted by each research effort?

3. Design and Implementation of GraphCoBots System

In this section, the design of GraphCoBots system is presented. In particular a brief overview of the system architecture, and the technologies and services that were utilized and developed are provided. Furthermore, the development of three chatbots and two KGs, the dialog design approach and the implementation of the collaboration mechanism among the three chatbots is explained. Finally, a brief overview of the web administration tool developed to collect users' interactions, is also provided.

3.1. GraphCoBots System Architecture

GraphCoBots System Architecture was designed to fulfill the distributed conversational collaboration concept among different chatbot services. Each architectural layer was implemented to support dialogue control, knowledge access, and coordination responsibilities of the system. Figure 1 illustrates the multi-chatbot architecture GraphCoBots, which is structured in six layers, that separate user interaction, conversational processing and knowledge access:

 doi:10.20944/preprints202606.0203.v1

- **Server Layer.** The system is deployed on a cloud-based infrastructure (Microsoft Azure) to support modular services and scalable deployment. Hosting the system components in a shared environment allows different chatbot services and knowledge bases operate autonomously while staying loosely connected.
- **Chatbot service layer.** This layer hosts three Rasa based chatbot instances each representing a domain-specific conversational agent. Each chatbot maintains its own dialogue models, training data, connected knowledge sources, and action services, reinforcing domain autonomy. This choice allows collaboration to occur through explicit coordination rather than shared internal state.
- **Knowledge base layer.** This layer consists of a PostgreSQL database for interaction logging and Neo4j KGs [38] that provide structured domain knowledge to chatbot instances. Each thematic chatbot is connected to each own KG, enforcing knowledge boundaries and preventing implicit data sharing.
- **Web sources layer.** This layer serves as an additional knowledge layer. The system can connect to external web-based knowledge sources through APIs such as Open-Meteo weather API [39] and Wikidata[40]. These sources extend the informational reach of individual agents while maintaining the principle that knowledge access remains mediated by the chatbot responsible for the interaction.
- **System admin layer.** This layer contains administrative tools that support the monitoring, analysis, and maintenance of chatbot interactions. The Azure internal monitoring service and a Streamlit-based Python Administration tool enable developers to continuously assess system performance without interrupting ongoing collaborative operations.
- **User Applications layer.** Users can interact with the chatbot via audio or text through multiple channels, including webpages and social media chat services. Additionally, the chatbot can be integrated into VR/AR and mobile apps through an API service. The separation of presentation and conversational logic allows multiple interaction channels to share the same distributed backend architecture, ensuring that collaboration mechanisms remain independent of the user interface.

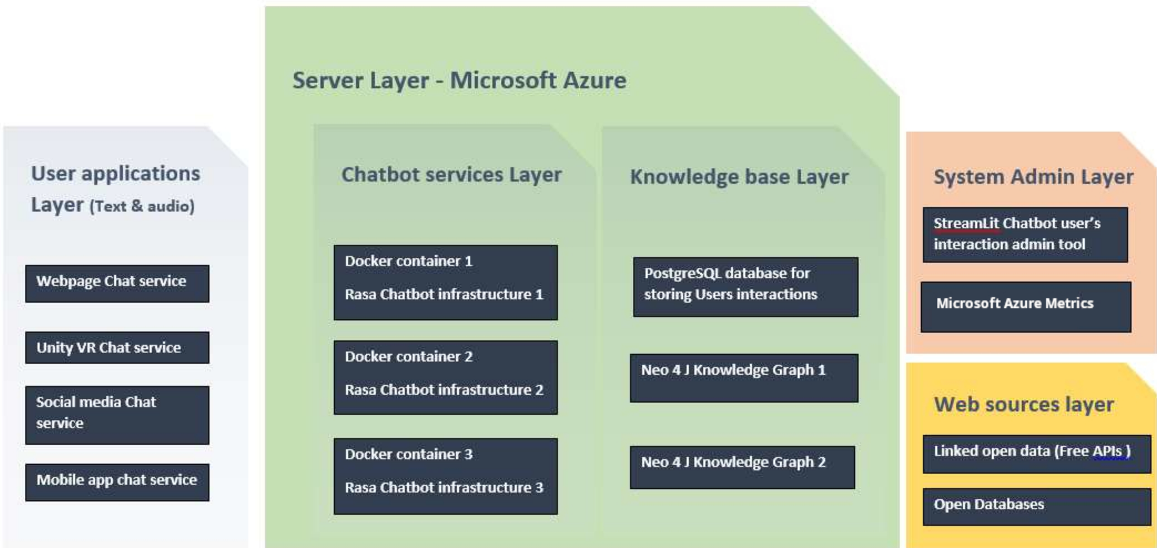

Figure 1. GraphCoBots - Multi-chatbot graph-based Architecture Layers

Figure 2 illustrates the interaction model between visitors (human agents) and a distributed and collaborative system of three specialized chatbots (AI agents). The detailed architecture was fully presented in our previous article [5]. During the implementation of the architecture with our chosen technologies, a clearer thematic specialization was introduced by assigning each chatbot to its own KG. This refinement enables a more explicit study of domain-based collaboration and routing behavior.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0203.v1

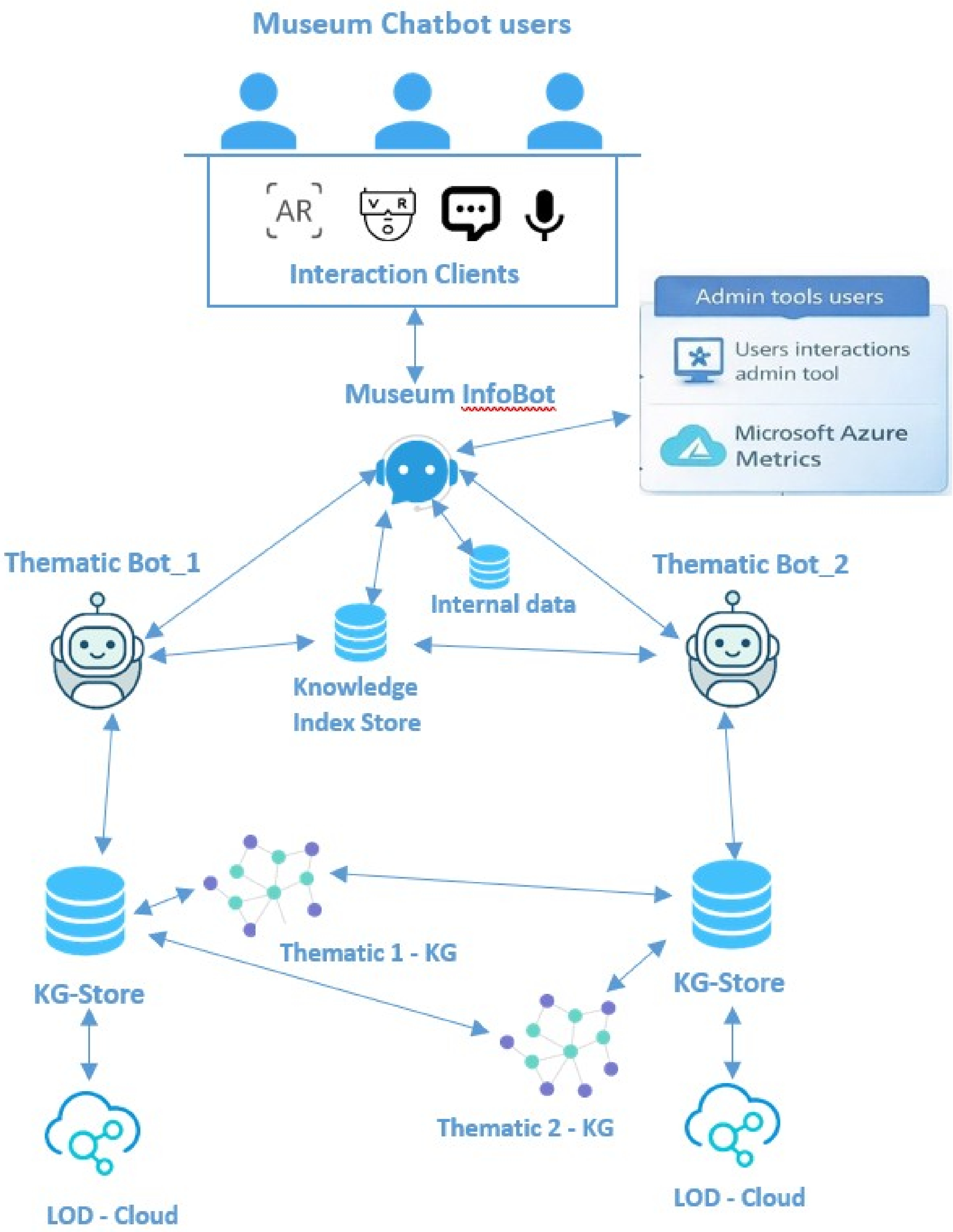

Figure 2. GraphCoBots - Multi-chatbot graph-based architecture and flow diagram.

User queries are processed through a structured routing pipeline that determines how a query should be handled. A user query is first interpreted at the intent level, which functions as a symbolic coordination mechanism. Based on this interpretation, the system routes the request to one of three sources of knowledge:

1. internal intent-based responses,
2. the domain-specific KG associated with the active chatbot, or
3. external linked data sources accessed through structured queries.

Because routing decisions are made before knowledge retrieval, conversational responsibility is assigned at the dialogue management level rather than emerging from shared data access. This design has an important research implication: KGs do not collaborate directly at the data level. Instead, collaboration occurs through conversational control and routing decisions. While tighter integration of multiple KGs could enable deeper cross-domain reasoning, the present architecture intentionally prioritizes modularity. This constraint makes inter-agent coordination explicit and observable, allowing the system to serve as a platform for studying distributed conversational collaboration.

3.2. Front-End and Back-End Implementation of GraphCoBots System

In the following paragraphs, the implementation of GraphCoBots system is presented. The creation of the three chatbots and their associated KGs is described, along with the process of integrating each chatbot with its corresponding KGs and LOD sources .

3.2.1. Creating Conversational Collaborative Chatbots with the Use of Rasa Open Source Framework

Our architecture includes three chatbots focused on Nikos Kazantzakis. The first is an infobot designed for the Museum of Nikos Kazantzakis. The second provides information about Kazantzakis's works and life, while the third chatbot focuses on selected artifacts presented in the museum. The latter two chatbots are KG-based chatbots and derive their knowledge from a KG. The infobot serves as the host of the distributed collaborative system. Chatbots were developed with the Rasa open source framework [11], a Python-based conversational framework that uses YAML files for configuration. Rasa can be connected to KG databases such as Neo4j or GraphDB , enabling the delivery of accurate, domain-specific information. Its Dual Intent and Entity Transformer (DIET) is a lightweight transformer-based model that provides accurate intent classification and strong generalization to unseen user utterances. When combined with specialized entity extractors, DIET supports robust and flexible natural language understanding across diverse type of user input. [41]. Furthermore, Rasa supports distributed collaboration architectures, allowing the development of multi-chatbot systems, under an orchestration agent. [32].

Museum of Nikos Kazantzakis (MNK) infobot

MNK infobot was designed and developed to provide users with comprehensive assistance regarding the museum. Users can inquire about the location, ticket prices, the museum's founder, educational programs, events, the presence of a café, pet policies, opening hours, and a variety of other useful information. These questions were formulated by exploring the museum's webpage and incorporating feedback from its curators. All knowledge was stored in the YAML files that were generated during the Rasa installation. The chatbot was also enhanced with two additional features. The first enhancement involves integrating word embeddings tailored for the Greek language [42] using the Greek model from SpaCy, an open-source library for NLP in Python. These embeddings help represent words as vectors in a high-dimensional space, where semantically similar words are located closer together. The second feature that was implemented involved the integration of an external LOD source to enrich the chatbots responses. Specifically, the "weather" intent was connected to the Open-Meteo weather API[39], which provides both real-time and forecasted weather data.

Nikos Kazantzakis Life and Work (NKLW) KGbot

The NKLW KGbot was developed similarly to the MNK Infobot and can answer questions about Kazantzakis' life, including his relatives, marriage, friends, studies, and travels, as well as general and specific inquiries regarding his works. Five intents, implemented with Python actions and Cypher queries, enable retrieval from the connected KG of relevant information, such as listing all novels or identifying a specific work by year. These intents are designed to be reusable, allowing users to specify different types of works or years. Additionally, a transference intent determines whether a question should be routed to other collaborative bots, supporting effective multi-bot coordination. The NKLW KGbot is further enhanced with access to structured data about Nikos Kazantzakis stored in Wikidata via two intents that use Python actions to send SPARQL queries to the Wikidata SPARQL

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0203.v1

Query Service [40] . Once the results are received, they are presented as responses to the user. The SPARQL queries are provided on the GitHub repository [1].

Nikos Kazantzakis Museum Artefacts (NKMA) KGbot

The NKMA KGbot was developed similarly and can answer general and specific questions about the artifacts and collections displayed across the museum's seven halls. Five intents implemented with Python actions and Cypher queries, enable the retrieval from the connected KG of relevant items - for example, listing all personal items in the museum or showing artifacts in a specific showcase and collection. These intents are reusable, allowing variations in user queries through changes in showcase or collection references. Furthermore, an additional intent uses an index to determine whether a question should be routed to one of the other two collaborative bots. During the design of the collaborative system, challenges emerged due to keyword overlaps with the NKLW KGbot index; these were resolved by reviewing and refining intent utterances and the selection process, ensuring accurate intent routing and seamless multi-bot collaboration.

3.2.2. KG Development for the GraphCoBots System

As described in our architecture, KGs are a primary data source for each chatbot service, enabling the retrieval and presentation of relevant knowledge to the users. KGs represent an emerging trend in AI, evolving from KBs and the Semantic Web (SW). KGs provide an effective and flexible means of representing interconnected descriptions of real-world entities and events, giving chatbots broad access to extensive knowledge stores [43,44]. A KG mainly describes real-world entities and their interrelations, organized in a large, directed graph. Both the schema and the corresponding data are represented in the graph, often using the widely recognized RDF model, in the form of triple statements. The integration of KGs to conversational AI (usually called hybrid AI) enhances dialogue quality that could eventually lead to more meaningful AI chatbot applications [43,45]. For the design and implementation of our KGs, we used Neo4j, a leading graph database system [38]. Neo4j was selected for its efficient management of highly connected data and its strong performance in traversal-intensive and query operations, as reported in comparative evaluations of native graph databases and RDF-based systems [46,47]. Neo4j is an ACID-compliant, transactional database with native graph storage and processing, allowing complex, interconnected data to be stored and queried efficiently using the Cypher query language.By using a graph structure instead of traditional relational tables, Neo4j enables efficient representation and analysis of complex relationships between entities [48].

In this work we present two KGs. The first captures the life and the works of the Greek writer Nikos Kazantzakis and related entities, while the second encompasses selected artifacts associated with him, displayed in Nikos Kazantzakis Museum at Myrtia, Heraklion, Crete. For the development of the two KGs, we followed the Collaborative and Hybrid Engineering of KGs (CHEKG) methodology introduced by Angelis et al., a hybrid schema-centric and data-driven approach that supports iterative, modular, and collaborative design of domain-specific KGs across the entire engineering lifecycle[49]. The KGs were developed initially in Protégé using OWL and aligned with established Semantic Web standards and vocabularies, including RDF/RDFS, FOAF, and Dublin Core, to ensure semantic consistency, interoperability, and reuse.

NK life and works KG

For the design of the Life and Works KG, family, professional, and friendship relationships were first mapped. Subsequently, all types of Kazantzakis' works were modeled and linked to these relationships. The Cypher query language was used to define entities and their interconnections. Figure 3 presents a graphical representation of the NK Life and Works KG. The data were developed in both Greek and English, with English terms labeled using the prefix "en." The complete Cypher code has been uploaded to a Neo4J database hosted on Microsoft Azure and is available in a GitHub repository [1]. To test the KGs integration with the chatbot infrastructure, three Cypher queries were created, which are available in the same GitHub repository. The last two queries parameterized, allowing the values to be modified, making them reusable for a wide range of user questions.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0203.v1

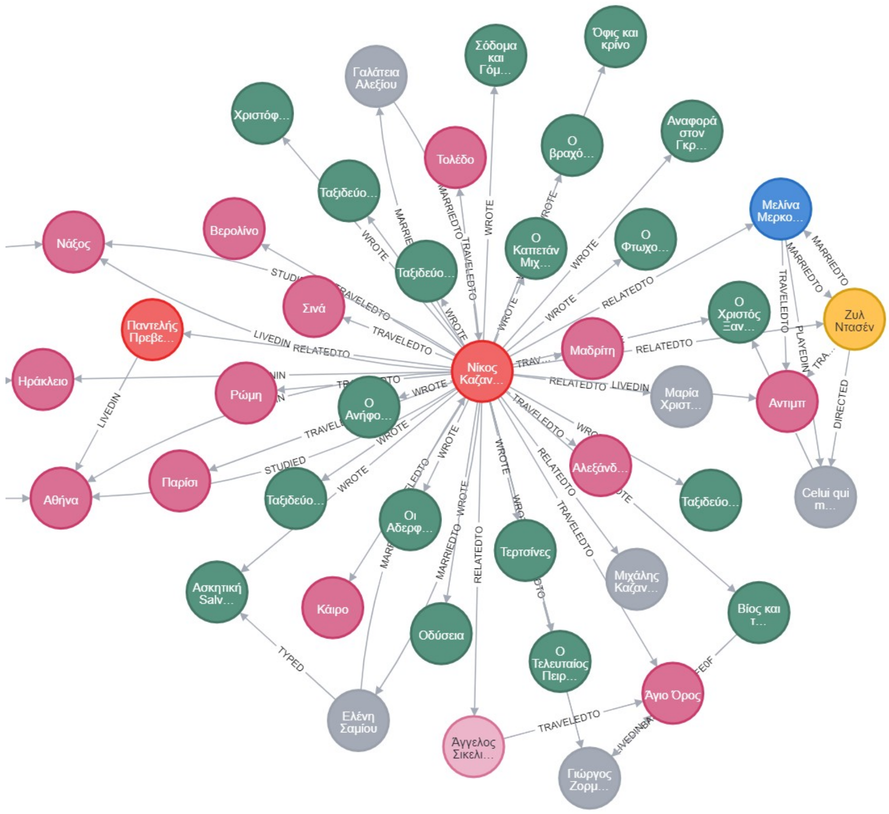

Figure 3. KG centered on Nikos Kazantzakis, illustrating relationships among places, persons, and works. Nodes represent entities and edges denote relations such as lived in, traveled to, and wrote. Key labels are provided in English (e.g., Sikelianos, Madrid, Rome, Athens, Zorba the Greek, Odysseia

Exhibits of Museum Nikos Kazantzakis KG

For the design of the KG related to the exhibits presented at the Museum of Nikos Kazantzakis, all seven halls of the Museum were modeled, and a total of 70 exhibits were selected with their relationships established. The data were developed in two languages: Greek and English, with English terms labeled with the prefix "en." The complete Cypher code has been uploaded to a Neo4J database hosted on Microsoft Azure and is available in a GitHub repository [1]. To test the KGs integration with the chatbot infrastructure, four Cypher queries were created and are available in the same GitHub repository. Queries 1 and 3 parameterized, allowing the values to be modified and making them reusable for a wide range of user questions.

3.3. Dialog Design with Agile Creation Process

The design and creation of dialogue for a chatbot is a crucial factor in achieving human-like interactions. In this work, dialogue development followed an agile and collaborative process involving content editors and Machine Learning (ML) engineers, enabling iterative refinement of intents, utterances and responses. A number of key aspects were considered during this process. First, user intent identification required a thorough understanding of both the domain content and anticipated user needs. Second, for each, a diverse set of utterances - typically up to 15- was created to support robust intent classification during training. Third, a response formulation focused on ensuring grammatical correctness, semantic alignment with all utterances, and consistency in informational content. During the training phase, ML engineers trained and evaluated the chatbot, introducing iterative adjustments as needed. This phase included the creation of additional utterances. identification of semantic overlaps between intents, recommendations for merging or splitting intents, and refinement of responses that were unclear or ineffective. The overall process was repeated iteratively - up to three times - until satisfactory performance was achieved. In addition to domain-specific intents, particular intention

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0203.v1

was given to the design of dialogue management intents such as greetings, welcome messages, compliments, out of scope responses, and user challenge handling. These elements contribute to a more human-like conversational experience. To further support user engagement, storytelling techniques and interactive elements (e.g., buttons. follow-up questions or knowledge paths) were incorporated. Together, these design practices enabled more fluid, coherent, and engaging interactions, aligning the chatbots behavior with user expectations in a cultural heritage context.

3.4. Developing the Collaborative Chatbot Service

As defined in the system architecture [5], a Knowledge Index Store was implemented and integrated as an intent within the configuration of each chatbot. This index contains curated phrases and keywords associated with the intents of the other collaborative chatbots, enabling cross-bot intent identification. All the intents that work as indexes are available in the source code of the three chatbots on GitHub repositories[1]. When a user query matches an entry in the Knowledge Index Store, the chatbot detects that the request falls outside its own knowledge scope and notifies the user that the question can be handled by another chatbot. The user is then prompted to confirm whether they wish to be transferred. Upon confirmation, a Python-based action terminates the current webchat session and initiates a new session with the appropriate collaborative chatbot, while preserving and transferring the identified user intent. As a result, the user receives a response from the collaborative chatbot at the start of the new session. Although effective, this collaboration mechanism requires further refinement at the webchat level, as users currently experience a brief delay during the transition between chatbot services.

The architecture also supports collaboration at the KG level, allowing a chatbot to access information from multiple KGs when required. In practice, however, knowledge access is governed by intent classification and routing through the Knowledge Index Store intent, which determines the responsible chatbot and, by extension, the KG to be queried. An alternative design-connecting each chatbot to all available KGs-was considered but ultimately rejected, as it would weaken explicit chatbot collaboration and increase the complexity of intent routing by distributing KG-related intents across multiple agents. Although KGs can be interconnected at the relationship level, knowledge access and routing are always performed through a chatbot selected at a higher orchestration level. The adopted approach preserves clear knowledge boundaries while enabling coordinated, modular collaboration among the chatbots.

3.5. GraphCoBots Administration Tool

The final component integrated into the proposed architecture is a multi-chatbot administration tool designed to support system monitoring, maintenance, and iterative improvement. The tool provides access to the YAML files of all three chatbots, enabling administrators to review and modify dialogue configurations as needed. In addition, it presents a consolidated view of user interactions, allowing administrators to monitor conversational behavior and identify potential errors or suboptimal chatbot responses. The administration tool offers usage statistics in both numerical and graphical formats, including metrics such as the number of active users and the most frequently triggered intents. These insights support data-driven refinements of the chatbot configurations, contributing to improved system performance and conversational quality. Furthermore,the tool enables administrators to assess system load and usage patterns, informing decisions related to scaling the underlying web hosting infrastructure, such as upgrading or downgrading service plans based on demand. The tool was developed using Streamlit, a Python-based web framework . The chatbots YAML configuration files are hosted on a Azure web service, while user interaction logs are stored in a PostgreSQL database. Access to the tool is provided via a web interface with authorized login [50].

 doi:10.20944/preprints202606.0203.v1

4. Evaluation of GraphCoBots System

In this section, we present an empirical evaluation of the GraphCoBots system, focusing on usability, interaction quality, and user expectations. The study combines an expert-based SUS assessment with a broader user study involving participants from diverse backgrounds. Together, these approaches provide both quantitative and qualitative insights into real-world interaction with the system.

4.1. Experts Evaluation

This experts evaluation involved 15 participants, aligning with established guidelines for usability testing and pilot study methodologies [51,52]. The aim was to identify usability issues and evaluate user experience rather than to produce statistically generalizable results. Participants were divided into three user categories: A) five developers specializing in ML Engineering, Data Engineering, and Full-stack Engineering, B) five content creators with expertise in chatbot and web application content, and C) five general users with bachelor-level education aged between 40 and 45. Participants were first given a brief introduction to the GraphCoBots system and its purpose, and then invited to freely interact with the system by asking their own questions. The evaluation was carried out over three days during the first days of August, with participants working in three teams and using the system simultaneously for 15 minutes before completing the questionnaire. After the interaction, they completed a two-part survey describing their overall experience. The first part consisted of the standard System Usability Scale (SUS) questionnaire [53], to assess perceived usability of, while the second part included five open questions about the system's strengths, weaknesses, and general impressions.

4.1.1. Experts Evaluation Results of the SUS Questionnaire

The SUS questionnaire yielded favorable results, with an overall score of 76.16/100, corresponding to grade B and placing GraphCoBots system in the 78 percentile rank [54]. Groups A and B, whose members had prior experience with chatbots, reported the same average score (73), whereas group C (general users) gave a higher average score of 82.5. Notably, the two lowest individual scores (57.5 and 55) were derived from Groups A and B, while the highest score (92.5) by a participant in Group C. In terms of the Network Promoter Score (NPS), general users (Group C) can be characterized as Promoters of GraphCoBots system, while developers (Group A) and content creators (Group B) are more likely to be passive and not promote GraphCoBots system [54].

Figures 4 and 5 illustrate the distribution of responses to individual SUS items. Figure 4 presents the odd-numbered (positively worded) items, which are expected to receive higher scores, while Figure 5 shows the even-numbered (negatively worded) items, which should receive lower scores. Participants, as it is presented on Figure 4, rated the GraphCoBots system as easy to use (mean=4.0) and easy to learn (mean=4.27), while indicated that it was neither overly complex (mean=1.87) nor cumbersome (mean=1.6). They also reported that they would not require any help to use the system in the future (mean=1.33) and would not need to learn many thing before using it (mean=1.6). However, responses to negatively worded items, presented on Figure 5, suggest that some functionalities were perceived as not fully well integrated (mean=3.4). Users reported moderate confidence while using GraphCoBots system (mean=3.8) and a moderate willingness to use it frequently (mean=3.47), while expressing some concern regarding system consistency (mean=2.2).

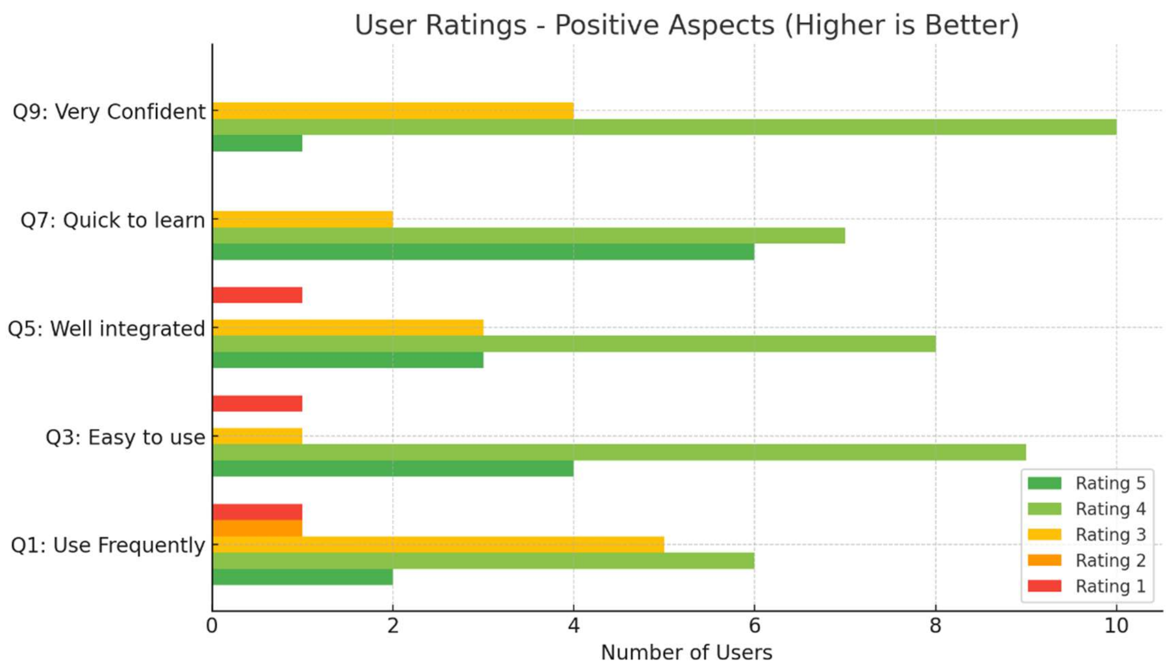

Figure 4. SUS positive aspects of experts evaluation

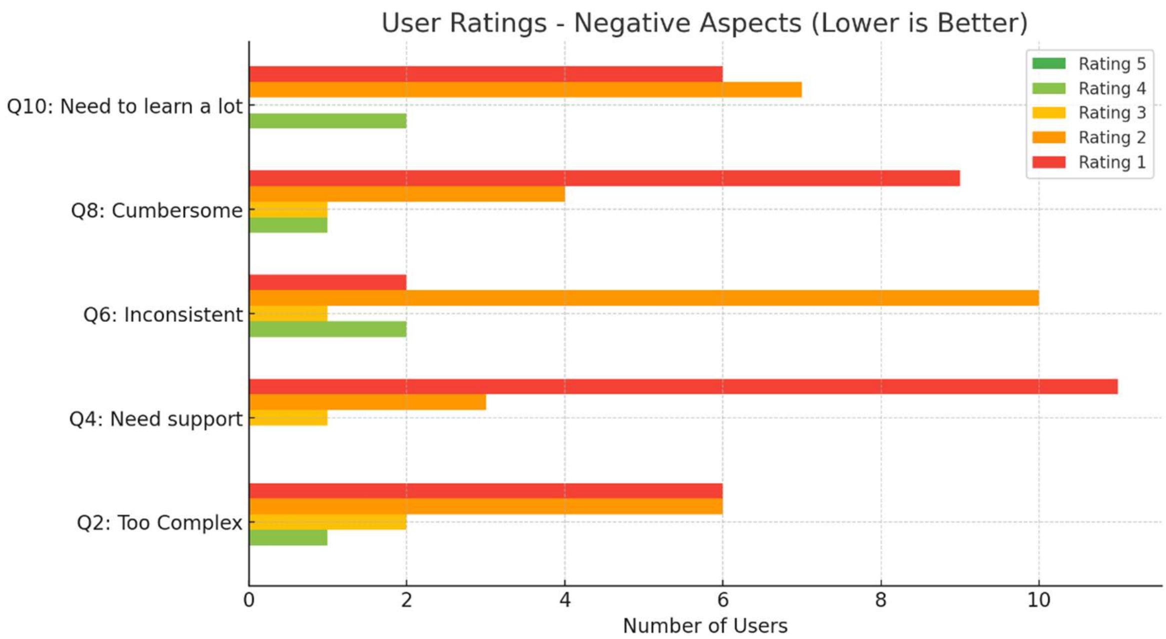

Figure 5. SUS negative aspects of experts evaluation

4.1.2. Qualitative Findings from Open-Ended User Feedback

We devote this section to summarize the answers to the open questions of the questionnaire.

Q1. What are the main strengths of GraphCoBots system?

Participants primarily highlighted the system's ease of use, intuitive interface, and fast response time. Many appreciated the rich, reliable content and the inclusion of visual material such as images and documents. A small number suggested further enhancing the experience through full voice interaction capabilities.

Q2. What are the main weaknesses of GraphCoBots system?

The most common concerns related to answer accuracy, limited support for complex or follow-up queries, and sensitivity to phrasing. Some users noted missing conversational context retention and inconsistencies across the collaborative chatbots. A few participants also reported technical issues or described the interaction style as insufficiently natural.

Q3. Can you think of any additional features to be included in GraphCoBots system?

Users proposed expanding content coverage, improving natural language understanding, and supporting multilingual interaction. Additional recommendations included voice features (STT/TTS), spatial and navigation support within and beyond the museum, QR-based access, richer visual material, and practical services such as ticketing and opening-hours information.

Q4. Can you think of any additional queries for GraphCoBots system?

Participants suggested deeper biographical and literary questions about Nikos Kazantzakis, as well as more detailed interpretations of his works. They also requested museum-related guidance (e.g., key exhibits, group visits) and post-visit recommendations, with some emphasizing the value of spatial awareness and the ability to leave user feedback.

Q5. What would you add to increase the accuracy/comprehensiveness of the information returned by GraphCoBots system?

Users recommended better entity disambiguation, integration of multiple data sources, and inclusion of excerpts and scholarly commentary. Other suggestions included clarification prompts when confidence is low, personalization based on user profiling, more visual references and external links, and a more conversational response style.

The statistical evaluation results, responses to the open-ended questions, and interaction logs collected through the GraphCoBots administration tool were analyzed to identify system improvements. The collected data of the administration tool did not consist of performance metrics, but mainly of user interactions, with particular focus on unanswered questions. Insights from this analysis informed refinements to the chatbot system, which are presented in the following chapter.

4.1.3. Evaluation Insights and Improvements

Expert evaluation of the GraphCoBots system provided the research team with valuable insights into its strengths and weaknesses. The two expert groups conducted a more rigorous evaluation often influenced by their specific content or technological expertise. In contrast, the general users primarily focused on the usability and the quality of content aspects, expressing overall satisfaction with the system. Additionally, it was noted that some users across all groups exhibited a bias influenced by Generative AI technologies [20,55]. Their perspective was influenced by the strong conversational capabilities commonly found in generative AI chatbot applications. They expressed dissatisfaction when the chatbot failed to fully understand their intent, required them to rephrase questions, was unable to handle follow-up queries or more complex questions, or did not provide answers to irrelevant inputs.

The research team conducted a final assessment of the results and implemented the following improvements: a) Users are no longer required to provide affirmation for the collaboration process. The chatbot collaboration mechanism operates seamlessly, and users generally are unaware of which chatbot is responding, though they may notice changes in the chatbot name, color, and icon, b) the content of the three chatbots and the two KGs was enriched with additional information, c) the speech to text feature was enabled, d) design and research efforts are ongoing to determine how to implement and integrate text-to-speech functionality, and e) support for the English language was incorporated, though it is not yet enabled.

The research team also recognized the need to integrate generative AI features into the chatbot infrastructure, while ensuring that knowledge reliability, data accuracy, and security remain uncompromised [20,56]. Potential implementations include: a) Employing generative AI to handle fallback and out of scope questions, b) using generative AI to automatically convert user intents into Cypher queries for the KG and facilitate the creation of appropriate responses, c) using Generative AI to dynamically populate the implemented KGs with new knowledge, d) leveraging generative AI to improve routing

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0203.v1

of user intents to the most suitable chatbot, and e) utilizing generative AI to produce varied responses derived from the predefined answers offered by the conversational AI, and f) leveraging generative AI, particularly Retrieval-Augmented Generation (RAG), as an additional knowledge source.

As part of the system's iterative development, the team deployed a generative AI service across the three collaborative chatbots to handle fallback and out-of-scope interactions. The second chatbot was further enhanced with a Retrieval-Augmented Generation (RAG) component, enabling access to specialized knowledge about the life and works of Nikos Kazantzakis. Following these enhancements and overall system refinements, the GraphCoBots system underwent its final evaluation through a large-scale user study.

4.2. Broader Audience-Users Study

The broader audience users study involved 46 participants and was conducted over seven days in late-January. Its aim was to further examine usability, user experience, and interaction quality in a larger and more diverse sample, without seeking statistically generalizable results. Participants were invited to remotely access the AI-powered digital assistant of the Nikos Kazantzakis Museum, which integrates the three collaborating chatbots. Before the interaction, participants received short usage guidelines and were asked to interact with the system by sequentially adopting three roles: (A) a visitor organizing a family trip to the museum, (B) a literature student exploring the author's life and work, and (C) a specialized visitor or researcher seeking detailed information about exhibits and thematic collections. The system automatically routed questions to the most appropriate chatbot and participants were encouraged to interact for approximately 15 minutes (about five minutes per role) using both simple and more advanced queries. After completing the interaction, the participants filled in an online evaluation questionnaire [2]. The survey included the standard 10-item SUS questionnaire to assess perceived usability, eight open-ended questions, three of them related to the research questions, and 12 additional questions focusing on overall user experience.

In the following paragraphs, we present a summary of the initial results of the user study. Specifically, we report findings from the SUS questionnaire, open-ended responses, and a set of basic chatbot usage metrics recorded through the system's administration tool.

4.2.1. User Study Results of the SUS Questionnaire

The SUS questionnaire yielded very positive results, with an overall score of 80.43/100, corresponding to grade A. This score indicates a high level of perceived usability and suggests that the participants considered the system easy to use, well designed, and suitable for regular use. Figures 6 and 7 illustrate the distribution of responses to the individual SUS items. As expected, the odd-numbered (positively worded) items received high scores, while the even-numbered (negatively worded) items received low scores, demonstrating consistent and reliable response patterns.

Participants rated the system on positively worded items (Figure 6) as easy to use (Q3 mean = 4.54) and reported a strong sense of confidence while interacting with it (Q9 mean = 3.98). They also agreed that the system's functions were well integrated (Q5 mean = 3.87) and expressed a clear willingness to use the system frequently (Q1 mean = 3.76). Learnability was positively evaluated and the users indicated that the system was easy to learn (Q7 mean = 4.48).

Responses to negatively worded items (Figure 7) further reinforce this interpretation. Participants strongly disagreed that the system was unnecessarily complex (Q2 mean = 1.67) or cumbersome to use (Q8 mean = 1.52). They also indicated that they would not require technical assistance to use the system (Q4 mean = 1.48) and that they would not need to learn many things before getting started (Q10 mean = 1.43). Perceived inconsistency was also low (Q6 mean = 2.35), suggesting that most users experienced stable and predictable system behavior.

[2] Link to the questionnaire: https://forms.gle/Sk5fM9Yu47WsEZfz5.

Overall, the SUS findings indicate that the system provides a high-quality user experience, combining strong learnability with effective, confidence-inspiring usability. These results suggest that the system is likely to be easily adopted and used efficiently by its intended audience.

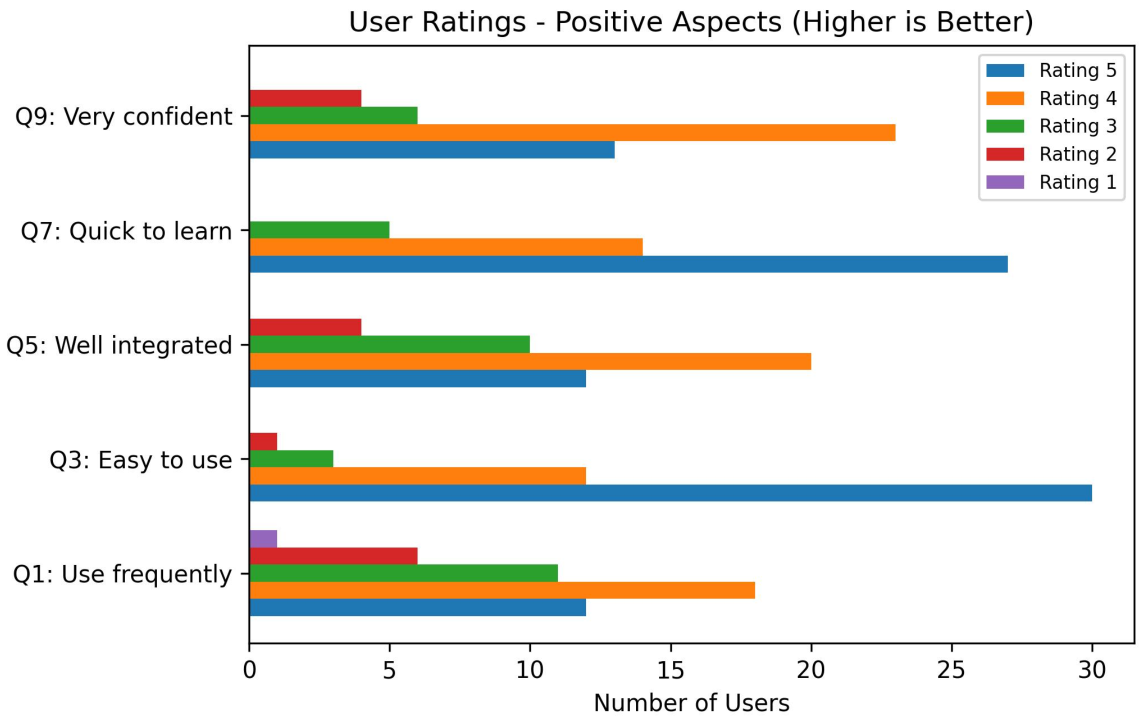

Figure 6. SUS positive aspects of broader user study

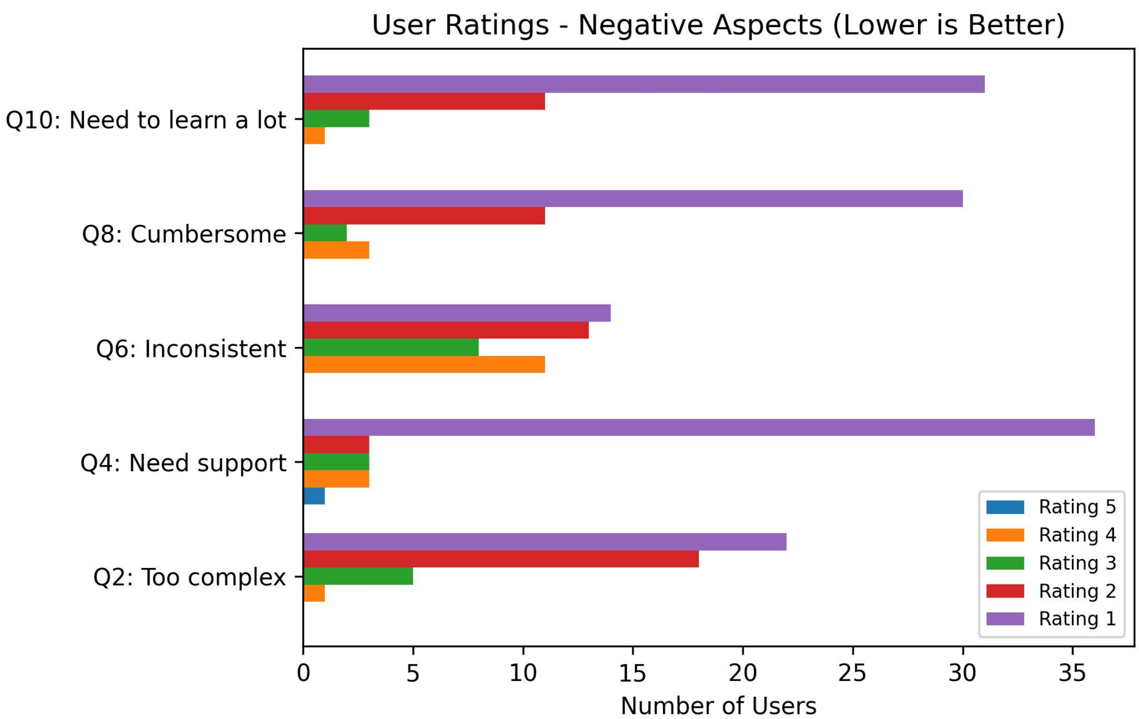

Figure 7. SUS negative aspects of broader user study

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0203.v1

4.2.2. Qualitative Findings from Open-Ended User Feedback

We devote this section to summarize the answers to the open questions of the questionnaire.

Q1. What are the main strengths of the GraphCoBots system?

Users highlighted the system's fast and immediate responses as a major strength, along with the accuracy and reliability of its information, particularly concerning the museum's archival content. The system was widely considered easy to use, featuring a clear and user-friendly interface. Participants also appreciated its interactive nature and the collaboration of multiple specialized chatbots, which supported well-organized and comprehensive information delivery.

Q2. What are the main weaknesses of the GraphCoBots system?

Participants reported limitations in chatbots ability to understand complex, open-ended, or loosely phrased questions, often resulting in unanswered or inaccurate responses. Some responses were perceived as overly generic, too long, or occasionally imprecise. Users also noted interaction issues, including the need to rephrase the questions, delayed responses, and loss of conversational context. Finally, feedback highlighted a lack of personalization, gaps in specialized information, and minor technical or interface-related usability issues.

Q3. Can you think of any additional features to be included in the GraphCoBots system?

Participants suggested several enhancements to expand the system's functionality and content coverage. Many recommendations focused on improving natural language understanding, enabling multilingual support, and allowing more complex or combined queries. Users also proposed adding voice interaction features, richer visual material, and integration with navigation or location-based services. Additional suggestions included practical museum-related services, such as ticketing, opening hours, and QR-based access to exhibits.

Q4. Can you think of any additional queries for the GraphCoBots system?

Participants proposed additional queries mainly related to deeper information about Nikos Kazantzakis, including his philosophy, literary influences, daily writing habits, and more detailed analysis of his works and characters. Many suggestions also concerned practical museum information, concerning special questions on ticket prices, visiting schedules, accessibility, transportation, and optimal visiting times. Users further requested more detailed questions about specific exhibits, family-oriented activities, and recommended visit routes based on available time. Finally, participants proposed broader and more exploratory questions, including modern interpretations of Kazantzakis' thought, educational programs, and nearby cultural or leisure options.

Q5. What would you add to increase the accuracy/comprehensiveness of the information returned by GraphCoBots system?

Participants emphasized the need for clearer source citation and transparency, suggesting references to specific works or pages to enhance credibility. Many also recommended improving natural language understanding, particularly for complex or imperfectly phrased questions. A frequent request concerned enriching and updating the system's knowledge base with more detailed and comprehensive content related to Kazantzakis and museum information. Finally, users proposed enhanced interactivity features, such as feedback mechanisms, clarification prompts, multimedia elements, multilingual support, and comparative or timeline-based content.

Q6. To what extent do you believe that KGs (i.e., intelligent structured databases), in combination with LOD sources, improve the coherence and continuous updating of the chatbots knowledge?

Most participants believed that KGs combined with LOD sources significantly improve the coherence and continuous updating of the chatbots knowledge. They emphasized that this integration enhances reliability, supports more accurate and up-to-date responses, and generally improves answer quality. One respondent noted that KGs could help transform the GraphCoBots system into an agent capable of understanding concepts rather than only keywords. At the same time, some users

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0203.v1

stressed the importance of validating information sources, while a few indicated that they did not feel sufficiently familiar with the topic to provide a confident assessment.

Q7. Do you think that Generative AI technologies can enhance the ease of dialogue and the overall interaction experience, while maintaining the accuracy and reliability of responses?

Most participants agreed that Generative AI technologies can enhance dialogue by making interactions more natural, fluid, and human-like, thereby improving the overall user experience. At the same time, they stressed that maintaining accuracy and reliability depends on proper safeguards, such as combining generative models with KGs and retrieval-based techniques to reduce the risk of incorrect or hallucinated responses. Users also noted that Generative AI increases the system's usefulness and encourages greater user engagement in the conversation.

Q8. Do you believe that collaboration among multiple specialized chatbots improves the user experience, and what limitations do you think may arise from this approach?

Most participants believed that collaboration among multiple chatbots can improve user experience by providing more accurate, targeted, and comprehensive responses across a wider range of topics. However, they also highlighted functional limitations, including increased response time, potential inconsistencies in tone or persona, and technical complexity in coordinating multiple systems. User-related concerns included potential confusion, particularly among less experienced users, when switching between assistants or when system roles lack transparency. Participants emphasized that the approach is most effective when transitions are seamless and responses remain coherent. A few respondents suggested that a single, unified chatbot might sometimes be preferable.

4.2.3. Basic Chatbot Usage Metrics

The operational metrics, gathered by the Administrator tool and the Microsoft Azure metrics component, further support the usability findings by demonstrating consistent and sustained interaction with the GraphCoBots system. During the six-day evaluation period, the system recorded 75 unique users, with an average of 12.5 users per day. In total 1253 requests were processed, indicating active engagement with the system. Performance indicators show that the average response time was 0,68 seconds with a minimum of 0,025 seconds and a maximum of 5,48 seconds, suggesting stable system behavior under real usage conditions. These results indicate that the system was able to support efficient concurrent interactions efficiently while maintaining a responsive user experience.

More specifically, analytics across the system's core modules reveal clear and meaningful usage patterns. The Conversational AI component handled 934 requests (approximately 155,6 per day), with an average response time of 0.39 seconds. The most frequent user intents related to practical and informational queries, including ticket prices, museum timetables, location details, book descriptions, works, writer's death, exhibition halls, collections information and museum artefacts. Collaboration features accounted for 273 requests, reflecting active interaction among the system's thematic agents. Cross-bot communication included 120 requests from the Museum Infobot to the other two thematic bots, 67 from LifeBot, and 86 from the Exhibit Bot, demonstrating coordinated knowledge exchange within the multi-agent architecture.

KG services supported 65 requests, both maintaining an average response time of 0,62 seconds, while interactions with LOD resources totaled 23 requests, with a slightly higher average response time of 0,97 seconds. Finally, the Generative AI module processed 230 requests, with response times ranging from 0,92 to 5,48 seconds (average 1,73 seconds), reflecting the higher computational demands of generative processing. Overall, these metrics indicate that the system was not only perceived as usable but was also actively and effectively utilized across multiple functional layers, confirming its practical value in a real-world cultural heritage environment. As a next step in the analysis of the evaluation results, the study will focus on examining the success rates of system responses. This stage moves beyond usage and performance indicators to assess response accuracy and task completion effectiveness. In particular, a detailed component-level analysis will be carried out, evaluating each

 doi:10.20944/preprints202606.0203.v1

chatbot individually in order to identify strengths, limitations, and performance differences among the system's specialized agents.

4.2.4. User Study General Conclusions

The evaluation of the GraphCoBots system provided the research team with valuable insights into both its strengths and its current limitations. Results from the broader user study showed that general users primarily emphasized usability, clarity of interaction, and the quality of the information provided, reporting a high level of overall satisfaction.

At the same time, more experienced or academically oriented participants tended to evaluate the system more critically, focusing on response precision, conversational depth, and the system's ability to handle complex or loosely structured queries. Across user groups, expectations appeared to be influenced by familiarity with modern generative AI systems, which often set a high benchmark for conversational fluency. As a result, some users expressed dissatisfaction when the system required question reformulation, struggled with follow-up or multi-part queries, or failed to interpret vague or exploratory inputs.

These findings highlight the growing gap between user expectations shaped by large-scale generative models and the constraints of structured, knowledge-driven conversational systems in specialized cultural heritage domains.

5. Discussion

During the recent surge of interest in generative AI, the research team critically examined the rationale for continuing to implement the KG-based distributed and collaborative multi-chatbot GraphCoBots system using solely established ML and NLP techniques. The introduction of Generative Pre-Training Transformer (GPT) models [55] and related chatbot services significantly raised user expectations, particularly regarding the effective use of contextual information to accurately interpret queries and generate coherent, open-domain responses in real time [20].

Recent research acknowledges that KGs and task or domain oriented chatbots, now often referred to as Agents or Agentic AI chatbots [57] are increasingly embraced and integrated within generative AI solutions. Our proposed architecture proactively addresses the need of reliable, continuously updated knowledge by incorporating KGs and other open data sources. It also emphasizes the importance of specialized chatbots (agents) capable of providing specific and performing certain tasks. Finally, it supports collaboration among these agents to enhance overall effectiveness.

Reflecting on our previous work [5] on the design and implementation of KG-based distributed and collaborative multi-chatbot system, we revisit the core motivations and the fundamental requirements that guided its development. The system was designed for use by small/medium museum and cultural organizations, which typically operate with limited budgets. Consequently, it needed to be: a)secure, ensuring that artifacts and museum knowledge - some of which remain unpublished or under study- are protected from misuse or unauthorized commercial exploitation, b) reliable in the accuracy and quality of the knowledge it provides, c) cost – effective and affordable, as an intended objective, with design choices expected to help reduce costs for small and medium-sized museums, d) user-friendly for all users, while allowing for continuous monitoring and updating and e) capable of delivering a natural, engaging conversational experience that feels more like interacting with a person than a machine. These requirements are supported at the architectural level through the use of open-source components, modular deployment, controlled access to KGs content, and the use of curated data sources. The GraphCoBots system is currently addressing these requirements, with ongoing enhancements being pursued by the research team. Generative AI techniques are integrated into the process due to their potential to address the majority of the identified requirements.

The research questions introduced at the beginning of this article are now addressed:

1. Do KGs contribute to more robust, reusable, and continuously up-datable knowledge management in a distributed multi-chatbot system for museums?

 doi:10.20944/preprints202606.0203.v1

The answer is affirmative as the KGs could continuously be enriched with robust and up-to-date information. Furthermore, recent literature discussed in this article, emphasizes that the integration of KGs in conversational AI systems is highly beneficial, with structured knowledge being favored by all stakeholders, including developers, museum curators, content creators, and end users. User feedback further also supports this conclusion, while several users also noted that structured knowledge could possibly help the system understand concepts rather than simply matching keywords. Moreover, The domain-independent architecture allows KGs to be adapted to different museums and extended with additional graphs and chatbots, supporting scalability to larger and more diverse collections.

2. Can collaborative multi-chatbot architecture, concerning museums domain, be effectively supported through knowledge obtained from external Web resources?
   The answer is also affirmative as the chatbots were successfully connected, through Python-based actions and queries, to open data sources such as wikidata, Open Meteo API, and Neo4J KG databases. Additional integrations with structured JSON/XML files and XLS files, were also explored, demonstrating the system's ability to incorporate heterogeneous external knowledge sources. User responses indicate that this integration contributed to richer and more comprehensive answers. At the same time, some participants highlighted the need for careful verification of external data to ensure reliability and consistency across responses. It should be noted that LOD sources and KGs operate as independent knowledge providers, with access handled through intent-based routing and source reliability. Schema-level integration and cross-source reasoning are beyond the scope of the current architecture.
3. Does the collaborative behavior of multiple chatbots enhance user experience in museums, and what limitations or drawbacks emerge?
   The evaluation demonstrated that users are primarily focused on receiving accurate and useful answers, with limited interest in the underlying collaboration mechanisms. For this reason, the research team worked toward making chatbot collaboration smoother and less intrusive, while continuing to refine the interaction design to enhance the overall user experience.
4. Can Generative AI be safely and effectively integrated into a KG–driven multi-chatbot architecture while preserving curatorial control and factual reliability over museum artifacts and cultural data?
   Following the initial hype in generative AI, researchers and organizations have recognized that current solutions often require substantial infrastructure, may raise data security concerns, can produce hallucinations or inaccurate information and tend to be expensive. As a result, the current focus has shifted toward developing more secure, cost-effective and reliable generative AI tools and solutions for organizations of any domain [20,55]. User responses reflected a similar perspective: participants noted that generative AI can enhance conversational naturalness and engagement, but emphasized the need for safeguards to ensure accuracy and reliability. Similar findings have been reported in recent studies showing that LLM-based chatbots can enhance visitor engagement and overall experience in cultural exhibition contexts [58]. Several suggested combining generative models with KGs and retrieval-based techniques to mitigate hallucinations and errors. In this context, the research team is progressively incorporating generative AI services into the proposed system in a controlled manner, aiming to enhance conversational naturalness while maintaining factual grounding.

It should be noted that the present evaluation was exploratory in nature and focused primarily on usability and user perception within a specific museum context. The study did not include controlled comparisons with alternative architectures or large-scale deployment scenarios. These limitations will be addressed in future work through broader evaluations and comparative studies.

 doi:10.20944/preprints202606.0203.v1

6. Conclusion and Future Work

The evaluation of the GraphCoBots system suggests that the integration of KGs, LOD sources, conversational components, and generative AI features can support a usable and effective multi-chatbot solution for cultural heritage environments. Both quantitative results and qualitative feedback show that users were able to retrieve meaningful information efficiently across diverse scenarios, including museum visits, biographical exploration, and exhibit-related inquiries. The multi-chatbot architecture, grounded in structured knowledge sources, delivered coherent and contextually relevant responses under realistic usage conditions. At the same time, the findings highlight increasing user expectations influenced by recent advances in generative AI, particularly regarding conversational flexibility and the handling of complex or ambiguous queries. Future work will address both interaction quality and system capabilities. Planned improvements include enhanced natural language understanding for loosely phrased queries, stronger conversational context management, more adaptive personalization, and smoother transitions between collaborating chatbots. The exploration of multi-modal interaction, such as voice and visual interfaces, is also foreseen.

From a technical perspective, ongoing enrichment of KGs and LOD sources will expand content coverage and improve factual precision, while further optimization of multi-chatbot coordination and routing will enhance response consistency and efficiency. A key research direction involves the controlled integration of generative AI within a hybrid architecture, where Large Language Models[59] and Retrieval-Augmented Generation techniques[60] support intent interpretation, fallback handling, and response generation while remaining grounded in curated and verifiable knowledge. Particular attention will be given to known risks such as hallucinations and overconfident errors through monitoring, validation mechanisms, and human oversight.

Future research will also include a more comprehensive quantitative assessment of the architecture, including controlled comparison with a single-chatbot baseline and the analysis of additional performance metrics such as task success rate, response latency, handoff effectiveness, and system behavior under concurrent use. In parallel, long-term deployment aspects for small and medium-sized cultural institutions will be investigated, focusing on content management workflows, operational sustainability, security, and cost efficiency. In this context, the potential use of locally deployable Small Language Models will be explored as a means to reduce operational costs while maintaining acceptable levels of accuracy and robustness [61]. The proposed approach establishes a foundation for future research on hybrid, knowledge-grounded conversational systems for sustainable and trustworthy cultural heritage applications.

Author Contributions: Conceptualization, S.V. and K.K.; methodology, S.V.; software, A.S., G.T. and I.G.; validation, S.V. and A.S.; investigation, S.V., G.T. and I.G.; resources, A.S.; data curation, S.V.; writing—original draft preparation, S.V.; writing—review and editing, K.K., G.T. and I.G.; visualization, S.V. and A.S.; supervision, K.K.; . All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: According to national policy Decision No. 23296, Government Gazette Issue B' 3469/17.06.2024, paragraph: Amendment of the Code of Ethics and Good Practice of the University of the Aegean, Paragraph 4.2.3. Obligations Regarding Specialized Areas of Research, such usability studies that involve anonymous feedback on non-sensitive topics and pose no risk to participants do not require formal ethics board approval.

Informed Consent Statement: Informed consent was obtained from all participants involved in the experts evaluation and the broader user study. Participation was voluntary, and all responses were collected anonymously.

Funding: This research has been partially implemented and co-financed by the European Union and Greek national funds through the Program COMPETITIVENESS under the call RESEARCH – INNOVATE (Project title: Destination AI Action bot platform - Project code: EKPAR03-0011707)

Data Availability Statement: The data concerning the expert evaluation and the broader user study and which are supporting the conclusions of this article will be made available by the authors on request. Restrictions apply to the availability of the cultural data concerning the two KGs of the GraphCoBots system. These data were

 doi:10.20944/preprints202606.0203.v1

obtained from the Nikos Kazantzakis Museum and are available from the authors upon reasonable request, with the permission of the museum.

Acknowledgments: The authors would like to thank the curators of the Museum of Nikos Kazantzakis for their valuable insights and domain expertise. We also acknowledge Evangelia Metoxianaki (Museologist), Varvara Tsaka, and Eleni Siakati (Aegean Solutions S.A.), content experts, for their contributions to dialogue design and knowledge documentation. We thank Alexandros Tzortzakakis (Aegean Solutions S.A.) for his overall support in providing the necessary resources. We further acknowledge Apostolos Patsidiotis, Machine Learning Engineer, for his contribution to the initial implementation of the multi-chatbot system and the Python-based administration tool, and Dimitrios Augoustopoulos, Front-end Developer, for the development of the web chat application. We also thank Adamantia Zimou, Graphic Designer, for the UX/UI design of the web chat interface. Finally, we would like to thank Stamatia Ladikou, Museologist and PhD Candidate at the Department of Cultural Technology and Communication, University of the Aegean, for her research support and contributions to the project.

Conflicts of Interest: Author Gkika Ioanna, Tsakiris Georgios and Skamagkis Aristotelis were employed by the company Aegean Solutions S.A. All three authors worked and collaborated on the article as part of their research activities during the implementation of the following research program: "MEMOBOT project – Conversations with Nikos Kazantzakis: Intelligent Mixed Reality Experiences – implemented within the framework of the Action "RESEARCH – CREATE – INNOVATE", 2nd Cycle, and is co-financed by the European Regional Development Fund (ERDF) of the European Union and national resources through the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (EPAnEK), NSRF 2014–2020 (Project Code: T2EDK-03334). This research effort is additionally supported through the framework of the following research program: Destination AI Action bot platform", implemented within the framework of the Action "RESEARCH – INNOVATE", co-financed European Union and national resources through the Operational Programme "Competitiveness" (Project Code: EKPAR03-0011707). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

KG	Knowledge Graph
AI	Artificial Intelligence
SUS	System Usability Scale
NLP	Natural Language Processing
NLU	Natural Language Understanding
ML	Machine Learning
LOD	Linked Opend Data
YAML	YAML Ain't Markup Language
DIET	Dual Intent and Entity Transformer
MNK	Museum of Nikos Kazantzakis
NKLW	Nikos Kazantzakis Life and Work
NKMA	Nikos Kazantzakis Museum Artefacts
SW	Semantic Web
NPS	Network Promoter Score
GPT	Generative Pre-Training Transformer

References

1. Kiourexidou, M.; Stamou, S. Interactive Heritage: The Role of Artificial Intelligence in Digital Museums. *Electronics* **2025**, *14*, 1884. https://doi.org/10.3390/electronics14091884.
2. Hao, X.; Xu, J.; Wang, Y. How Generative AI Shapes User Perceived Value and Adoption Intention in Digital Museum Experiences. *npj Heritage Science* **2025**, *13*, 608. https://doi.org/10.1038/s40494-025-02194-9.
3. Sánchez-Martín, J.M.; Guillén-Peñafiel, R.; Hernández-Carretero, A.M. Artificial Intelligence in Heritage Tourism: Innovation, Accessibility, and Sustainability in the Digital Age. *Heritage* **2025**, *8*, 428. https://doi.org/10.3390/heritage8100428.

4. Brown, K. Artificial Exchange: Chatbots and the Ethics of Museum Experience Design. In *Transformative Museum Experiences*; Calvi, L.; Vermeeren, A.; Sabiescu, A., Eds.; Springer Series on Cultural Computing, Springer: Cham, 2026. https://doi.org/10.1007/978-3-031-89521-0_13.
5. Varitimiadis, S.; Kotis, K.; Pittou, D.; Konstantakis, G. Graph-Based Conversational AI: Towards a Distributed and Collaborative Multi-Chatbot Approach for Museums. *Applied Sciences* **2021**, *11*.
6. Williamson, S.M.; Prybutok, V. The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation. *Information* **2024**, *15*, 299. https://doi.org/10.3390/info15060299.
7. Nikos Kazantzakis Museum. Nikos Kazantzakis Museum. Available online: https://www.kazantzaki.gr/, 2025. Accessed: 2026-02-07.
8. Nikos Kazantzakis Museum. Digital Nikos Kazantzakis. Available online: https://digital.kazantzaki.gr/, 2025. Accessed: 2026-02-07.
9. Hallili, A. Toward an Ontology-Based Chatbot Endowed with Natural Language Processing and Generation. In Proceedings of the 26th European Summer School in Logic, Language & Information, Tübingen, Germany, 2014. August 11–22.
10. Al-Zubaide, H.; Issa, A. OntBot: Ontology based chatbot. In Proceedings of the International Symposium on Innovations in Information and Communications Technology, Amman, Jordan, 2011; pp. 7–12.
11. Musen, M. The protégé project. *AI Matters* **2015**, *1*, 4–12.
12. Boné, J.; Ferreira, J.; Ribeiro, R.; Cadete, G. DisBot: A Portuguese Disaster Support Dynamic Knowledge Chatbot. *Applied Sciences* **2020**, *10*, 9082.
13. Yoo, S.; Jeong, O. Auto-Growing knowledge graph-based intelligent chatbot using BERT. *ICIC Express Letters* **2020**, *14*, 67–73.
14. Vegesna, A.; Jain, P.; Porwal, D. Ontology based Chatbot (For E-Commerce Website). *International Journal of Computer Applications* **2018**, *179*, 51–55.
15. Quamar, A.; Lei, C.; Miller, D.; Ozcan, F.; Kreulen, J.; Moore, R.; Efthymiou, V. An Ontology-Based Conversation System for Knowledge Bases. In Proceedings of the Proceedings of the ACM SIGMOD International Conference on Management of Data, Portland, OR, USA, 2020; pp. 361–376.
16. Ait-Mlouk, A.; Jiang, L. KBot: A Knowledge Graph Based ChatBot for Natural Language Understanding Over Linked Data. *IEEE Access* **2020**, *8*, 149220–149230.
17. Vasilevich, A.; Wetzel, M.; Sedlbauer, G.; Hubmer, K. Language-Agnostic Knowledge Graphs for Smarter Multilingual Chatbots. In Proceedings of the SEMANTiCS 2022 EU: 18th International Conference on Semantic Systems, Vienna, Austria, 2022.
18. Boroghina, G.; Corlatescu, D.; Dascalu, M. Multi-Microworld Conversational Agent with RDF Knowledge Graph Integration. *Information* **2022**, *13*, 539. https://doi.org/10.3390/info13110539.
19. Meloni, A.; Angioni, S.; Salatino, A.; Osborne, F.; Recupero, D.; Motta, E. Integrating Conversational Agents and Knowledge Graphs within the Scholarly Domain. *IEEE Access* **2023**, *11*, 22468–22489. https://doi.org/10.1109/ACCESS.2023.3253388.
20. Meloni, A.; Angioni, S.; Salatino, A.; Osborne, F.; Birukou, A.; Recupero, D.; Motta, E. AIDA-Bot 2.0: Enhancing Conversational Agents with Knowledge Graphs for Analysing the Research Landscape. In Proceedings of the The Semantic Web – ISWC 2023: 22nd International Semantic Web Conference, Athens, Greece, 2023; pp. 400–418. https://doi.org/10.1007/978-3-031-47243-5_22.
21. Patsoulis, G.; Promikyridis, R.; Tambouris, E. Integration of Chatbots with Knowledge Graphs in eGovernment: The Case of Getting a Passport. In Proceedings of the 25th Pan-Hellenic Conference on Informatics (PCI 2021), Association for Computing Machinery, New York, NY, USA, 2021; pp. 425–429. https://doi.org/10.1145/3503823.3503901.
22. Bartza, E.; Promikyridis, R.; Tambouris, E. On the Use of Chatbots and Knowledge Graphs for Public Service Information Provision Based on Life Events: The Case of Travelling Abroad. In Proceedings of the Proceedings of EGOV-CeDEM-ePart 2023, Corvinus University of Budapest, Hungary, September 2023. Conference dates: 05–07 September 2023.
23. Chen, Y.; Sinha, B.; Ye, F.; Shen, B.; et al.. Prostate Cancer Management with Lifestyle Intervention: From Knowledge Graph to Chatbot. *Clinical and Translational Discovery* **2022**, *2*, e29. https://doi.org/10.1002/ctd2.29.
24. Bao, Q.; Ni, L.; Liu, J. HHH: An Online Medical Chatbot System Based on Knowledge Graph and Hierarchical Bi-Directional Attention. In Proceedings of the Proceedings of the Australasian Computer Science Week Multiconference (ACSW 2020), Melbourne, VIC, Australia, 2020. February 4–6.

25. Martin, C.; Schreckenghost, D.; Bonasso, P.; Kortenkamp, D.; Milam, T.; Thronesbery, C. An Environment for Distributed Collaboration Among Humans and Software Agents. In Proceedings of the Proceedings of the AAMAS'03—Second International Joint Conference on Autonomous Agents and Multiagent Systems, Melbourne, Australia, 2003; pp. 1062–1063.
26. Bosse, S. Distributed Serverless Chat Bot Networks Using Mobile Agents: A Distributed Database Model for Social Networking and Data Analytics. In Proceedings of the Proceedings of the ICAART 2021—13th International Conference on Agents and Artificial Intelligence, Virtual, 2021; pp. 398–405.
27. Chaves, A.; Gerosa, M. Single or Multiple Conversational Agents? An Interactional Coherence Comparison. In Proceedings of the Proceedings of the Conference on Human Factors in Computing Systems (CHI 2018), Montreal, Canada, 2018; Vol. 191, pp. 1–13.
28. Pinhanez, C.; Candello, H.; Pichiliani, M.; Vasconcelos, M.; Guerra, M.; de Bayser, M.; Cavalin, P. Different but Equal: Comparing User Collaboration with Digital Personal Assistants vs. Teams of Expert Agents. arXiv preprint arXiv:1808.08157, 2018.
29. Kantharaju, R.; Pelachaud, C. Towards Developing a Model to Handle Multiparty Conversations for Healthcare Agents. In Proceedings of the Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications, Stockholm, Sweden, 2018; pp. 30–34. Available at http://ceur-ws.org/Vol-2338/paper4.pdf.
30. Subramaniam, S.; Aggarwal, P.; Dasgupta, G.; Paradkar, A. COBOTS—A Cognitive Multi-Bot Conversational Framework for Technical Support. In Proceedings of the Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS), Stockholm, Sweden, 2018; pp. 597–604.
31. OpenDialog. Manifesto. Available online: https://opendialog.ai/manifesto/, 2025. Accessed on April 1, 2025.
32. Rasa. Rasa AI Orchestration. Available online:https://rasa.com/orchestration, 2026. Accessed: 2026-02-07.
33. Kore.ai. Universal Bots. Available online: https://developer.kore.ai/docs/bots/advanced-topics/universal-bot/universal-bots/, 2025. Accessed on April 1, 2025.
34. Tan, S.M.; Liew, T. Multi-Chatbot or Single-Chatbot? The Effects of M-Commerce Chatbot Interface on Source Credibility, Social Presence, Trust, and Purchase Intention. *Human Behavior and Emerging Technologies* **2022**, p. 14 pages. https://doi.org/10.1155/2022/2501538.
35. Briel. Toward an eclectic and malleable multiagent educational assistant. *Computer Applications in Engineering Education* **2022**, *30*, 163–173.
36. Clarke, C.; Peper, J.; Krishnamurthy, K.; Talamonti, W.; Leach, K.; Lasecki, W.; Kang, Y.; Tang, L.; Mars, J. One Agent to Rule Them All: Towards Multi-Agent Conversational AI. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, 2022; pp. 3258–3267.
37. Jiang, Z.; Rashik, M.; Panchal, K.; Jasim, M.; Sarvghad, A.; Riahi, P.; DeWitt, E.; Thurber, F.; Mahyar, N. CommunityBots: Creating and Evaluating A Multi-Agent Chatbot Platform for Public Input Elicitation. *Proceedings of the ACM on Human-Computer Interaction* **2023**, *7*, 1–32. https://doi.org/10.1145/3579469.
38. Neo4j. A Highly Scalable Native Graph Database, 2025. Available online: https://neo4j.com/, Accessed: April 1, 2025.
39. Open Meteo. Free Open-source Weather API, 2025. Available online: https://open-meteo.com/, Accessed: April 1, 2025.
40. Wikidata Contributors. Wikidata's SPARQL Query Service Endpoint. https://query.wikidata.org/, 2025. Accessed: 2025-04-01.
41. Arevalillo-Herráez, M.; Arnau-González, P.; Ramzan, N. On Adapting the DIET Architecture and the Rasa Conversational Toolkit for the Sentiment Analysis Task. *IEEE Access* **2022**, *10*, 107477–107487. https://doi.org/10.1109/ACCESS.2022.3213061.
42. Outsios, S.; Karatsalos, C.; Skianis, K.; Vazirgiannis, M. Evaluation of Greek Word Embeddings. In Proceedings of the Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020), Marseille, France, 2020; pp. 2543–2551.
43. Bonatti, P.; Decker, S.; Polleres, A.; Presutti, V. Knowledge graphs: New directions for knowledge representation on the semantic web. *Report Dagstuhl Seminar* **2019**, *8*, 29–111.
44. Kotis, K.; Zachila, K.; Paparidis, E. Machine Learning Meets the Semantic Web. *Artificial Intelligence Advances* **2021**, *3*, 1.
45. Yan, J.; Wang, C.; Cheng, W.; Gao, M.; Zhou, A. A retrospective of knowledge graphs. *Journal of Asian Architecture and Building Engineering* **2006**, *5*, 67–74.

46. Monteiro, J.; Sá, F.; Bernardino, J. Experimental Evaluation of Graph Databases: JanusGraph, Nebula Graph, Neo4j, and TigerGraph. *Applied Sciences* **2023**, *13*, 5770. https://doi.org/10.3390/app13095770.
47. Pujante-Otalora, L.; Campos, M.; Juarez, J.M.; Vidal, M.E. Comparison Between Graph Databases and RDF Engines for Modelling Epidemiological Investigation of Nosocomial Infections. In Proceedings of the Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 2: HEALTHINF. SCITEPRESS, 2024, pp. 23–36. https://doi.org/10.5220/0012319900003657.
48. Robinson, I.; Webber, J.; Eifrem, E. *Graph Databases*, 2nd ed.; O'Reilly Media: Sebastopol, CA, 2015.
49. Angelis, S.; Moraitou, E.; Caridakis, G.; Kotis, K. CHEKG: a collaborative and hybrid methodology for engineering modular and fair domain-specific knowledge graphs. *Knowledge and Information Systems* **2024**, *66*, 4899–4925. https://doi.org/10.1007/s10115-024-02110-w.
50. Varitimadis.; all. GraphCoBots Administration Tool. https://memobot-dashboard.azurewebsites.net/. Accessed: 2026-02-01.
51. Nielsen, J. *Usability Engineering*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1994.
52. Van Teijlingen, E.; Hundley, V. The importance of pilot studies. *Social Research Update* **2001**, *35*.
53. Brooke, J. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*; Jordan, P.W.; Thomas, B.; Weerdmeester, B.A.; McClelland, I., Eds.; Taylor and Francis: London, UK, 1996; pp. 189–194.
54. Sauro, J. Interpreting a SUS score. Available online: https://measuringu.com/interpret-sus-score/, 2018. Accessed: 2025-04-01.
55. Yenduri, Y.G.; et al. GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access* **2024**, *12*, 54608–54649. https://doi.org/10.1109/ACCESS.2024.3389497.
56. Huang, X.; Ruan, W.; Huang, W.; Jin, G.; Dong, Y.; Wu, C.; Bensalem, S.; Mu, R.; Qi, Y.; Zhao, X.; et al. A Survey of Safety and Trustworthiness of Large Language Models through the Lens of Verification and Validation. *Artificial Intelligence Review* **2024**, *57*, 175. https://doi.org/10.1007/s10462-024-10824-0.
57. Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; Yang, H.; Zhang, J.; Chen, Z.; Tang, J.; Chen, X.; Lin, Y.; et al. A Survey on Large Language Model Based Autonomous Agents. *Frontiers of Computer Science* **2024**, *18*, 186345. https://doi.org/10.1007/s11704-024-40231-1.
58. Trichopoulos, G.; Ordoumpozanis, K.; Caridakis, G. An Evaluation of LLM-based Chatbots for Enhancing the Visitor's User Experience at Cultural Exhibits. *Journal on Computing and Cultural Heritage* **2025**. Accepted for publication, https://doi.org/10.1145/3775062.
59. Kumar, P. Large Language Models (LLMs): Survey, Technical Frameworks, and Future Challenges. *Artificial Intelligence Review* **2024**, *57*, 260. https://doi.org/10.1007/s10462-024-10888-y.
60. Zhao, P.; Zhang, H.; Yu, Q.; Wang, Z.; Geng, Y.; Fu, F.; Yang, L.; Zhang, W.; Jiang, J.; Cui, B. Retrieval-Augmented Generation for AI-Generated Content: A Survey. *Data Science and Engineering* **2026**. https://doi.org/10.1007/s41019-025-00335-5.
61. Popov, R.O.; Karpenko, N.V.; Gerasimov, V.V. Overview of small language models in practice. In Proceedings of the CS&SE@SW 2024: 7th Workshop for Young Scientists in Computer Science & Software Engineering, Kryvyi Rih, Ukraine, 2025. December 27, 2024.