HHAI 2024: Hybrid Human AI Systems for the Social Good
F. Lorig et al. (Eds.)

doi:10.3233/FAIA240193

Assessing the *HI-ness* of Virtual Heritage Applications with Knowledge Engineering

Delaram JAVDANI RIKHTEHGAR [a,1], Ilaria TIDDI [b], Shenghui WANG [a], Stefan SCHLOBACH [b], and Dirk HEYLEN [a]

[a] *University of Twente, Enschede, The Netherlands*
[b] *Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

ORCiD ID: Delaram Javdani Rikhtehgar https://orcid.org/0009-0007-8266-8395, Ilaria Tiddi https://orcid.org/0000-0001-7116-9338, Shenghui Wang https://orcid.org/0000-0003-0583-6969, Stefan Schlobach https://orcid.org/0000-0002-3282-1597, Dirk Heylen https://orcid.org/0000-0003-4288-3334

Abstract. Virtual Heritage exhibitions aim to engage a diverse audience through the integration of Virtual Reality and various AI technologies, including Artificial Agents, and Knowledge Graphs. Understanding the nuances of human-agent interactions is crucial to fully harness the potential of these technologies and deliver personalized and captivating experiences. Evaluating the alignment of Virtual Heritage applications with the vision of Hybrid Intelligence – where humans and machines collaborate toward a common goal – presents a significant challenge. In this paper, we investigate the assessment of Hybrid Intelligence within the Virtual Heritage domain using Knowledge Engineering methods. Through the analysis of six different scenarios presented as workflows of tasks and input/output data, we identify and compare classical Knowledge Engineering tasks with HI-specific tasks to measure the level of HI-ness achieved. Our study focuses on evaluating the synergy achieved by mixed teams during various tasks as a measure of HI-ness. The findings provide insights into the effectiveness of Knowledge Engineering to identify HI aspects within existing applications, the potential for quantifying and improving HI-ness in an application, and the identification of modeling limitations.

Keywords. Virtual Heritage, Knowledge Engineering, Hybrid Intelligence, Human-Computer Interaction, Personalization

1. Introduction

We investigate the use of Knowledge Engineering methods to assess the nature of Hybrid Intelligence [1] in a virtual reality-based application within the domain of Cultural Heritage (CH). Recent advancements in Virtual Reality (VR) technologies have opened up new possibilities for attracting and engaging end-users through immersive and interactive experiences in the realm of art, history, and culture exploration [2,3]. To enhance accessibility to cultural heritage, institutions have integrated multimodal technologies to pro-

[1]Corresponding Author: Delaram Javdani Rikhtehgar, d.javdanirikhtehgar@utwente.nl

vide a more realistic and immersive visitor experience [4], attract diverse audiences [5], foster collaborative meaning-making [6], and elicit more natural interactions [7].

With the growing instances of Virtual Heritage (VH) [8] applications, virtual agents within intelligent environments play an important role in improving user experiences and satisfaction by facilitating effective communication and interactions. It is therefore crucial to understand the intricacies of human-agent interactions to leverage the full potential of AI technologies in VR and provide personalized and engaging experiences.

In pursuit of this understanding, we have designed an ontology that allows us to model complex scenarios involving heterogeneous data, models, agents, and interactions. This ontology facilitates the formalization of scenarios, such as (1) a virtual museum guide equipped with Theory of Mind capability (i.e., the cognitive ability to attribute and understand mental states such as beliefs, intentions, and emotions, enabling the interpretation and prediction of behavior based on these mental states [9]), assessing a visitor's emotional states in real-time and adapting its painting description to increase engagement; (2) a collaborative and immersive exploration of the exhibition, where the virtual agent delivers a rich, educational experience based on user preferences, suggesting new paintings serendipitously and observing user reactions; and (3) a trustworthy museum tour where the user learns to trust the agent's recommendations, as it provides transparent explanations on preferences, historical context, and the overall narrative flow of the exhibition.

The symbiotic collaboration of heterogeneous actors toward shared goals, the adaptability to each other's capabilities, and the use of heterogeneous data and processing methods makes Virtual Heritage applications typical Hybrid Intelligence scenarios. The field of Hybrid Intelligence (HI) envisions the development of collaborative systems, where humans and intelligent machines operate in mixed teams collaboratively, synergistically, and proactively to achieve shared goals [1]. In a typical HI scenario, humans and artificial agents complement each other's limitations — i.e. human stereotyping, error proneness, in-group favoritism, or short memory are mitigated by machine-driven decision-making, while human feedback ensures machine fairness, sample efficiency, and task generalizability.

Making a VR exhibition a hybrid intelligence application is crucial for maximizing collaboration between human visitors and AI-driven agents, particularly in the context where their collaboration is centered on enhancing the overall experience, facilitating exploration, and providing personalized tour. By leveraging the diverse strengths of both humans and machines, the exhibition can effectively address their common goal of engaging visitors and transferring knowledge. This collaborative approach amplifies the exhibition's ability to tailor content delivery, guide immersive exploration, and maximize understanding and enjoyment of its content, ultimately fostering deeper engagement and satisfaction among visitors. In essence, integrating hybrid intelligence elevates the exhibition's interactive capabilities and enriches the overall user experience.

In this context, a full understanding of the inner workings of the hybrid interaction between users and virtual agents in our scenario is still missing, making us lose the potential of a fully-fledged HI application. In other words, the challenge we face is to determine whether our scenarios are HI-compliant, and more specifically, how much *HI-ness* manifests in the Virtual Guided Tour (VGT) application. We focus on assessing the level of HI-ness in terms of synergy between humans and machines; the stronger the HI, the greater the synergy in collaboration.

To tackle this problem, this paper suggests the use of Knowledge Engineering [10] to measure the HI-ness of an existing application, focusing on the Virtual Guided Tour application. Knowledge Engineering allows to elicit, structure, formalize, and operationalize the information, knowledge, and tasks involved in knowledge-intensive applications. Methods such as CommonKADS [11] supported engineers in clarifying the structure of complex applications in the past. Its adaptation was recently used to define the most typical tasks, inputs/outputs, and knowledge roles (called the application's *Knowledge Model*) in Hybrid Intelligence applications [12]. Here, we bring forward the idea that the HI Knowledge Model can be used as an analytical toolkit to measure the HI-ness of the VGT application. We call it the Hybrid Intelligence Knowledge Engineering toolkit (or HIKE), and aim at answering the following two research questions: (1) *can we identify the HI aspects in our scenarios using HIKE, and if so, which ones?* and (2) *how can we use HIKE to evaluate the HI-ness of an application?* In order to answer these questions, we start from a set of Virtual Guided Tour scenarios and then formally describe all processes, inputs and outputs involved therein to create a Knowledge Model for the VGT application. We then decompose the scenarios into workflows of tasks and inputs/outputs, and evaluate (i) the amount of specific HI tasks as compared to classical KE tasks and (ii) the strength of this HI-ness using the tasks' qualitative analysis. The workflows ultimately help us assessing whether the Hybrid Intelligence Knowledge Engineering toolkit helps us sufficiently describe the multiple HI aspects of our scenarios.

2. Related Work

Our research is related to Virtual Heritage applications and Knowledge Engineering methods for Hybrid Intelligence.

Virtual Reality and Virtual Heritage Applications. The potential of extended reality (XR), notably augmented reality (AR) and virtual reality (VR), has been widely acknowledged by museums [13]. VR, in particular, has demonstrated its effectiveness in facilitating learning [14,15,16,17,18] and fostering student motivation [19]. In [20], a novel didactic evaluation method for Digital Cultural Heritage (DCH) learning in higher education, underscoring VR's value in teaching, was introduced. The research highlights the experiential and engaging aspects of VR, emphasizing its compensatory role in guiding learners compared to traditional teaching methods.

VR has also proven instrumental in enhancing visitor experiences within museums [21,22]. [23] found that VR, by capturing users' attention, strengthens engagement, leads to an improved overall user experience. The immersive nature of VR allows users to explore the realms of active imagination, creating a heightened sense of presence within the virtual environment [24]. [25] demonstrated that VR can stimulate the intention for physical visits to museums, emphasizing the profound impact of VR experiences. Authors of [26] emphasized the compelling opportunities VR provides for museum visitors to engage with places or objects that may be challenging to exhibit physically due to budget constraints, limited space, or staffing issues. The role of VR in creating interactive audience experiences, emphasizing how understanding user attention and behavior in VR informs the creative process, is discussed in [27].

The field of Virtual Heritage integrates various interactive technologies aimed at engaging a diverse audience and creating personalized visitor experiences [28]. One category involves multimodal technologies which contribute to the immersive nature of VR experiences, thereby enhancing visitors' overall experience [29]. Conversational agents have been incorporated within virtual museums to serve as guides to visitors and substantially improve the accessibility of information [30,31,32]. Furthermore, eye-tracking technology has emerged as a valuable tool in VR to offer insights into users' behavior and interactions. This technology enables the creation of tailored experiences by monitoring and understanding visitors' gaze patterns, allowing for the customization of content based on individual preferences [33,34]. Beyond behavioral analysis, eye-tracking is employed to discern visitors' emotional states [35]. Additionally, eye-tracking technology has been leveraged to monitor visitor learning experience [36].

Knowledge Engineering and Hybrid Intelligence. Knowledge Engineering allows to design and formalize knowledge-based systems following software engineering methodologies [10]. Different modeling paradigms address different engineering aspects, i.e. CommonKADS [11] allows to describe models, MIKE [37] formalizes the execution of the models, and PROTÉGÉ [38] allows collaborative knowledge acquisition and reasoning. CommonKADS has been applied to characterize a variety of single-agent scenarios (e-governance, smart grid management and robot control) [39,40,41], or multi-agent ones (e.g. supply chain management and traffic simulations) [42,43]. These scenarios are limited to the design of classical KE tasks, such as diagnosis, assessment, planning. New KE tasks that adapt to the open-endness of the modern applications have only been proposed in the Semantic Web and Hybrid Intelligence areas [12,44].

With the rise of Hybrid Intelligence as a field, several formalizations appeared. An official conceptualization of HI as a field (in the lines of the Human-Computer Interaction ontology [45]) is yet to be achieved [12]. The HI research agenda presented in [1] defines research challenges and solutions in HI (Theory of Mind for synergetic collaboration, reinforcement learning for adaptability to changing environments, societal and personal value-awareness integrated in the systems, knowledge graphs as background knowledge to increase trust in the team). A taxonomy to design HI systems is proposed in [46]. Standard models and design patterns for human-machine team collaboration were presented in [47,48,49]. A preliminary model to characterize and evaluating HI systems is presented in [50], based on a user-study that identifies the importance of team properties (boundedness, interdependence, competency, purposefulness etc.). An ontology describing actors, interactions and information processing types, with a set of common HI tasks and sub-tasks, is presented in [12]. Our goal is not to propose a new HI terminology, but rather apply the existing terminologies to assess the HI-ness of our application.

3. Preliminaries

In order to show how to use the method of [12] to measure the HI-ness of our Virtual Guided Tour application, this section introduces a few preliminaries, namely a set of relevant VGT scenarios and the basics of the HIKE toolkit.

3.1. Scenarios

In the VGT, a virtual agent accompanies a visitor (let us call her Sarah) through a VR exhibition. Information about the exhibition's objects (e.g. paintings) is stored in a cultural knowledge graph that the agent uses to guide Sarah throughout, recommending her relevant art objects and answering questions she may have. Sarah interacts with the agent, creating a personalized tour where her preferences and interest are considered.

S1. Multimodal Interactions. The agent captures Sarah's gaze, facial expressions, and speech inputs, employing multimodal analysis and reasoning to understand her interests. Based on the interpreted multimodal inputs, the agent dynamically tailors its interaction (e.g. highlighting specific objects in the painting to guide visual attention or encourage hand grabbing, offering additional cultural background, recommending related paintings), thus providing Sarah with personalized information according to her interests.

S2. Real-time Understanding of User States. The agent employs Theory of Mind to assess Sarah's beliefs, intentions, and emotions as she interacts with the agent within the virtual environment. The agent dynamically adjusts its responses to align with Sarah's beliefs or intentions, provides emotional support to address her feelings, or changes its own beliefs, intentions, and emotions to improve future interactions.

S3. Memory Integration for Personalization. The agent, equipped with memory capabilities, observes Sarah's actions, recalling her preferences (e.g. a favorite painter from a past session), and integrates them with information from previous users to enhance the exhibition experience. The agent adapts its explanations based on known user patterns and preferences, giving more emphasis to artists and styles resonating with users sharing similar tastes. The agent actively learns from its past actions, storing this knowledge in its memory, to inform and improve future interactions with users.

S4. Perception Building. An ongoing communication between Sarah and the agent takes place. The agent employs analytics, e.g. gaze-based measures, to learn information about Sarah's interests or knowledge level. Simultaneously, Sarah gains insights into the agent's capabilities and limitations, understanding the type of information it can and cannot provide, shaping her perception of the agent's abilities.

S5. Transparency and Explainability. Intrigued by the agent's highlighted paintings, Sarah asks for the reasons behind these choices. In response, the agent provides a transparent explanation, detailing the information it utilized (e.g. Sarah's interest and preferences, her past interactions, and insights gathered from other users' engagements) and the overall narrative flow of the exhibition. This transparent communication enhances Sarah's trust in the agent's guidance, enriching her overall experience.

S6. Collaborative Pursuit of Goals. Sarah and the agent engage in a collaborative exploration. The agent's goal is to foster understanding of the exhibition's theme by showcasing curated paintings, while Sarah focuses on learning different art styles. The agent adapts its guidance to accommodate both educational goal and artistic appreciation. Their collaborative efforts result in a comprehensive and immersive exploration, satisfying both individual and shared objectives.

3.2. HIKE – The Hybrid Intelligence Knowledge Engineering Toolkit

Taking inspiration from CommonKADS' templates to describe the behavior of knowledge-intensive applications, the work of [12] proposes a Knowledge Model for Hybrid Intelligence applications.

In CommonKADS, a Knowledge Model usually specifies the vocabulary used, i.e. the main classes of the domain (agents, users, paintings in our case) and the processing task(s) performed over such classes (e.g. diagnosis, classification, etc.). These are articulated in three layers: (i) a task layer including the main tasks to be solved, (ii) an inference layer, where these tasks are decomposed into more fine-grained, minimal functionalities (called inferences) and (iii) a domain layer including an ontology with functional terms that serve as inputs and outputs of these functionalities.

Following this structure, the HI Knowledge Model includes a high-level ontology of that describes the most important components in an HI application (actors, processing information, interaction, context), and a taxonomy of HI-relevant tasks (including reasoning, prediction, recognition, actions) that can be achieved by humans and intelligent machines operating in a team. This model is ultimately used as a toolkit to describe and compare HI applications. Similarly, we suggest to use HIKE to measure the HI-ness of the Virtual Guided Tour application. In order to do that, a Knowledge Model of the Virtual Guided Tour needs to be defined, as presented in the next section.

4. The Knowledge Model of the Virtual Guided Tour application

To build the Virtual Guided Tour Knowledge Model, we use a *middle-out* strategy: we first define the application's domain ontology (Section 4.1), and then identify the tasks and sub-tasks by linking the domain ontology to the HI one (Section 4.2).

4.1. An Ontology for Virtual Guided Tour

Our Virtual Guided Tour (VGT) ontology represents users (visitors), their actions and interactions with the virtual agent (guide) in a VR environment, following principles of Human-Computer Interaction (HCI) [51] and user modeling [52]. As illustrated in Figure 1, our ontology delineates six key components: `Actor`, `Measures`, `Environment`, `Event`, `Action`, and `Object`, capturing their relationships.[2]

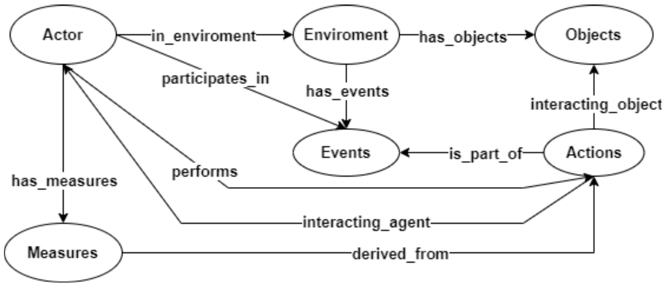


Figure 1. Key VGT components: `Actor`, `Measures`, `Environment`, `Event`, `Action`, and `Object`.

The `Actor` component encapsulates information about users or agents navigating the virtual environment, participating in events, and executing actions. This includes a

[2]Due to space limitations, we show here the core elements of the ontology and discuss its specific components in the text. The full ontology can be found in our repository (cfr. Section 5).

comprehensive set of user details such as familiarity with the VR environment, knowledge level about the exhibition, and demographic information (e.g. age, culture, education, gender, and language). Such information allows the agent to adapt to users' individual preferences and analyze diverse user behaviors to understand underlying patterns. Additionally, details about agents involved in user interactions, such as their capabilities, tasks, and configurations, are also recorded. Each actor performs specific actions during an event, such as the user viewing an area of interest within a painting, or the agent answering users' questions.

The `Action` concept can be further categorized into types such as `Feedback`, `Recommendation`, `Answer`, `Question`, `Comment`, `View`, `Explanation`, `Read`, and `Click`. These actions encompass modalities like `Gesture`, `Voice`, `Gaze`, and `Facial Expression`, each associated with start time, duration, and the actor's location in the environment during the action. These actions are geared towards achieving a specific `Goal` [51], driven by the actor's `Intentions`, rooted in their overarching goal, giving rise to an `Interpretation` for an interacting actor.

Inferences about users can be drawn based on various `Measures` such as gaze-based metrics (fixation count, duration, dwell time, transitions, and scan paths), as well as measures like the number of questions asked and response time. These measures provide valuable insights into information processing, decision-making intricacies, attention patterns, and search strategies [53]. Additionally, they can indicate aspects of user curiosity, attention, and interest [54,55], helping determine emotional states, interests, cognitive characteristics, and personality [56,57,58,59,60,61,62].

The `Environment` section encompasses various aspects of the virtual environment, including the VR space map, room configurations, domain context, and contextual information about the exhibition. Furthermore, it addresses the goal of the environment, such as conveying the intended message, promoting diversity and inclusion, and ethical considerations related to the virtual environment.

The `Object` segment provides detailed information about elements within the environment, such as paintings and their textual descriptions. These details include attributes like name, description, story, and coordinates that indicate the object's position in the environment. Additionally, we cover `Area of Interest` (AOI) within objects, along with coordinates and associated semantics. These AOIs can represent specific elements (e.g. a person or a building) whose details – such as who he is, what he is proficient in, or the name of the building and its location – can be linked to specific cultural heritage knowledge graphs.

4.2. *Inferences and Tasks*

In the second phase, we link the VGT ontology to the HI Knowledge Model (refer to Figure 2) to determine the main processes our scenarios deal with.

Firstly, the terms of the domain ontology are mapped to the terms of the ontology of HI knowledge roles. These consist of a few abstract classes and relationships indicating the roles entities play in the reasoning process of any HI application. These terms can be used as inputs and outputs of the minimal processes (inferences) of the application, allowing us to reconstruct the general process in what is called 'task decomposition.' By creating these mappings, we achieve two goals: (1) we can describe VGT scenarios in HI terms — including mixed actors, their capabilities, their interactions, and processing

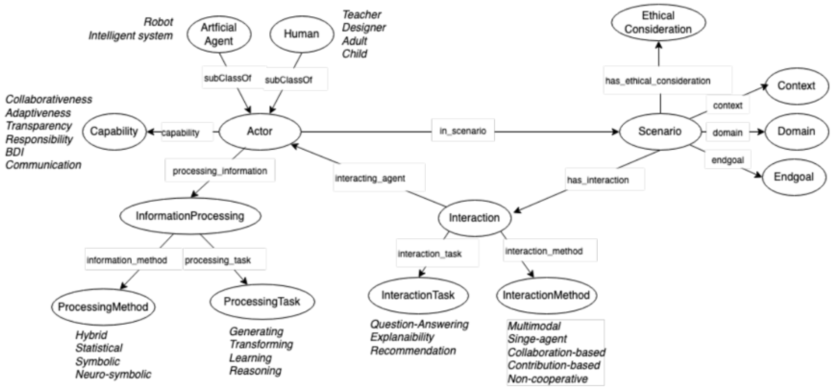


Figure 2. HI knowledge model [12]

methods, and (2) we can link these terms to the minimal functionalities and derive task decompositions for our VGT scenarios.

Specifically, the classes `vgt:Actor` and `vgt:Capability` are linked respectively to `hi:Actor` and `hi:Capability` through an `rdfs:subClassOf` relationship. Through inheritance, we can describe the agent in terms of the `hi:Interaction` happening between them (through the classes `hi:InteractionTask` and `hi:InteractionMethod`) and the `hi:ProcessingTask` and `hi:ProcessingMethod` they undertake (associated with the class `hi:InformationProcessing`). To map between the classes `vgt:Environment` and `hi:Scenario`, we use the relationship `skos:broader` given its broader semantics. This allows us to describe `hi:Domain`, `hi:Context`, `hi:EndGoal`, and potential `hi:EthicalConsiderations`, if needed. Table 1 summarizes the knowledge roles and tasks in our scenarios. For example, in Scenario 1 (S1), the processing method is "Multimodal Analysis & Neurosymbolic Reasoning," and the identified processing task is "Reasoning." This reflects the cognitive process the virtual agent uses to interpret Sarah's multimodal cues and customize the interaction accordingly.

Table 1. Knowledge roles of the Virtual Guided Tour. The classes `hi:Actor` and `hi:Domain` are omitted, as all scenarios deal with Sarah and the virtual agent, operating in a Virtual Guided Tour application.

	hi:EndGoal	hi:ProcessingMethod	hi:ProcessingTask
S1	Multimodal Interaction	Multimodal Analysis & Neurosymbolic	Reasoning
S2	Real-time State Understanding	Theory of Mind	Generate
S3	Memory Integration	Symbolic techniques	Transform
S4	Perception Building	Multimodal Analysis & Theory of Mind	Learning
S5	Explainability	Statistical techniques	Generate
S6	Collaborative Pursuit of Goals	Reinforcement Learning	Learning, Generate

The next step is to determine a set of relevant fine- and coarse-grained processes that can help us build the task layer of the Virtual Guided Tour application. Recall that, in a task decomposition process, coarse-grained tasks are decomposed into smaller, fine-grained sub-tasks hierarchically, and the leaves at the lowest level are the primitive processes (inferences) linked to the terms of the ontology (the inputs and outputs of such processes). As suggested by HIKE, our primitives include (i) "transfer functions" that allow communication between agents (*obtain*, *receive*, *present*, *provide* a given input/output), (ii) classical KE processes (*abstract*, *select*, *transform* certain data), and (iii) VR-specific processes such as *perceive* through a VR headset or *interact* with virtual objects. We also maintain the distinction between static inputs (e.g. the cultural heritage Knowledge Graph, a room in the VR exhibition) and dynamic inputs (e.g. an explanation to Sarah, Sarah's eye gaze, etc.). Finally, our scenarios rely on both the physical and the computational component, hence we use all the high-level tasks proposed by the HIKE toolkit, e.g. *recognition* of Sarah's mood, *prediction* of the next painting to recommend, *reasoning* over Sarah's interests (deductively based on historical data, or inductively based on her current visual attention), and *action* (including both cognitive actions based on Theory of Mind or memory and physical actions such as gaze or movements).

5. Validation

The VGT Knowledge Model is now used to answer our research questions: namely, whether and which HI aspects can be identified using the HIKE toolkit (Section 5.1), and how to use HIKE to evaluate HI-ness of our scenarios (Section 5.2).

5.1. RQ1 : Task Decomposition of VGT scenarios

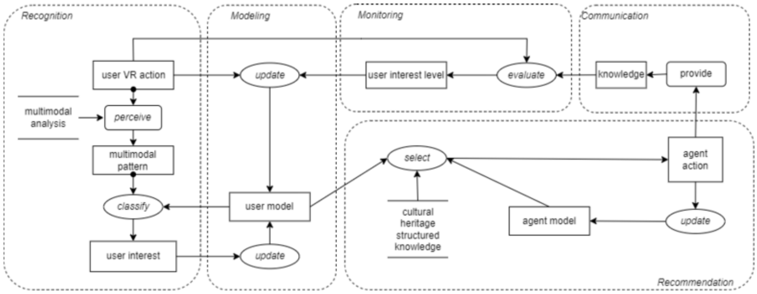
Our goal here is to assess the effectiveness of HIKE in identifying aspects of Hybrid Intelligence within the VGT scenarios. To achieve this, we use the VGT Knowledge Model to represent our scenarios in the form of processing workflows[3]. We then compare the number of classical KE tasks vs. the HI tasks identified by the HIKE toolkit within these workflows. Given space constraints, presenting decompositions for all six tasks is impractical. Therefore, for illustrative purposes, Figure 3 presents S1, S2, S3, and readers are encouraged to refer to our online repository[4] to access the complete set of the material. Addtionally, the high-level tasks of the scenarios are outlined in Table 2.

We recognize in our scenarios recurring tasks such as *Recognition, Communication, Monitoring, Modeling* and *Adaptation*. Additionally, we observe *Recommendation, Recollection, Explanation, Decision Making*, and *Perception* in all scenarios but S2. In Table 2, we mark HI-specific tasks as ✓✓(*Communication, Explanation, Adaptation, Perception, Collaborative Decision Making*), open-ended knowledge engineering tasks as ✓(*Recollection, Recognition, Recommendation*), and classical CommonKADS tasks without checkmark (*Monitoring, Modeling*). All scenarios include at least one HI task, which not only demonstrates the existence of HI-specific tasks, but also reflects the HI concepts of adaptability, collaborativeness, and explainability at the core of the HI agenda [1].

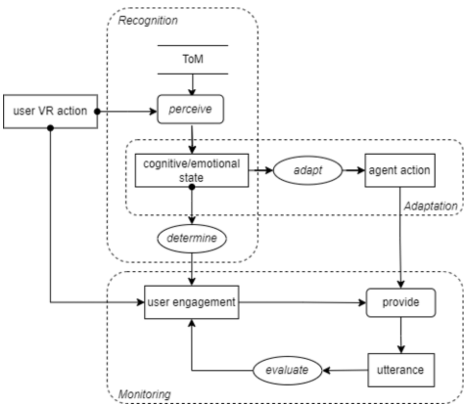
Additionally, the decompositions in Figure 3 reveal new, recurrent elements that are typical of hybrid environment: sub-tasks such as *perceive* multimodal inputs through a VR headset (all scenarios), *adapt* behavior (S2, S3), and *update* memory (S1, S6); inputs/outputs types such as multimodal (VR) actions (all scenarios) and Theory of Mind

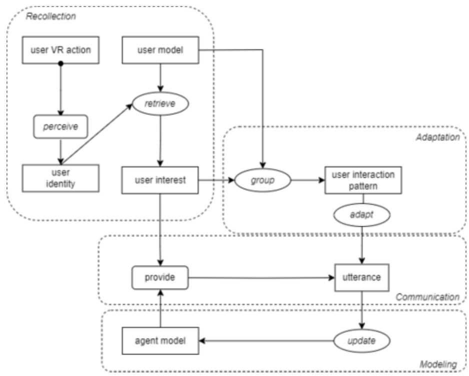
[3]Note the recommended UML notation: ovals (inferences), rounded rectangles (transfer functions), full squares (dynamic inputs/outputs), horizontal lines (static inputs/outputs), dotted arrows (lists), dashed boxes (more general tasks).

[4]https://github.com/DelJavdani/HHAI2024-Assessing-the-HI-ness-of-Virtual%2DHeritage-applications-with-Knowledge-Engineering.git



(a) S1: Multimodal Interaction.





(b) S2: Real-time Understanding of User States. (c) S3: Memory Integration for Personalization.

Figure 3. Task Decomposition for Scenarios S1, S2 and S3.

(S2, S6). These new aspects also confirm the suitability of the task decomposition to identify HI-ness of an application, but suggest that additional terms might be needed within HIKE. For example, we face challenges in explicitly attributing tasks to specific actors (agent, user, or the creator of the virtual exhibition), hampering the use of inferences such alter virtual environment or interact with virtual object.

5.2. RQ2: Assessing HI-ness of VGT Scenarios

To address our second question, we focus on assessing the level of HI-ness in terms of synergy between humans and machines (i.e. the stronger the HI, the greater the synergy in collaboration). In Table 2, we describe the single tasks we identified in each scenario, and mark them with a Weak HI level if a task involves a mixed team without collaboration toward the same goal, and a Strong HI level if the teams actually create synergy. Ultimately, we provide an overall percentage to assess the HI-ness in the entire scenario.

We observe that the *Modeling* tasks (S1, S3) primarily involve data storage and retrieval of a single agent, representing minimal synergy and hence a weak level of Hybrid Intelligence. Similarly, tasks categorized under *Recognition* (S1, S2, S5, S6) primarily focus on pattern recognition or acquiring user information, indicating slightly increased collaboration within the team, but still demonstrating weak HI due to a primarily data-driven interaction rather than true collaboration.

Table 2. VGT scenarios, task description, task HI-ness specificity and level, and total scenario HI-ness

S*	Task	Task Description	HI Task	HI Level	HI-ness
S1	Modeling	Storing user VR multimodal actions, user interest, agent actions, and communication style of the agent		Weak	20%
	Recognition	Determining user interest based on VR action and updating the user model	✓	Weak	
	Recommendation	Selecting agent actions based on agent and user models	✓	Strong	
	Communication	Providing knowledge to the user	✓✓	Weak	
	Monitoring	Evaluating user interest based on current user actions after providing knowledge		Weak	
S2	Recognition	Determining the user's cognitive state through Theory of Mind techniques	✓	Weak	33%
	Adaptation	Adapting agent actions based on recognized user state	✓✓	Strong	
	Monitoring	Evaluating user engagement based on current actions after providing knowledge		Weak	
S3	Recollection	Recollecting information about the user, such as interests, previous interactions	✓	Strong	50%
	Adaptation	Adapting agent utterances based on collective patterns of users stored in memory	✓✓	Strong	
	Communication	Providing knowledge to the user	✓✓	Weak	
	Modeling	Storing agent actions		Weak	
S4	Perception	Shaping perception about the agent's capabilities	✓✓	Strong	100%
	Communication	Providing information about capability to the user	✓✓	Strong	
S5	Recognition	Recognizing the intent of the user's question	✓	Weak	50%
	Explanation	Providing the reason behind the agent's action	✓✓	Strong	
S6	Recognition	Recognizing the goal of the user based on their actions	✓	Weak	50%
	Decision Making	Deciding the learning path based on system and user objectives	✓	Strong	

To enhance these weaker tasks and bolster HI, it is essential to incorporate mechanisms where the agent seeks clarification or confirmation from the user when uncertain about its decisions regarding user interests or mental state. By integrating user confirmation, ideally coupled with clear explanations, the agent can refine its decision-making strategies, thereby fostering better judgment in analogous situations in the future. This interactive process ultimately culminates in a heightened level of hybrid intelligence, characterized by enhanced capabilities in terms of collaboration, adaptivity, and explainability.

In contrast, the *Recommendation* tasks (S1), where the agent chooses actions based on both user and agent models, stand out as a strong HI component. Here, the agent refines its actions and adapts information to the user, showcasing a high degree of synergy. The *Adaptation* tasks (S2, S3) also show strong HI-ness, by dynamically adapting agent actions based on user states and collective patterns from memory. Conversely, the *Com-*

munication (S1, S3) and *Monitoring* tasks (S1, S2) lean towards weak HI-ness, emphasizing one-way information transfer and post-information evaluation. *Communication* in S4 can be considered strong as the agent provides information about its capabilities and limitations (e.g. "I can only answer factual questions", "I am an expert in 17th-century paintings" or "I am not familiar with the fashion style back then"). This helps the user understand the agent's capabilities and foster *Team Awareness* – an important HI capability. Examining the *Explanation* task (S5) reveals a strong HI aspect, as providing explanations for the agent's actions involves building trust and understanding. Similarly, the *Decision Making* task (S6) demonstrates strong HI-ness by considering both agent and user goals, fostering collaboration and mutual satisfaction.

Overall, 4 out of 6 scenarios present at least half of their tasks with strong HI-ness (50% or above). The weakest scenarios (S1, S2) involve classical KE tasks, but still include HI aspects such as adaptivity, Theory of Mind reasoning and user-agent shared understanding. In the scenarios with 50% HI-ness, collaboration through communication appear to be the key HI component. S4, involving building perception jointly through communication, finally appear to be the scenario with the strongest HI-ness.

Notice that HIKE serves not only as an analytical tool, but also as an active method to recommend adaptations and improvements of one's scenarios. For instance, S1 includes a HI task (*Communication*) whose the implementation is still evaluated as weak. This could suggest that the design of the Communication module should be modified, not only to increase the VGT's HI-ness, but potentially the overall quality of an application. Similarly, HIKE could help improving scenarios involving non-specific HI tasks. In the case of e.g. *Recognition* in S6, the curator might decide strengthen the role of the user by, for instance, using different methods for recognition (MTurks, explicit user-based models, etc.). This analysis confirms the idea that a Knowledge Engineering method such as HIKE can be used to assess, and possibly improve, the HI-ness of an application.

6. Conclusions

In this paper, we employed Knowledge Engineering to assess the level of HI-ness of a Vitural Guided Tour application designed for the Virtual Heritage domain. We described 6 different scenarios in the form of workflows of tasks and input/outputs, and identified the specific Hybrid Intelligence tasks performed within each scenario. We ultimately suggest to assess the HI-ness of the scenarios by (i) comparing the number of classical Knowledge Engineering tasks vs. HI-specific tasks, and (ii) measuring the strength of synergy achieved by the mixed team in the various tasks. Our study reveals interesting insights, including the usability of the Hybrid Intelligence Knowledge Engineering (HIKE) toolkit to identify HI aspects in an existing application, the possibility of measuring and improving its level of HI-ness, as well as the need of improving certain modeling aspects within HIKE. Furthermore, our findings indicate that HIKE can function as a design tool to facilitate the creation of HI scenarios. Future work will revolve around improving HIKE through validation with new scenarios, strengthening the concept of HI-ness, and standardization of the HI task templates (intended as recurrent combination of tasks and input/outputs) across HI initiatives. Additionally, the paper mainly focuses on the AI agent, with Table 2 presenting tasks from the agent's perspective. For future work, our aim is to investigate how Table 2 would appear when structured from the human task perspective.

References

[1] Akata Z, Balliet D, De Rijke M, Dignum F, Dignum V, Eiben G, et al. A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. Computer. 2020;53(8):18-28.

[2] Cecotti H. Cultural Heritage in Fully Immersive Virtual Reality. Virtual Worlds. 2022;1(1):82-102. Available from: https://www.mdpi.com/2813-2084/1/1/6.

[3] Marto A, Gonçalves A, Melo M, Bessa M. A survey of multisensory VR and AR applications for cultural heritage. Computers Graphics. 2022;102:426-40. Available from: https://www.sciencedirect.com/science/article/pii/S0097849321002120.

[4] Liarokapis F, Petridis P, Andrews D, de Freitas S. In: Ioannides M, Magnenat-Thalmann N, Papagiannakis G, editors. Multimodal Serious Games Technologies for Cultural Heritage. Cham: Springer International Publishing; 2017. p. 371-92. Available from: https://doi.org/10.1007/978-3-319-49607-8_15.

[5] C S, F P, G R, B L. A multimodal mobile museum guide for all. In: Mobile HCI'07 - 9th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 21–25, Singapore, 9-12 September 2007; 2007. .

[6] Ho CML, Nelson ME, Müeller-Wittig W. Design and implementation of a student-generated virtual museum in a language curriculum to enhance collaborative multimodal meaning-making. Computers Education. 2011;57(1):1083-97. Available from: https://www.sciencedirect.com/science/article/pii/S0360131510003520.

[7] Cao ZHPCea L. Real-time multimodal interaction in virtual reality - a case study with a large virtual interface. Multimed Tools Appl. 2023;82:25427–25448.

[8] Virtual Heritage: A Guide. Ubiquity Press; 2021. Available from: http://www.jstor.org/stable/j.ctv2dt5m8g.

[9] Premack D, Woodruff G. Does the chimpanzee have a theory of mind? Behavioral and brain sciences. 1978;1(4):515-26.

[10] Studer R, Benjamins VR, Fensel D. Knowledge engineering: Principles and methods. Data & knowledge engineering. 1998;25(1-2):161-97.

[11] Schreiber G, Wielinga B, de Hoog R, Akkermans H, Van de Velde W. CommonKADS: A comprehensive methodology for KBS development. IEEE expert. 1994;9(6):28-37.

[12] Tiddi I, De Boer V, Schlobach S, Meyer-Vitali A. Knowledge Engineering for Hybrid Intelligence. In: Proceedings of the 12th Knowledge Capture Conference 2023; 2023. p. 75-82.

[13] Shehade M, Stylianou-Lambert T. Virtual reality in museums: Exploring the experiences of museum professionals. Applied sciences. 2020;10(11):4031.

[14] Puig A, Rodríguez I, Arcos JL, Rodríguez-Aguilar JA, Cebrián S, Bogdanovych A, et al. Lessons learned from supplementing archaeological museum exhibitions with virtual reality. Virtual Reality. 2020;24:343-58.

[15] Hu-Au E, Lee JJ. Virtual reality in education: a tool for learning in the experience age. International Journal of Innovation in Education. 2017;4(4):215-26.

[16] Huang KT, Ball C, Francis J, Ratan R, Boumis J, Fordham J. Augmented versus virtual reality in education: an exploratory study examining science knowledge retention when using augmented reality/virtual reality mobile applications. Cyberpsychology, Behavior, and Social Networking. 2019;22(2):105-10.

[17] Gao L, Wan B, Liu G, Xie G, Huang J, Meng G. Investigating the effectiveness of virtual reality for culture learning. International Journal of Human–Computer Interaction. 2021;37(18):1771-81.

[18] Souchet AD, Philippe S, Lourdeaux D, Leroy L. Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. International Journal of Human–Computer Interaction. 2022;38(9):801-24.

[19] Lau KW, Lee PY. The use of virtual reality for creating unusual environmental stimulation to motivate students to explore creative ideas. Interactive Learning Environments. 2015;23(1):3-18.

[20] Paolanti M, Puggioni M, Frontoni E, Giannandrea L, Pierdicca R. Evaluating Learning Outcomes of Virtual Reality Applications in Education: a Proposal for Digital Cultural Heritage. ACM Journal on Computing and Cultural Heritage. 2023.

[21] Jung T, tom Dieck MC, Lee H, Chung N. Effects of virtual reality and augmented reality on visitor experiences in museum. In: Information and Communication Technologies in Tourism 2016: Proceedings of the International Conference in Bilbao, Spain, February 2-5, 2016. Springer; 2016. p. 621-35.

[22] Trunfio M, Lucia MD, Campana S, Magnelli A. Innovating the cultural heritage museum service model through virtual reality and augmented reality: The effects on the overall visitor experience and satisfaction. Journal of Heritage Tourism. 2022;17(1):1-19.
[23] Chang HL, Shih YC, Wang K, Tsaih RH, Lin Z. Using virtual reality for museum exhibitions: the effects of attention and engagement for National Palace Museum. 2018.
[24] Krassmann AL, Rocha Mazzuco A, Melo MC, Bessa M, Bercht M. Usability and sense of presence in virtual worlds for distance education: a case study with virtual reality experts. 2020.
[25] Lee H, Jung TH, tom Dieck MC, Chung N. Experiencing immersive virtual reality in museums. Information & Management. 2020;57(5):103229.
[26] Schofield G, Beale G, Beale N, Fell M, Hadley D, Hook J, et al. Viking VR: designing a virtual reality experience for a museum. In: Proceedings of the 2018 Designing Interactive Systems Conference; 2018. p. 805-15.
[27] Mu M, Dohan M, Goodyear A, Hill G, Johns C, Mauthe A. User attention and behaviour in virtual reality art encounter. Multimedia Tools and Applications. 2022:1-30.
[28] Champion E. Virtual heritage: a guide. Ubiquity Press; 2021.
[29] Liarokapis F, Petridis P, Andrews D, de Freitas S. Multimodal serious games technologies for cultural heritage. Mixed reality and gamification for cultural heritage. 2017:371-92.
[30] Liu D. Knowledge Graph Driven Conversational Virtual Museum Guide. University of Twente; 2021.
[31] Duguleană M, Briciu VA, Duduman IA, Machidon OM. A virtual assistant for natural interactions in museums. Sustainability. 2020;12(17):6958.
[32] De Carolis B, Macchiarulo N, Valenziano C. Marta: A virtual guide for the national archaeological museum of taranto. In: Proceedings of the 2022 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage. CEUR-WS. org; 2022. .
[33] Mokatren M, Kuflik T, Shimshoni I. Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage. Future generation computer systems. 2018;81:528-41.
[34] Yi T, Chang M, Hong S, Lee JH. Use of eye-tracking in artworks to understand information needs of visitors. International Journal of Human–Computer Interaction. 2021;37(3):220-33.
[35] Pelowski M, Leder H, Mitschke V, Specker E, Gerger G, Tinio PP, et al. Capturing aesthetic experiences with installation art: An empirical assessment of emotion, evaluations, and mobile eye tracking in Olafur Eliasson's "Baroque, Baroque!". Frontiers in Psychology. 2018;9:1255.
[36] Rainoldi M, Neuhofer B, Jooss M. Mobile eyetracking of museum learning experiences. In: Information and Communication Technologies in Tourism 2018: Proceedings of the International Conference in Jönköping, Sweden, January 24-26, 2018. Springer; 2018. p. 473-85.
[37] Angele J, Fensel D, Landes D, Studer R. Developing KB-systems with MIKE. In: Domain Modelling for Interactive Systems Design. Springer; 1998. p. 9-38.
[38] Gennari JH, Musen MA, Fergerson RW, Grosso WE, Crubézy M, Eriksson H, et al. The Evolution of Protégé: an Environment for Knowledge-Based Systems Development. International Journal of Human-computer studies. 2003;58(1):89 123.
[39] Yang D, Tong L, Ye Y, Wu H. Applying CommonKADS and Semantic Web Technologies to Ontology-based e-government Knowledge Systems. In: ASWC. Springer; 2006. p. 336 342.
[40] Zhou M, Ren J, Qi J, Niu D, Li G. CommonKADS Methodology for Developing Power Grid Switching Orders Systems. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer; 2007. p. 87.
[41] Henao M, Soler J, Botti V. Developing a Mobile Robot Control Application with CommonKADS-RT. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer; 2001. p. 651.
[42] Nguyen J, Farrenkopf T, Guckert M, Powers ST, Urquhart N. Using Semantic Technology To Model Persona For Adaptable Agents. In: ECMS; 2021. .
[43] Sabri M, Garakani MF. Modeling an Agent-mediated Supply Chain Management System using MAS-CommonKADS Methodology. International Journal of Mechatronics, Electrical and Computer Technology. 2012;2(3):58-75.
[44] van Harmelen F, ten Teije A, Wache H. Knowledge Engineering Rediscovered: towards Reasoning Patterns for the Semantic Web. In: Proceedings of the fifth international conference on Knowledge capture; 2009. p. 81 88.
[45] Costa SD, Barcellos MP, de Almeida Falbo R, Conte T, de Oliveira KM. A core ontology on the Human–Computer Interaction phenomenon. Data Knowledge Engineering. 2022;138:101977. Available

from: https://www.sciencedirect.com/science/article/pii/S0169023X21000951.

[46] Dellermann D, Calma A, Lipusch N, Weber T, Weigel S, Ebel P. The Future of Human-AI Collaboration: a Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In: Hawaii international conference on system sciences (HICSS); 2019. .

[47] Dubey A, Abhinav K, Jain S, Arora V, Puttaveerana A. HACO: a framework for developing human-AI teaming. In: Proceedings of the 13th Innovations in Software Engineering Conference on Formerly known as India Software Engineering Conference; 2020. p. 1-9.

[48] Madni AM, Madni CC. Architectural framework for exploring adaptive human-machine teaming options in simulated dynamic environments. Systems. 2018;6(4):44.

[49] Meyer-Vitali A, Mulder W. Trustworthy Hybrid Team Decision-Support. In: Proceedings of the First International Conference on Hybrid Human-Machine Intelligence. Workshop on the Representation, Sharing and Evaluation of Multimodal Agent Interaction, befindet sich International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 13-17, Amsterdam, Netherlands. Frontiers of AI. IOS Press; 2022. .

[50] Dell'Anna D, Murukannaiah P, Dudzik B, Grossi D, Jonker C, Oertel C, et al. Toward a Quality Model for Hybrid Intelligence Teams. In: Proceedings of the 2024 International Conference on Autonomous Agents and Multiagent Systems. AAMAS '24. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems; 2024. p. 000–000.

[51] Costa SD, Barcellos MP, de Almeida Falbo R, Conte T, de Oliveira KM. A core ontology on the Human–Computer Interaction phenomenon. Data & Knowledge Engineering. 2022;138:101977.

[52] Heckmann D, Schwartz T, Brandherm B, Schmitz M, von Wilamowitz-Moellendorff M. Gumo–the general user model ontology. In: User Modeling 2005: 10th International Conference, UM 2005, Edinburgh, Scotland, UK, July 24-29, 2005. Proceedings 10. Springer; 2005. p. 428-32.

[53] Rahal RM, Fiedler S. Understanding cognitive and affective mechanisms in social psychology through eye-tracking. Journal of Experimental Social Psychology. 2019;85:103842.

[54] Guyer JJ, Fabrigar LR, Vaughan-Johnston TI. Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. Personality and Social Psychology Bulletin. 2019;45(3):389-405.

[55] Jirout JJ. Supporting early scientific thinking through curiosity. Frontiers in Psychology. 2020;11:1717.

[56] Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: taxonomy, review and current challenges. Sensors. 2020;20(8):2384.

[57] Berkovsky S, Taib R, Koprinska I, Wang E, Zeng Y, Li J, et al. Detecting personality traits using eye-tracking data. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; 2019. p. 1-12.

[58] Tarnowski P, Kołodziej M, Majkowski A, Rak RJ. Eye-tracking analysis for emotion recognition. Computational intelligence and neuroscience. 2020;2020.

[59] David-John B, Peacock C, Zhang T, Murdison TS, Benko H, Jonker TR. Towards gaze-based prediction of the intent to interact in virtual reality. In: ACM Symposium on Eye Tracking Research and Applications; 2021. p. 1-7.

[60] Zhang L, Wade J, Swanson A, Weitlauf A, Warren Z, Sarkar N. Cognitive state measurement from eye gaze analysis in an intelligent virtual reality driving system for autism intervention. In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE; 2015. p. 532-8.

[61] Peacock CE, David-John B, Zhang T, Murdison TS, Boring MJ, Benko H, et al. Gaze signatures decode the onset of working memory encoding. In: CHI2021 Eye Movements as an Interface to Cognitive State (EMICS) Workshop Proceedings. ACM; 2021. .

[62] Garbutt M, East S, Spehar B, Estrada-Gonzalez V, Carson-Ewart B, Touma J. The embodied gaze: Exploring applications for mobile eye tracking in the art museum. Visitor Studies. 2020;23(1):82-100.