sustainability

MDPI

Article

A Virtual Assistant for Natural Interactions in Museums

Mihai Duguleană [1,*], Victor-Alexandru Briciu [2], Ionuț-Alexandru Duduman [3] and Octavian Mihai Machidon [3]

[1] Department of Vehicles and Transportation, Faculty of Mechanical Engineering, Transilvania University of Brașov Eroilor 29, 500036 Brașov, Romania
[2] Department of Social Sciences and Communication, Faculty of Sociology and Communication, Transilvania University of Brașov, Eroilor 29, 500036 Brașov, Romania; victor.briciu@unitbv.ro
[3] Department of Electronics and Computers, Faculty of Electrical Engineering and Computer Science, Transilvania University of Brașov, Eroilor 29, 500036 Brașov, Romania; dudumanionut96@gmail.com (I.-A.D.); octavian.machidon@unitbv.ro (O.M.M.)
* Correspondence: mihai.duguleana@unitbv.ro

Received: 23 July 2020; Accepted: 21 August 2020; Published: 26 August 2020

Abstract: Artificial Intelligence (AI) and its real-life applications are among the most effervescent research topics of the last couple of years. In the past decade, stakeholders such as private companies, public institutions, non-profit entities, and even individuals, have developed and used various AI algorithms to solve a wide range of problems. Given the extended applicability and the disruption potential of this technology, it was just a matter of time until it connected to the field of cultural heritage. This paper presents the development of an intelligent conversational agent which was built to improve the accessibility to information inside a history museum. We present the cultural context, the application architecture, the implementation challenges, and the visitors' feedback. We created a smart virtual agent that interacts with users in natural spoken language. This involved the design and implementation of the artificial intelligence-based core responsible for understanding the Romanian language. A short survey regarding the tourist acceptance of the system was conducted at the premises of our partners, the Museum "Casa Mureșenilor" from Brașov, shows good acceptance levels from both visitors and museum staff. Given the flexibility of the implementation, the system can be used by a large array of stakeholders with minor modifications.

Keywords: cultural heritage; virtual reality; artificial intelligence; conversational agent

1. Introduction

Every once in a while, humanity has the chance to change its destiny by making a technological leap from the slowly-rising progress curve on which it is situated at that moment, to a newer, steeper one. This process is happening, thanks to innovation. Innovation has helped us to overcome obstacles that were once thought to be impossible to solve. People now travel, learn and entertain in ways which 30 years ago seemed to be dreamy.

In recent years there has been a significant advancement in the field of personal virtual assistants, based mostly on the core technologies used behind such as virtual and augmented reality. These have become widely accepted by the general public, and are used daily in personal and professional settings. Due to the general overload and thus lack of time, adults are unable to focus on social interactions, spend quality family time, or engage in recreational activities. Moreover, the increasing life expectancy in Europe causes people of a higher age to develop more health-related

conditions and limitations [1], thus requiring personal care and assistance. In these cases, intelligent virtual assistants can make a difference.

Today, we find ourselves reaching the final days of the EU's Horizon 2020 research and innovation program. Based on the experience of the first half of the initiative, the EU aims to boost the impact of its actions by setting new priorities such as delimiting the specific focus areas, investing in market-creating innovation measures, and supporting the access to open data. 2020 also marks an essential milestone in the expansion of interactive technologies, as these align with the priorities presented above. These technologies reached maturity thanks to new advances in commercial equipment, backed mainly by the entertainment industry. We can use today's wireless 3D glasses, inertial chairs, and wearable gadgets to build Virtual Reality (VR) apps that are more immersive than ever. These systems can react not only to the physical parameters of the users but also to their physiological state. There are already computer vision tools which determine whether users are enjoying what they are doing, and to what degree [2].

Artificial Intelligence (AI) is now on the verge of changing the way we live our lives. This technology allows us to provide a personalized experience in virtually any domain. We can use AI to work faster, learn more efficiently, and even shop cheaper. Using AI will become trivial in the following years. Amazon Alexa and Google Assistant already help us watch movies worth watching, read books worth reading, or visit places worth visiting.

However good AI may be, adoption is one of the most common issues faced by this technology. That is why we see linking AI with VR as one of the most important objectives of the entire scientific community activating in these two research fields. People are attracted to new ways of interaction. They like to communicate and share information as freely as possible. Seeing a human 3D avatar, having a dialog in natural language with a digital companion, and sharing emotions and thoughts with a software assistant can be possible if we connect these concepts. However, there are serious problems related to the design and implementation of such a system.

Our study assumes some of these challenges. From the human-computer interaction point of view, we believe it is essential to have systems that can communicate with people in their natural language. One useful real-life application of such technology is in the physical space of a museum. The finite number of actions and solutions is a fertile ground for a successful AI application.

Social and cultural inclusion of users is a critical element that such a system aims to foster by offering a virtual exploration of cultural content. This enables users to explore diverse cultural content from around the world and provides enhanced experiences by having increased accessibility through the unique interface that our system, titled "IA", integrates: an avatar engaging the user through spoken natural language interaction. This is of particular importance to the cultural engagement use case since research has shown that the presence of an avatar acting as a social agent in such an environment has positive and catalytic effects on user engagement, sense of immersion, and learning effectiveness. The sense of social presence supports the acquisition of knowledge [3], making this interactive cultural scenario more likely to increase cultural knowledge transfer.

2. Background and Related Work

2.1. Previous Work on Virtual Museum Guides

The idea of implementing such a system is not new. More than a decade ago, studies were already discussing the challenges posed by the development of an intelligent museum guiding system [4]. However, back then, there were technological limitations related to the processing power and sensorial capabilities. Within a few years, however, technology evolved rapidly and allowed researchers to focus on fine-tuning this paradigm to achieve good conversational agents. In [5], scientists were investigating conversational agents in real-life museum settings, mostly focusing on categorizing different parts of the agent's speech options. This is particularly helpful for improving the conversation (the Human-Computer Interaction—HCI), but it still eludes employing a smart conversational agent.

One of the most cited works linked to this study is about Ada and Grace, two virtual assistants that were developed in 2010 to increase the engagement of middle-school students in museum tours [6]. Although revolutionary, this endeavor had limited applicability, as the system was installed in a single museum and for only a short period (the Museum of Science from Boston, MA, USA). The two virtual bots could answer visitors only if the questions were precomputed. Following this study, researchers felt that, when it comes to multiple agents, a study about conversation coherence is necessary. Results showed that multiple agents are not entirely effective [7].

Speaking of efficiency, the research community found out that it is one of the hardest things to quantify. In 2011, the Boston Museum of Science showcased Coach Mike, a pedagogical agent, aimed to improve various experiences (visiting, learning to code, problem solving). The authors presented a preliminary study in [8] and concluded that further investigation is needed in order to see the possible improvement of the interaction.

In 2013, a research team built a virtual robot guide named Tinker, which, besides synthesized language, used nonverbal gestures, empathy, and other relational behavior to establish a reliable connection with its users [9]. Unfortunately, Tinker was a schematic avatar that users could interact with mainly with a keyboard, not in natural language.

The limited interaction possibilities urged researchers to find alternative ways of deploying the idea of intelligent museum guides. In 2015, a private company designed a virtual museum assistant that could be seen with the help of a smartphone [10]. However, the project had limited expansion capabilities. Moreover, the assistant was not available to the general public, had no support to foreign languages, and was thought only as a mobile app.

Using multimodal data to improve a user model was further exploited in the same year [11]. However, this study did not imply any of the benefits obtained from using some form of AI algorithm. Almost a decade later, the intelligence feature is added in studies such as [12]. Speech recognition, object recognition, eye gaze, and natural speech synthesis are some of the elements proposed by this new framework. Unfortunately, the authors found that implementing AI in such a framework is quite tricky, as did some of the following works. For example, in [13], the authors focused on user intention detection and the dialog management components. However, the use case presented in this paper relies mostly on statistical data and not on dynamic connections that can be achieved by using a true AI approach. The same can be said about the system modeled in [14], with the notable exception that this study strived on making recommendations based on the measured satisfaction of the user, almost in a similar manner to the reinforcement learning paradigm.

The study of Tavcar et al. theorizes on the architecture of a virtual museum guide [15]. However, there were no practical implementations to test whether the system was feasible, especially in locations where only the national language is spoken. In [16], the authors develop I3-CPM, an intelligent Mixed Reality platform for museums around the world. Although the idea was ahead of its time, the impact was rather small, as the platform was unusable by museums from non-English speaking countries.

In recent years, many researchers concluded that not only the virtual agents but also the interaction between them and users should be as natural and humane as possible. This idea was presented in [17] and [18]. Moreover, in [19], authors also took into account emotions, and even personality, to make systems more coherent. These traits however, do not change the absence of the "intelligent" nature of a conversational agent (which can only be achieved using AI). They mostly solve some of the psychological barriers faced by users when interacting with such systems.

In order to make the whole assisted museum visiting experience memorable, researchers used storytelling coupled with a specific narrating style to enhance the features of virtual avatars [20]. Although this study is carried out in virtual environments (virtual museums), using a story to enrich an avatar also works in real life setups [21].

Perhaps one of the closest studies linked to this research is presented in [22]. Our work was focused here on creating an agent as a software product connected to Europeana, Europe's most extensive digital library. This study however, has a different context, input parameters, operation mode, and final scope.

As the era of big data is emerging, the research community could not miss the opportunity to enjoy the benefits of this technology in museum recommender systems [23]. This use-case is, however, only the tip of the iceberg. Big data can be used to manage large digital libraries, make accurate user models for an infinity of specific locations, register and store millions of interactions with items showcased by real and virtual museums, and so on [24]. As technologies such as 5G or Internet of Things (IoT) expand, so do connected areas. Indoor environments will become smarter, even in the case of museums [25]. Unfortunately, there is still a long way to go until we have systems that can enhance and even automate environments of such nature.

Aside from the problems outlined above, there is also the problematic issue of synthesizing voices in other languages. Much work has been done to develop great English natural language processors (NLP), but foreign languages with specific phonetic contexts, syllable, global, and variable morphosyntactic information, such as Romanian, remain under-researched [26,27].

As one can see, the research initiatives from this domain are not coherent, do not offer support for the Romanian language, and have limited applicability. Building a system based on these prior works is time-consuming and requires a significant financial investment. Museums and art galleries from Romania (as well as from other countries where English is not the primary language) cannot use these to build their virtual guides unless they dispose of serious resources, which are not available in most cases. The present study tries to give an answer to two fundamental questions: "Is a virtual intelligent museum guide system feasible for museums and art galleries?" and "What are the remaining difficulties for museum visitors, even in the case of successful implementation and demonstration of the concept?"

2.2. Location and Cultural Background

The social and cultural context of our research is quite complicated. Brașov is a city located in the central part of Romania [28], as part of Brașov County. It is a destination of "future tourism development" [29], about 166 km north of Bucharest, the capital of Romania.

Although it has a large number of touristic assets, varying from cultural-historic, sports, and entertainment attractions, Brașov fails to retain tourists in the city for more than two days [30]. This situation reflects weak exploitation of tourism facilities and a low contribution to the local economy.

In this specific social and touristic context of the city, the Museum "Casa Mureșenilor" from Brașov is one of the most important memorial places in Romania, as a collecting institution, a house museum with a plethora of culturally associated objects. The institution was founded in 1968, following the donation of the descendants of Iacob Mureșeanu a member of the Romanian Academy, pedagogue, journalist, and an exceptional Romanian politician of 19th century. Since 2015, the museum has been reoriented towards a more intense promotion of contemporary culture and arts. Confronted with a communication barrier in delivering their message to young audiences, the museum staff decided to use modern technologies, such as virtual museum guides, to come closer to their actual and potential visitors.

Having such an openness to new technologies made this museum a good fit for our newly developed intelligent assistant. Based on our previous experience within the project eHERITAGE ("Expanding the Research and Innovation Capacity in Cultural Heritage Virtual Reality Applications") [31], we made good connections with museum custodians and art gallery curators from national and international landmarks. During our brief collaboration with the Museum "Casa Mureșenilor" from Brașov, we have inferred what would be the ideal features of a VR assistant, which would operate as a visitor guide for such places. We are talking about the use of the local language, the methods of invocation (the use of a specific greeting phrase, also called wake-up words), finding the right location, the form of display, the design and implementation of a dedicated maintenance interface for the staff of the facility, and many others.

In conclusion, Romanian museums can and must adapt their exhibition discourse to the needs of the community in which they operate, constantly taking into account the national cultural strategic priorities, in an integrative and holistic way, prepared for the alternative promotion of the national

cultural heritage, but also of the contemporary creation. Our study contemplates this desiderate and offers a viable option in the context mentioned above.

3. Research Challenges

Building a VR museum guide involves providing answers to a series of challenging scientific questions: "What is an accurate mathematical model of a museum visitor?", "How can the system understand what the visitors are asking in Romanian?", "How can the system advise users to get to a destination from a current position?", "What does a museum exhibit represent? ", "How can the system relate one work of art to another?", or "How can the system be physically implemented?". Solving these challenges would be in the benefit of most Romanian museums and art exhibits in general, in the museum "Casa Mureșenilor" in particular.

The research methodology was built on a set of hypotheses, which are characteristic for the AI and VR applications:

- The virtual agent can be observed by many users, but only one user can interact with the system at any given time
- The virtual agent can be invoked at any given time using a "magic" word
- The virtual agent has a finite set of answers that it can provide to a finite set of questions.

Following these hypotheses, our work targets complementary unknown factors, such as what would be the best type of AI framework that could be used in developing an intelligent virtual assistant, what are the technological limitations posed by the particularities of such an AI algorithm, or how flexible would the final solution really be (could the final system be used in other context - e.g., for public administration?). Building an intelligent virtual assistant implies using state-of-the-art frameworks that each solve partial problems, and integrating them into a simple and unitary solution that can be deployed at the museum and used by its staff without extensive technical training. Our aim is to show a good development approach, to find a good system architecture, and to validate the result by conducting a user study.

4. System Development

4.1. Principle Simulation

After meeting with the representatives of the museum, we needed a quick way to see if we could deliver a system that matched their expectations. In order to demonstrate the feasibility of the project, we prepared a functionality simulation. The simulation aimed to demonstrate the ability to communicate vocally with a device and receive an audio response. As this was an initial test study, we didn't intend to illustrate in detail the functionality of each element of the actual project. The simulation provided an overview of how the final project should behave. For this reason, we didn't include AI aspects, or the transition from voice to text (speech-to-text).

The system developed within this project is a less familiar concept because its completion involves interacting with an intelligent conversational agent who will communicate through an Internet connection with a computing server (as the majority of the frameworks available at the moment require such a connection). Such agents already exist, but this project consists of a procurement module that will contain only the sound sensor, the unit that will host the HTTP server and an Output system. Data processing by the agent can be performed anywhere through the HTTP server endpoints. The simulation was conducted in Simulink. We decided to use any sound that could be interpreted as a human voice, regardless of the conversation's significance. The simulation plays a recorded sound, which is the response to a greeting. This recording will be played when the Input signal is received.

The simulation uses the personal computer's sound system to record and play an audio file to respond to a greeting. Because the detection of words in an audio signal is complex, voice detection was chosen for this simulation, so regardless of the phrase spoken, the answer will be the same. The simulation consists of a set of blocks, each with a well-defined purpose.

The first block used is called "Audio Device Reader" (see Figure 1). It accesses the recording device and starts recording sounds. The block returns a continuous stream of data, representing the captured signal.

The second block, connected after the first, is called "Buffer". Because the first block sends a continuous stream of data to its output, there is, due to the longer processing time than the signal period, the possibility of data loss. We use the "Buffer" module to ensure data integrity, storing the data received from the first block, and forwarding it when needed. It works on the principle of an electric capacitor.

Next, we deployed the "Voice Activity Detector". It detects the presence of a voice in an audio signal. We separate signals that may contain sounds with frequencies that fall within the human voice's frequency band with its help.

The previous block works similarly to the first block: it emits a continuous stream. This is why we used the "Buffer" block following the first block and following the "Voice Activity Detector" block.

We attached a block called "FFT" to the "Buffer". It aims to obtain the spectrum of the received signal after the analysis made by the previous blocks, to obtain further information, useful for framing in voice or noise signals.

After applying the Fast Fourier Transform (FFT), we obtain the signal spectrum. The frequency of the human voice is in the 100Hz–900Hz frequency band. Next, the block called "Bandpass Filter" was chosen. This block filters the obtained signal and identifies frequencies included in the frequency band of the human voice. Filtering aims to eliminate background noise from the sounds identified as voice signals by the "Voice Activity Detector" block.

The next block that provides bandwidth filtering is the block named "IFFT", which reverses the effect of the "FFT" block. This way, we get the signal cleared, and we can decide if it contains a voice.

After filtering, we have a signal that is recognized by the previous blocks as voice. The next block, "Change Detection", detects a change in the input signal. After the input signal is filtered, everything that represents noise in the input signal is eliminated in theory. If the signal contains voice, this is the only thing that remains as a component of it. "Change Detection" detects a signal change only if previous blocks have decided that the signal contains a voice. When a signal change is detected, it outputs a Boolean value. Since we need an integer numeric signal and not a Boolean, the next block is a simple Boolean to integer converter. The TRUE value is converted to 1, and the FALSE value, to 0.

If the signal contains voice, the system will have a signal with a value of 1 at the converter block's output. This signal reaches the decision gate of the "Enabled Subsystem" block. It will forward the "Hello" string. This is actually the text sent by the Node. Js server, after confirming that it was spoken, to the NLU Race server. The NLU Race server returns a text response that will be forwarded to the user. The response mode is transposed in the simulation by playing an audio file.

Next to the "Enabled Subsystem" block is another "Change Detection" block that will detect when the "Hello" string is passed on from the previous modules. "Change Detection" will return a Boolean signal, which in turn must be converted to an integer, which is why a block of "Boolean integer conversion" follows.

The last part of the system consists of an "Enable Subsystem" block that will forward the audio file containing the response to a greeting. The "Enabled Subsystem" block is controlled by detecting the word "Hello" in the previous blocks. Once the word is detected, the last block will send the sound file to the speakers (Output) to respond to the greeting.

The simulation (see Figure 2) shows the correct functionality of the concept in which the human voice can receive an answer from a system integrated into a personal computer. Although the system presented is not one that can respond interactively, the work carried out here proves the concept's feasibility.

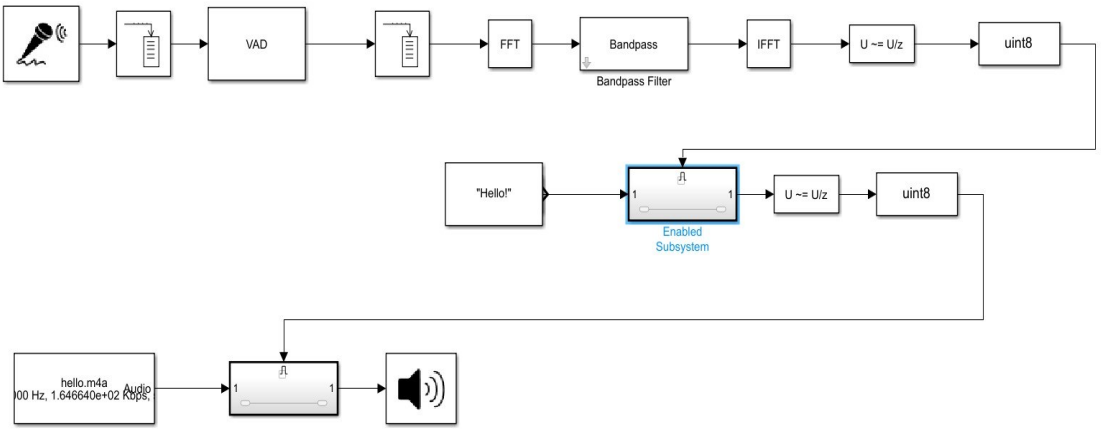

Figure 1. Simlink simulation.

4.2. System Architecture

The IA system architecture is presented in Figure 2. This architecture is based on the simulation described in the previous section and originates from a vast analysis of state-of-the-art solutions available at the moment.

The solution comprises of the following components:

- Google Cloud Speech to Text [32]. This module takes the audio recording from the user and turns it into text. Google offers an API that recognizes 120 languages, including Romanian. This service is based on the machine-learning technology refined by Google over decades of semantical data collection obtained through its other subsidiaries (Search, Translate, and so on) We decided to use Google Cloud Speech To Text after comparing this framework with several others, such as Vonage, Amazon Transcribe, IBM Watson, Microsoft Bing Speech API, AssemblyAI, Kaldi, or Sayint. Google Cloud Speech To Text is highly scalable and allows rapid development.
- The application developed using the NLU (Natural Language Understanding) RASA platform [33]. It takes the text and "understands" it, extracting the keywords and determining their semantical meaning. The answer is then processed and served to the user. The analysis and processing are performed directly on Romanian text, as the RASA NLU platform supports any language for its training pipeline. Any language for its training pipeline. We chose RASA NLU over other natural language processors such as NLTK, SpaCy, TensorFlow, Amazon Comprehend or Google Cloud Natural Language API, as this platform allows a facile re-configuration of the smart agent by easily adding new content/answers/narratives.
- The SitePal service [34]. SitePal offers avatars who can be directly displayed on websites and will play the text obtained by RASA NLU by voice (including lip-sync and facial and lip mimicry during speech). The avatar obtained through the SitePal service includes support for Text-to-Speech in Romanian. The SitePal service allows the creation of both bust and standing avatars. Users can choose from a wide range of models, but avatars can also be made based on specific photos. Moreover, the background can be customized or be left transparent. The only drawback of this service is that at the moment, the text-to-speech service in Romanian is available only with a female voice. Luckily, the representatives of the museum "Casa Mureșenilor" were only interested in this type of voice. We chose SitePal over similar services such as Living Actor or Media Semantics as the representatives of the museum found their avatars to be more appealing visually.

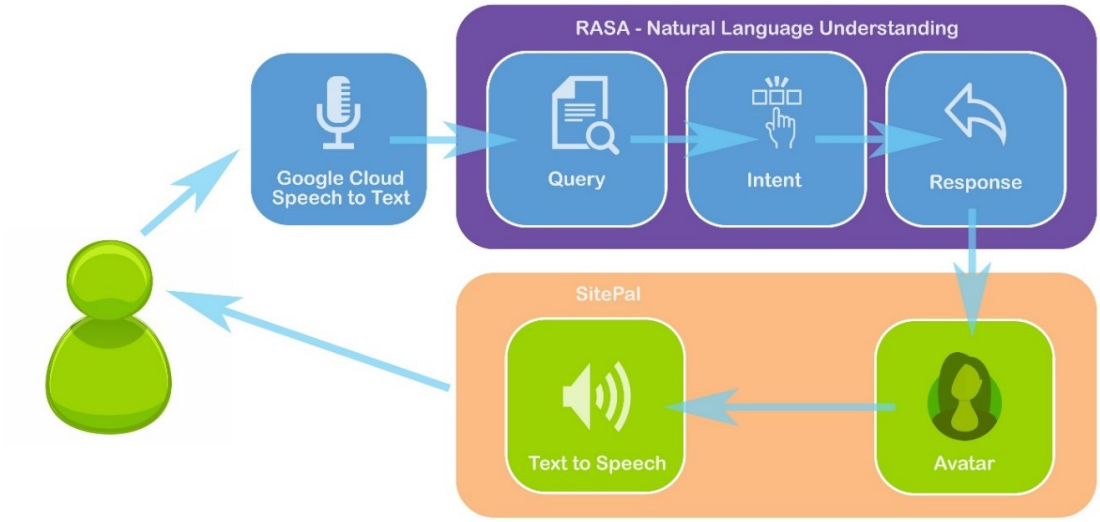

Figure 2. IA System Architecture.

4.3. Natural Language Understanding module

The Natural Language Understanding (NLU) kernel of the system was implemented using RASA NLU, a framework that allows the creation of a new project that implements the functionalities provided by Rasa Technologies.

The training data of the Rasa NLU framework are structured in four different parts:

- common examples
- synonyms (used to map several entities to the same name)
- regex formulas (sequences of characters that can be interpreted in different ways following a set of strict rules)
- lookup tables (used to generate regex syntax) [33].

Common examples are composed of three components: text, purpose, and entities. The text represents the user's message. This is mandatory, being the only date of entry. The purpose is a type of cataloging. Depending on the settings made in this field, our agent will make decisions regarding the data's framing. The last component of the common examples is the entity. This is used to allow multiple words to be placed in a context.

Synonyms are used to map multiple values to a single entity. An example of their use is mapping misspelled words so that they are recognized and associated with their correct form.

Regexes are mainly used for variable text framing. These allow you to sort strings defined according to certain rules. Their implementation in a chatbot is very useful because it offers the possibility to frame a wide range of entries that differ by specific differences in characters. These differences can be systematized and anticipated.

Lookup tables provide a function similar to that of synonyms. They offer the ability to create a list of values that can be stored in separate files. They are useful as the entity assigned to each value within such a table doesn't have to be specified.

All 4 features were used in our RASA NLU server implementation, as presented in Figure 3.

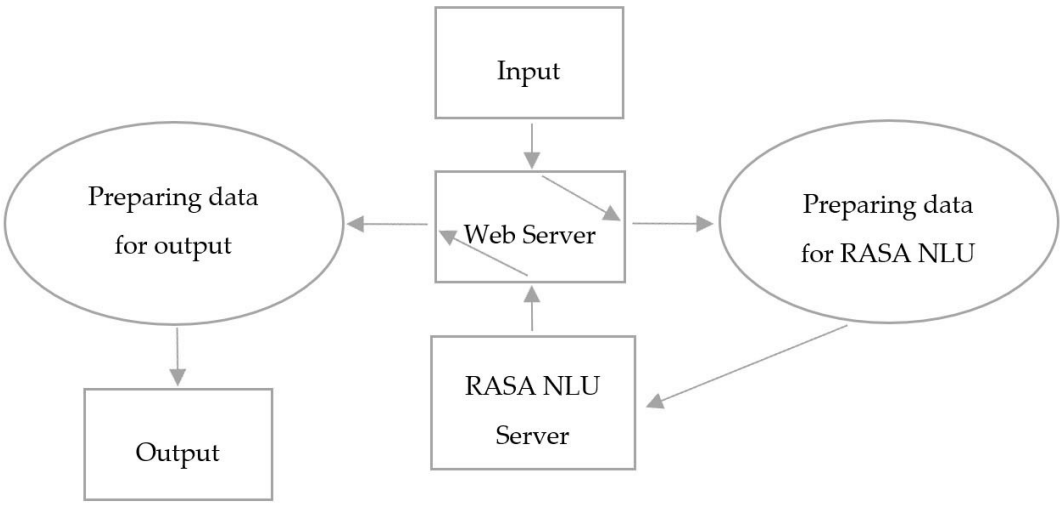

Figure 3. RASA NLU—IA project architecture.

The IA system is made attentive to the user by detecting "hot keyword" (invocation phrase), in our case, "Servus, IA!". For achieving this, we implemented an AI module based on Sonus/Snowboy [35]. Snowboy is a real-time keyword detector. It can analyze the signals provided by the audio sensor continuously without requiring an internet connection. A keyword can be considered a phrase that the computer searches for in the received audio signal. The functionality that Snowboy offers can be found in the word detection systems implemented on Amazon's Alexa assistant, or Google's assistant.

Snowboy offers the possibility of generating personalized "magic" phrases. This functionality is useful in developing an assistant application because it offers the possibility to customize the entire system. Most open-source word detectors provide only their library of developer-driven data.

A neural network was trained using several samples of "Servus, IA!". The Sonus system listens to the invocation phrase continuously. After meeting an invocation phrase, Sonus listens to the user input and connects in real-time to Google Speech to Text, obtaining the text corresponding to the input spoken by the user. That text is sent to the NLU RASA module via a GET request.

AI for NLU was made with RASA. We created an implementation in which we entered the training data (several variants for each question) and the desired answers. This AI core is available as a WEB service that communicates through GET requests with the modules responsible for text-to-speech and avatar control. The RASA NLU server, therefore, receives a text (a question) from Sonus/Snowboy module and provides an answer based on the priory trained AI core. This response is sent by the node.js server that manages the entire system to the frontend (via Socket.IO), more precisely to a JavaScript function that controls the avatar on the Web page through the SitePal JavaScript frontend API. This function commands the avatar to utter the answer received from RASA. The architecture of the AI functionality is presented in Figure 4.

The RASA NLU core was programmed to handle over 300 questions grouped on three main topics according to the topics of interest for the museum visitors: questions about the city of Brasov, about the Museum in general and its exhibits, and about the famous Mureseanu family. Some of the questions are presented in Table 1.

Table 1. Examples of questions that can be answered by IA, clustered into 3 categories.

Brașov	Museum	Mureșeanu Family
Who founded the city?	What is the visiting schedule?	Who were the members of the Mureșeanu family?
Who were the people who contributed to the development of the city?	Are all the pieces in the museum original?	What contribution did the Mureșeanu family have for Brașov?
How was Brasov in the past?	Where was the furniture brought from?	Did they live here?
Are there legends about Brasov?	How old are musical instruments?	Is there an artist in the family?
What events are in the city?	Where is the toilet?	Where was the anthem first sung?
What other cultural objectives are there in the city?	How many employees does the museum have?	

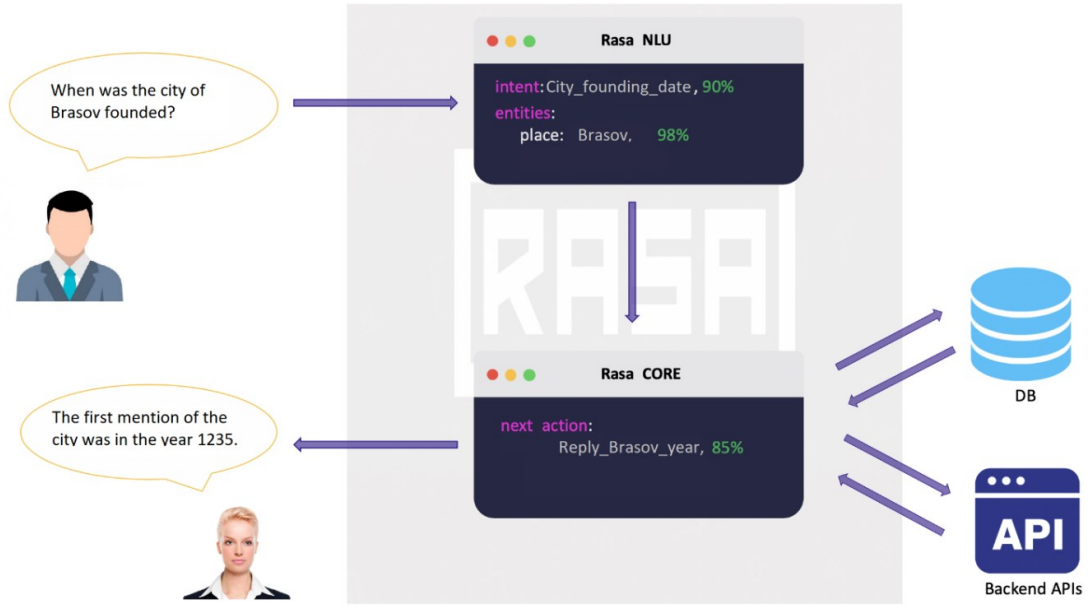

Figure 4. AI functionality with RASA.

4.4. Virtual Avatar

As presented in [36], choosing the right virtual avatar for a digital heritage application can be a daunting task. The avatar used in our system is called Kara, and is available directly on the SitePal platform (see Figure 5).

It was chosen following several meetings with the representatives of the museum, on the account that it offers an out-of-the-box functional 3D human body enhanced with gestures and lip-synching movements, textured, and animated autonomously. The avatar character and scene details (clothing, background) were chosen after several iterations and based on the feedback received from the museum staff.

Figure 5. Kara avatar from SitePal.

However, we have already made plans to improve this avatar by using a real-life model captured using photogrammetry, that can be rigged and animated in 3D development platforms such as Unity or Unreal, or that can be recorded in 3D for a finite set of questions/answers. The difference between these 2 approaches is that in the latter, we can better control the behavior of several features such as eye gaze, gestures, emotions, and personality.

4.5. Physical Stand

The physical stand consists of a large TV screen connected to a computer with internet, an omnidirectional microphone and speakers. For the duration of the experiment, the TV was placed at a height of 1.5 m to match the average height of the visitors of the museum (which were, in our case, mostly children between 13 and 18 years). The TV can be slightly moved up or down to match the height profile of a general user. Behind the flat screen lies a large convex wall banner that lists several questions that can be asked, thus familiarizing visitors with the type of interaction they could expect, and addressing an open invitation to use the system. The idea of using a wall emerged from the early testing stages. Prior to deploying the final system in the museum, we had several visits from the representatives of the museum. The system was also tested by several students from the Faculty of Electrical Engineering and Computer Science. Following these tests, it became clear that visitors seeing the assistant for the first time needed some sort of guidance that would help them with the process of putting questions. The initial experiments featured a handbook which contained all of the questions. However, this was abandoned for the wall, as this provided much better visual cues, and enough space to cluster questions around 3 categories: the city of Brașov, the museum and the family Mureșeanu (see Figure 6).

We have discovered that this wall improves dramatically the entire setup, while also increasing the number of interactions, basically diminishing the "leap of faith" required by any new technology from its potential beneficiaries. At the moment of research and implementation, the estimated costs for using the platforms are $0.006 for 15 s of audio converted to text (Google Cloud Speech to Text) and a $20 monthly subscription to SitePal, which covers 4000 audio streams (spoken phrases).

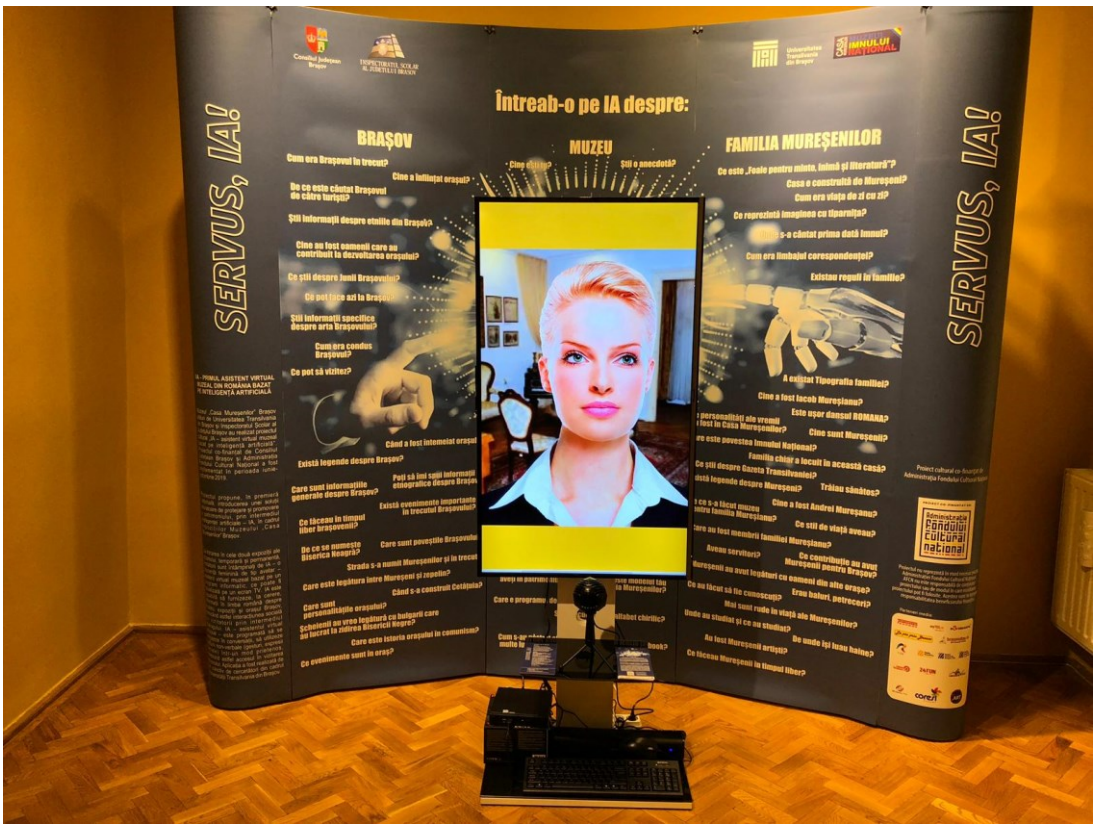

Figure 6. IA physical stand.

5. Ethical Compliance

This research was carried out under the Directive 1995/46/EC on the protection of individuals concerning the processing of personal data and on the free movement of such data, and Directive 2002/58/EC on privacy and electronic communications. The ethical aspects related to all visitors who participated in our study (informed consent, including parental agreement) were solved by the two institutions involved in the process (the School Inspectorate of Brașov County and the Museum "Casa Mureșenilor" from Brașov). All institutions, including our university, approved the user studies carried out under this research.

6. User Study

6.1. User Acceptance Evaluation

Following the partnership between the museum and the School Inspectorate of Brașov County, the visits of over 250 students were scheduled during October 2019, and their opinions were evaluated after the interaction with IA. The virtual assistant is a digital, interactive application, designed with human appearance and skills (verbal and gestural), which allows it to participate in a dynamic social environment. Thus, it was imperative to test the efficiency of this system using a quantitative method.

The questionnaire design is presented in Table 2. The staff of the museum supplied the data that resulted from this questionnaire. The first five questions had only three possible answers (Positive, Negative, or Neutral). Given the age and the specific features of the test group, this form of evaluation was preferred to a Likert Scale, mostly because of its simplicity and accessibility. The average time taken to fill the questionnaire was less than 2 min. The staff of the museum supplied the data that resulted from this questionnaire.

Table 2. Evaluation Questionnaire for IA.

1. Is it useful to have an Artificial Intelligence Guide in a museum?
2. Are the questions that IA can answer appropriate?
3. Are the answers provided by IA appropriate?
4. Is the avatar used by IA appropriate?
5. Do you consider the IA to be a success?
6. Please share any other useful information for this project.

The students included in this test came from a middle school (Johannes Honterus National College) and four high schools ("Andrei Mureșanu" High School, "Maria Baiulescu" Technical College in Brașov, "Johannes Honterus" National College, and "Dr. Ioan Meșotă" National College). Out of the total of the 250 participants, about 60% were on their first visit to the "Casa Mureșenilor" museum, and 40% had visited it before. The age of the surveyed participants was between 13–18 years, and the distribution by gender was 35% male and 65% female. This general data was collected before entering the museum.

To the question "Is it useful to have an Artificial Intelligence Guide in a museum?" 218 (87.2%) of the respondents rated it as good, 30 declared themselves neutral, and two answered negatively.

To the question "Are the questions that IA can answer appropriate?" 201 (80.4%) of the respondents rated it well, 47 declared themselves neutral, and two answered negatively.

To the question "Are the answers provided by IA appropriate?" 194 (77.6%) of the respondents rated it positively, and 56 declared themselves neutral.

To the question "Is the avatar used by IA appropriate?" 183 (73.2%) of the respondents rated it positively, 47 declared themselves neutral, and 20 answered negatively.

To the question "Do you consider the IA to be a success?" 225 (90%) of the respondents rated it as good, 23 said they were neutral, and two answered negatively.

6.2. A Comparison with Other Virtual Agents for Museums or Related Heritage Applications

In order to assess the degree of novelty of the system proposed here, we've constructed a comparison table (Table 3) which takes into consideration all notable initiatives carried out in this research area (to the best of our knowledge). This offers an explanation of how our system differs from the related work, not just in terms of combining some existing technologies in a unique way, but in terms of the new characteristics proposed by this work. The factors that weigh in are Humanoid look (whether the agent looks like a virtual human), Non-English Language (whether the agent can speak in a language different from English - any other foreign language), Conversational Interactivity (whether the agent can reply as in a normal conversation), Natural Gestures (whether the agent can make gestures while talking) and Learning Ability (whether the agent can improve the way it provides answers).

Table 3. Comparison of IA with related work.

Virtual Assistant	Humanoid Look	Non-English Language	Conversational Interactivity	Natural Gestures	Learning Ability
IA	Yes	Yes	Yes	Yes	Yes
Max [4]	Yes	No	Yes	No	No
Ada and Grace [6]	Yes	No	Yes	Yes	No
Coach Mike [8]	Yes	No	Yes	Yes	No
Tinker [9]	No	No	Yes	Yes	No
GEN [10]	Yes	Yes	Yes	No	No
Recommender [15]	No	Yes	No	No	No
CulturalERICA [22]	No	No	Yes	No	Yes

7. Results and Discussion

Figure 7 presents the outcome of the questions presented above in 5 pie charts. One can easily see that feedback was mostly positive in all areas. The only slightly arguable result is related to the appearance of the avatar, where respondents have indicated the negative choice in 8% of the cases, meaning that there is still room for improvement in this chapter.

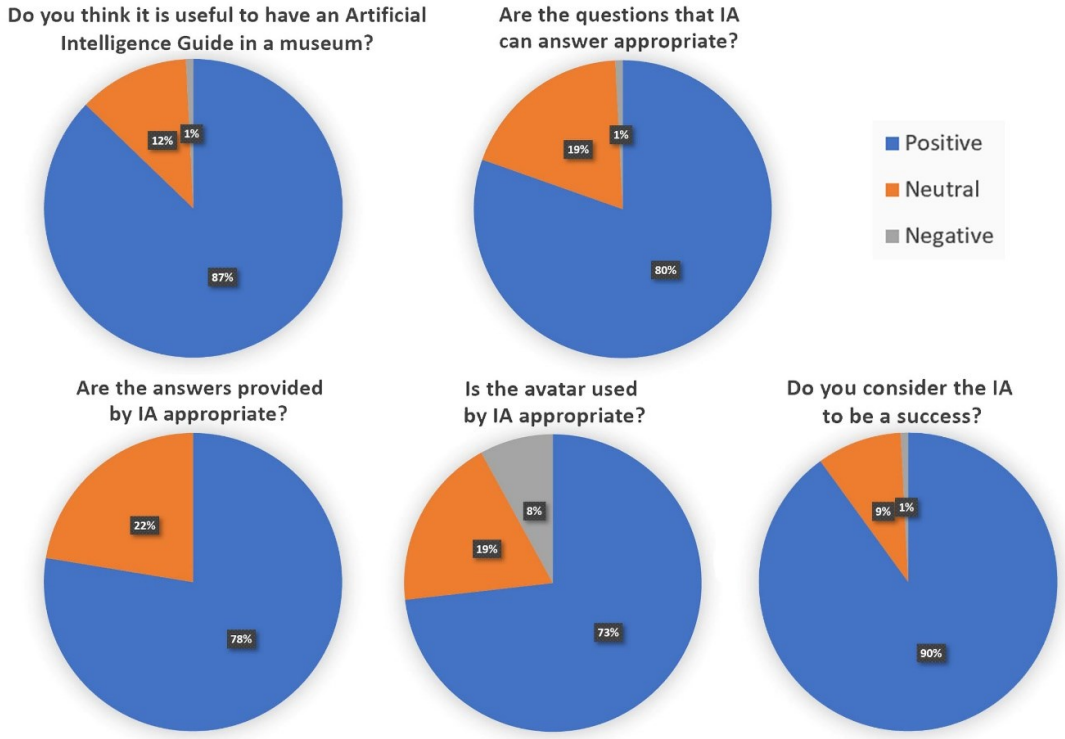

Figure 7. Questionnaire results.

An open question was asked at the end of the research questionnaire. Out of the over 250 respondents, only 45 (25%) answered. Table 4 presents an outline some of the responses registered at this section, structured under three categories: Aspect, Functionality and Miscellaneous.

Table 4. Additional comments about IA.

Category	Comment
Aspect	IA's blinking is dubious. The avatar looks fake.
	Make IA a brunette.
	IA should have long hair.
	IA looks like James Charles.
	IA should look friendlier.
	IA should not move its eyes, it's a little scary.
	You should use real human avatars. Perhaps the image of a Mureșanu family member would be more appropriate.
Functionality	IA should answer more correctly.
	The information provided should be more extensive. It should have a setting for a detailed or short answer.
	IA should know how to sing.
	Occasionally, it does not recognize the activation words.
	IA should be able to answer other questions that are not related to the museum; for example, what love means.
	IA should know how to make jokes.
Miscellaneous	Subjects should be more accessible to children.
	Its name should be changed.

Following the analysis of the research results on the 250 students who benefited from the visit of the temporary exhibition and who interacted with IA, the main conclusion is that the system was positively received by the target audience. Although this conclusion could be expected without conducting proper research, given the way young people interact with new technologies, the results presented in this report also reveal the following:

- AI could be the solution for knowledge transfer, especially in the case of young visitors to museums.
- 3D VR avatars are considered innovative by young audiences. Based on this technology, one can imagine a multitude of developments and related events that translate into a more efficient interpretation and promotion of cultural heritage in museums in Romania.
- The introduction of virtual elements allows to increase the level of interaction of visitors with museum products and services, which can lead to an increase in the number of visitors. Interaction is the most important feature of a successful user-centric system [37].
- Given the pioneering nature of this project at the national level and the relative novelty at the international level, the implementation can be expanded, and evaluations could be conducted among other categories of the visiting public.

8. Conclusions

This study is valuable from the applied research point of view, as we present a system architecture that can be developed fast and that is flexible enough to be used in multiple scenarios. Improving the engagement and the attractiveness of cultural institutions are two crucial objectives, not only from the social and psychological point of view but also economically. A thorough investigation of these issues represents significant progress in the technologies used to digitize cultural heritage. This provides a new perspective with an impact on the daily activities of thousands of people. The development of an intelligent virtual guide for museums represents a dynamic research endeavor that can be immediately applied, with sustainable benefits. The results presented here allow the opening of new research topics in digital cultural heritage and intelligent human-computer interfaces.

IA is based on the concept of using the latest advances in the field of artificial intelligence, 3D applications, image processing, which can be classified as "TRL 2—Technology concept and/or application formulated". Overall, the system developed can be quickly scaled and ready to demonstrate a technology of level "TRL 5—Technology validated in a relevant environment". This shows that market uptake could be very close to the finishing point of the project, translating it into a high impact initiative, culturally and sociologically speaking.

The user study presented in the section above shows that IA was well received. However, this is rather trivial, since new technologies usually hold in intrinsic appeal to the public in general, and the young in particular. In the near future we aim to conduct a new user study on several categories of visitors based on a quantitative and qualitative methodology following more sensitive aspects such as technical challenges, the spectrum of questions, and others. The purpose of a proper evaluation strategy is to differentiate between the natural public interest in new technologies and the real value brought by these. Based on the questionnaire conducted on a reasonably large set of respondents, we infer that IA can answer decently to a good amount of questions. The weakest feature that could be the subject of future developments is the avatar used by the system. Several visitors reported problems related to eye gaze, lip movement, or gestures. All of these can be optimized in future IA versions. Eye gaze could, for example, be set to track persons who are actively involved in the conversation.

The display metaphor is also a subject of change. Instead of viewing the upper body, we could shift the viewing angle to comprise only the head, or by the contrary, the entire body. Symbolic designs can emphasize the artificial nature of the agent, such as using shapeless forms or geometric figures. Other future work could be clustered around using an Augmented Reality (AR) or a holographic display to increase the feeling of presence and improve the entire HCI paradigm of the system. As for the "intelligence" of the system, the museum's staff could expand the database with

thousands of questions. One particularly useful comment received as a feedback from visitors is to allow them to choose among two different forms of answers: small (brief resumes) and large (more descriptive). All these developments can be quickly addressed and implemented, as proof that technologies used are mature enough to support and facilitate fast tuning.

Author Contributions: Conceptualization, M.D. and O.M.M.; methodology, O.M.M.; formal analysis, M.D., V.-A.B. and O.M.M.; investigation, M.D. and O.M.M.; resources, I.-A.D.; writing—original draft preparation, M.D., V.-A.B. and O.M.M.; writing—review and editing, M.D., V.-A.B., I.-A.D. and O.M.M.; supervision, M.D.; project administration, M.D. and O.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This project was co-funded by the Administration of the National Cultural Fund and the Brasov City Council during June–October, 2019.

Acknowledgments: We thank the Museum of Casa Mureșenilor for the continuous assistance and support throughout this study.

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

References

1. Chodzko-Zajko, W.J.; Proctor, D.N.; Singh, M.A.F.; Minson, C.T.; Nigg, C.R.; Salem, G.J.; Skinner, J.S. Exercise and physical activity for older adults. *Med. Sci. Sports Exerc.* **2009**, *41*, 1510–1530.
2. Gordon, J.C.; Shahid, K. Tailoring User Interface Presentations Based on User State. U.S. Patent No. 10,552,183, 4 February 2020.
3. Tan, B.K.; Hafizur, R. CAAD futures. *Virtual Heritage: Reality and Criticism*; De l'Université de Montréal: Montreal, Canada, 2009.
4. Kopp, S.; Gesellensetter, L.; Kramer, N.C.; Wachsmuth, I. A conversational agent as museum guide–Design and evaluation of a real-world application. In *International Workshop on Intelligent Virtual Agents*; Springer: Berlin, Germany, 2005; pp. 329–343.
5. Robinson, S.; Traum, D.R.; Ittycheriah, M.; Henderer, J. *What Would you Ask a Conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting*; LREC: Marseille, France, 2008.
6. Swartout, W.; Traum, D.; Artstein, R.; Noren, D.; Debevec, P.; Bronnenkant, K.; Williams, J.; Leuski, A.; Narayanan, S.; Piepol, D. Ada and Grace: Toward realistic and engaging virtual museum guides. In *International Conference on Intelligent Virtual Agents*; Springer: Berlin, Germany, 2010.
7. Chaves, A.P.; Aurelio Gerosa, M. Single or multiple conversational agents? An interactional coherence comparison. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018.
8. Chad Lane, H.; Noren, D.; Auerbach, D.; Birch, M.; Swartout, W. Intelligent tutoring goes to the museum in the big city: A pedagogical agent for informal science education. In *International Conference on Artificial Intelligence in Education*; Springer, Berlin, Germany, 2011.
9. Bickmore, T.W.; Laura, M.; Vardoulakis, P.; Schulman, D. Tinker: A relational agent museum guide. *Auton. Agents Multi-agent Syst.* **2013**, *27*, 254–276.
10. SJM Tech. GEN Project. Available online: http://www.sjmtech.net/portfolio/gen/ (accessed on 2 October 2019).
11. Mollaret, C.; Mekonnen, A.A.; Ferrané, I.; Pinquier, J.; Lerasle, F. Perceiving user's intention-for-interaction: A probabilistic multimodal data fusion scheme. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy, 29 June–3 July 2015.
12. Ghazanfar, A.; Le, H.Q.; Kim, J.; Hwang, S.W.; Hwang, J.I. Design of seamless multi-modal interaction framework for intelligent virtual agents in wearable mixed reality environment. In Proceedings of the 32nd International Conference on Computer Animation and Social Agents, Paris, France, 1–3 July 2019.
13. Schaffer, S.; Gustke, O.; Oldemeier, J.; Reithinger, N. *Towards Chatbots in the Museum*; mobileCH@ Mobile HCI: 2018. Barcelona, Spain.
14. Pavlidis, G. Towards a Novel User Satisfaction Modelling for Museum Visit Recommender Systems. In *International Conference on VR Technologies in Cultural Heritage*; Springer: Berlin, Germany, 2018.
15. Tavcar, A.; Antonya, C.; Butila, E.V. Recommender system for virtual assistant supported museum tours. *Informatica* **2016**, *40*, 279.
16. Longo, F.; Nicoletti, L.; Padovano, A. An interactive, interoperable and ubiquitous mixed reality application for a smart learning experience. *Int. J. Simul. Process Modell.* **2018**, *13*, 589–603.

17. Doyle, P.R.; Edwards, J.; Dumbleton, O.; Clark, L.; Cowan, B.R. Mapping perceptions of humanness in speech-based intelligent personal assistant interaction. *arXiv* **2019**, arXiv:1907.11585.
18. Rosales, R.; Castañón-Puga, M.; Lara-Rosano, F.; Flores-Parra, J.M.; Evans, R.; Osuna-Millan, N.; Gaxiola-Pacheco, C. Modelling the interaction levels in HCI using an intelligent hybrid system with interactive agents: A case study of an interactive museum exhibition module in Mexico. *Appl. Sci.* **2018**, *8*, 446.
19. Becker, C.; Kopp, S.; Wachsmuth, I. Why emotions should be integrated into conversational agents. *Conversat. Inf. Eng. Approach* **2007**, *2*, 49–68.
20. Sylaiou, S.; Kasapakis, V.; Gavalas, D.; Dzardanova, E. Avatars as storytellers: Affective narratives in virtual museums. *Pers. Ubiquitous Comp.* **2020**, doi:10.1007/s00779-019-01358-2.1–13.
21. Carrozzino, M.; Colombo, M.; Tecchia, F.; Evangelista, C.; Bergamasco, M. Comparing Different Storytelling Approaches for Virtual Guides in Digital Immersive Museums. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*; Springer: Berlin, Germany, 2018.
22. Machidon, O.M.; Tavčar, A.; Gams, M.; Duguleană, M. CulturalERICA: A conversational agent improving the exploration of European cultural heritage. *J. Cult. Herit.* **2020**, *41*, 152–165.
23. Amato, F.; Moscato, F.; Moscato, V.; Pascale, F.; Picariello, A. An Agent-Based approach for recommending cultural tours. *Pattern Recognit. Lett.* **2020**, *131*, 341–347.
24. Castiglione, A.; Colace, F.; Moscato, V.; Palmieri, F. CHIS: A big data infrastructure to manage digital cultural items. *Future Gener. Comp. Syst.* **2018**, *86*, 1134–1145.
25. Amato, F.; Chianese, A.; Moscato, V.; Picariello, A.; Sperli, G. SNOPS: A Smart Environment for Cultural Heritage Applications. In Proceedings of the Twelfth International Workshop on Web Information and Data Management, Maui, HI, USA, 29 October–2 November 2012; Association for Computing Machinery: New York, NY, USA, 2012.
26. Toderean, G.; Ovidiu, B.; Balogh, A. Text-to-speech systems for romanian language. In Proceedings of the 9th International Conference "Microelectronics and Computer Science " & The 6th Conference of Physicists of Moldova, Chișinău, Republic of Moldova, 19–21 October 2017.
27. Tiberiu, B.; Daniel Dumitrescu, S.; Pais, V. Tools and resources for Romanian text-to-speech and speech-to-text applications. *arXiv* **2018**, arXiv:1802.05583.
28. Nechita, F.; Demeter, R.; Briciu, V.A.; Kavoura, A.; Varelas, S. Analysing Projected destination images versus visitor-generated visual content in Brasov, Transylvania. In *Strategic Innovative Marketing and Tourism*; Kavoura, A., Kefallonitis, E., Giovanis, A., Eds.; Springer: Cham, Germany, 2019; pp. 613–622, doi:10.1007/978-3-030-12453-3_70.
29. Candrea, A.N.; Ispas, A. Promoting tourist destinations through sport events. The case of Brasov. *J. Tour.* **2010**, *10*, 61–67.
30. Candrea, A.N.; Constantin, C.; Ispas, A. Tourism market heterogeneity in Romanian urban destinations, the case of Brasov. *Tour Hosp. Manag.* **2012**, *18*, 55–68. Available online: https://hrcak.srce.hr/83822 (accessed on 12 February 2020).
31. eHERITAGE Project. Available online: http://www.eheritage.org/ (accessed on 20 April 2020).
32. Google Cloud Speech to Text. Available online: https://cloud.google.com/speech-to-text/ (accessed on 20 April 2020).
33. RASA NLU. Available online: https://www.rasa.com/docs/nlu/ (accessed on 20 April 2020).
34. SitePal. Available online: https://www.sitepal.com/ (accessed on 20 April 2020).
35. Snowboy Hotword Detection Engine. Available online: https://snowboy.kitt.ai/ (accessed on 20 April 2020).
36. Machidon, O.M.; Duguleana, M.; Carrozzino, M. Virtual humans in cultural heritage ICT applications: A review. *J. Cult. Herit.* **2018**, *33*, 249–260.
37. Carrozzino, M.; Voinea, G.D.; Duguleană, M.; Boboc, R.G.; Bergamasco, M. Comparing innovative XR systems in cultural heritage. A case study. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *1*, 373–378.