Intentions, Meanings, and Whys: Designing Content for Voice-based Conversational Museum Guides

Fabricio Barth
IBM Data & AI IBM Cloud
fbarth@br.ibm.com

Heloisa Candello
IBM Research
hcandello@br.ibm.com

Paulo Cavalin
IBM Research
pcavalin@br.ibm.com

Claudio Pinhanez
IBM Research
csantosp@br.ibm.com

ABSTRACT

If an artwork could talk, what would visitors ask? This paper explores what types of content voice-based AI conversational systems should have to attend visitors' expectations in a museum. The study analyses 142,463 conversation logs from 5,242 unique sessions of a nine-month long deployment of a voice-based interactive guide in a modern art museum in Brazil. In this experiment, visitors freely asked questions about seven different artworks of different styles. By grouping the visitor utterances into eight types of content, we determined that more than half of the visitors asked about the *meanings and intentions* behind the artwork, followed by *facts* about the artwork and *author*-related questions. We also determined that the types of questions were not affected by each artwork, the artwork style, or its physical location. We also saw some relationships between the visitor's overall evaluation of the experience with the types of questions she asked. Based on those results, we identified implications for designing content for voice-based conversational systems in museums.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; User studies.

KEYWORDS

Conversational interfaces, Content curation, User studies

1 INTRODUCTION

Museum and galleries aim to enhance connections between visitors and exhibits. Nowadays, visitors bring their own devices and not always engage with the art, a phenomenon called "heads down",

Figure 1: Visitors interacting with an artwork using the speech-based system "The Voice of Art" in the *Pinacoteca* museum.

looking at their screens and typing on their phones. Also, in some situations, art may seem unapproachable, through "do not touch the objects" or by not being accessible for everyone.

What are the ways art can connect with visitors in a more approachable way, using the same technologies? Several cultural and scientific institutions have adopted technologies to connect beyond the labels displayed next to the artworks. These include: chatbots [5], robots [24], QR codes [22], RFID tags [13], and augmented reality [4, 28]. This paper focuses on the use of conversational chatbots and, more specifically, on how to design their content. We approach the problem by studying a deployment of a voice-based conversational system in a modern art museum in Brazil where visitors can ask questions directly to the artworks.

Traditionally, conversation promotes a space for learning in museums: visitors engage with museum content and develop conversations with the exhibits [17]. To understand how to introduce and design conversations with chatbots in such contexts, researchers should know the purposes and intents of the design as well as the purposes and experience of visitors. Similarly, the content available to the visitors plays a central role in the experience of informal learning generated by exhibits.

In this work, we study the main content topics visitors asked to artworks when using a conversation system, and found that questions related to the intentions of the authors and the meanings of the artworks and their constituent elements play a central, fundamental
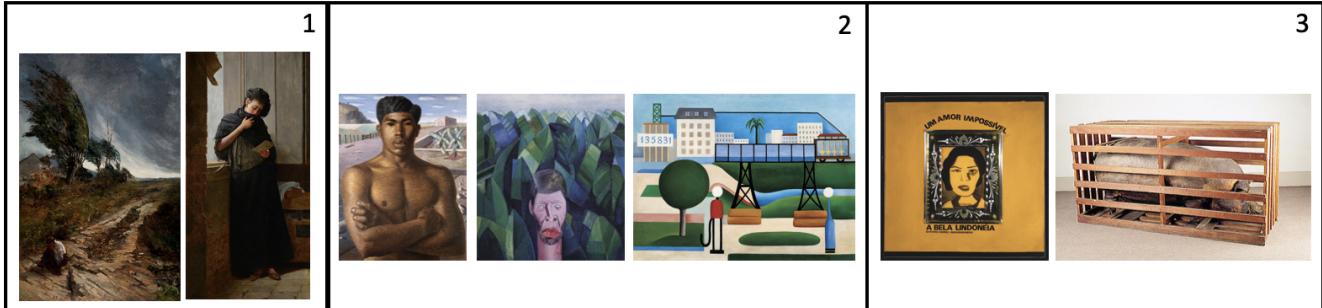
Figure 2: The artworks of "The Voice of Art" grouped by artistic movement styles: (1) Romanticism (*Ventania*; *Saudade*); (2) Modernism (*O Mestiço*; *Bananal*; *São Paulo*) (3) Brazilian Pop-art (*Lindonéia*, *O Porco*).

role. We also found relationships between the types of questions asked by visitors and the overall satisfaction of the visitor experience.

We consider here the conversation logs of 142,463 visitors who interacted with a voice-based conversational system called "The Voice of Art" which was used by visitors to explore seveb artworks of the *Pinacoteca Museum* in Brazil (see fig. 2). We also use the feedback visitors gave of their overall experience when answering a survey, based on *5-point Likert scale*-questions, available to them before returning the conversational device.

The visitors' voice was transcribed by the system in real time using *Speech to Text (STT)* technology; then the text was analyzed and mapped into 388 different *intents*, which triggered the system to reproduce a predefined answer associated to each intent. The conversational platform, *IBM Watson Assistant*, uses the paradigm of most of the conversational systems platforms built today, an *intent-action* approach [23]. The system is created by defining the intents, providing a basic set of user example questions, and associating the systems' responses which should match them. The term *intent* is adopted to describe the goal of a single group of example questions, so the essential task of the conversational platform is to identify the intent of a given question written or spoken by the user, and then output its associated answer or action. Each of the questions asked by visitors, and the determined intents, were logged.

For the study, we clustered the 388 different *intents* associated to the visitors questions into eight major *content types*: *fact, author, visual, style, context, meaning, play, outside* following a methodology we describe in detail later. The 7 artworks available in the "The Voice of Art" app belong to three different art movements: Romanticism, Modernism, and Brazilian Pop-art (fig. 2), and where displayed in 4 different areas of the museum. We analyzed the conversation logs to understand the popular topics of interest of the visitors and whether the distribution of content types varied according to the artwork (1), the artwork movement style (2), the visitor evaluation of the system (3), and the artwork location (4). Those lead to the five research questions this work aims to answer:

[RQ1] What kind of content does people ask to artworks using conversational AI systems?
[RQ2] Do different artwork styles affect the distribution of content type asked using conversational AI systems?

[RQ3] Do different artworks induce questions with the same content type distribution when visitors are using conversational AI systems?
[RQ4] Does the physical location of the artworks affect the distribution of content type distribution among visitors?
[RQ5] Does the visitor experience perception change according to content type when visitors are using conversational AI systems?

By focusing on those questions, we hope to draw implications to assist curators and exhibition designers to choose suitable types of content when designing voice-based conversational systems for museums. We start the paper describing related work in voice-based conversational systems in museums. Next, we provide an overview of how "The Voice of Art" app works, and then we describe our experimental procedures, followed by the analysis of the results as we try to answer our research questions. We finalize by offering a discussion, proposing a set of design implications, and exploring further work.

2 CONVERSATIONAL SYSTEMS IN MUSEUMS

Despite technological advances in conversational systems, the majority of museums still offer only traditional audio-guides to visitors with the expectation to provide learning and improved user experience. A review of the low usage of audio-guides unveiled that visitors have a high probability of dropping usage of their audio guide in the middle of the tour and that only a small number of visitors use them to the end [16]. That research also recommended that next generation audio guides should be comfortable to wear and provide convenient and personalized services, since the uniformly designed content and discomfort of using the devices seem to cause the low adoption rate for the audio guides studied.

At the same time, some cultural heritage places and museums are starting to adopt interactive conversational systems (e.g., such as chatbots, virtual agents, and conversational robots) to engage with the public and promote learning. Some cultural heritage locals have used devices with text-based interaction capabilities [5, 21, 26], others have used voice-based systems [3, 14, 20, 25], as well as hybrid interactions with text and audio features [7] were made available to the public.

Figure 3: One of the "The Voice of Art" screens.

Conversational systems have sometimes been employed to trigger more complex questions and to promote conversations among users. In [21], a chatbot was used as a medium to evoke emotional engagement, provocation, and transformation in archaeological sites. System-initiative chatbots [29], in which the chatbot drives the exploration of a theme instead of the visitor, have also been part of the *IRIS+* exhibition in the *Museum of Tomorrow* in Rio de Janeiro [8]. The IRIS+ aimed to inspire visitors to think about their role in society and truly participate in the search for more awareness, tolerance, and a sustainable tomorrow for society.

Conversational systems also are in the shape of interactive virtual agents, portraying famous personalities or characters [7, 15, 20, 25]. In those cases, agents interact directly with the museum visitors who usually address the virtual agent directly using second-person pronouns (you, yours, and yourself). In other cases, interactive virtual agents act like a museum guide giving information about the museum and its exhibitions [14]. In those cases, the interaction with the exhibits is not direct, and information and visitors expectations are filtered through what the curators expect them to ask. It is the case, for instance, of the Anne Frank house chatbot in Amsterdam [26]. Similarly, *Living Artworks in Museums* [3] shows the idea of a proactive chatbot questioning children next to an artwork. The chatbot follows a script and provides factual and sensorial information.

Games and treasure hunts in museums and cultural heritage sites, supported by mobile technology for more than two decades [12], still attract visitors, often providing immersive experiences with the exhibition content. Museums such as *Maxxi* [26] provide visitors a chatbot app which suggests options of trails with artwork quizzes, and visitors can collect coins as they complete each quiz.

Nevertheless, in all those works, how content was chosen by curators and the visitors' expectations about it are not well described and studied, especially for the design of the conversational systems. In fact, we did not find any clear guidance or a framework on how to design content for interactive conversational systems such as "The Voice of Art" in traditional museums. This paper aims to assist in selecting the content expected to engage visitors in cultural heritage places and art and science museums, filling in this gap in the literature. We start by describing in the next section, the app "The Voice of Art" used in our studies and follow by detailing our experimental procedures and results.

3 THE VOICE OF ART

"The Voice of Art" is an artificial intelligence voice-based interactive guide which allows visitors to ask questions to artworks in the Pinacoteca Museum in Brazil. Pinacoteca is a very popular contemporary art museum [1], receiving around 2,000 visitors a day. "The Voice of Art" was available for public use for nine months in the museum in 2017 and more than five thousand people interacted with the app in this period.

The interactive experience was voice-based and used a user-initiative approach; that is, the system did not ask questions to visitors [29]. At the entrance, visitors received a smartphone and a headset with the "The Voice of Art" installed. In the app, visitors had a map of the floor and could choose one of seven artworks; using a microphone icon they could ask questions to that artwork (see 3). All users had to sign a consent term when given the device by the staff of the museum. In this term, visitors gave the right to use their data, anonymized for any future use for the museum, IBM, and research initiatives.

To gather content to the app, an internal website was created where invited collaborators could type questions to the seven artwork pictures. Overall more than 1,750 questions were collected and organized in 326 intents, thus enabling the training of the system. The *IBM Watson Assistant API* was employed to detect the intent and retrieve a pre-stored artwork response created by the museum curators. The curatorial team at the museum wrote the answers, and developers added those to the *IBM Watson Assistant* system.

The recognition of user intents (questions to artworks) was trained for six months before the exhibition's start, collecting questions from potential visitors and volunteers. A supervised learning process was adopted during the exhibition period to avoid the system from learning mistakes and cleaning possible data that could confound the system. At the end of the exhibition period, the system had 388 intents and approximately 3,150 questions.

Other *IBM Watson API services* were also used to generate the narrated audio of the answers. Next to the artwork, location beacons were placed for each artwork. The beacons were recognized by the smartphones within 5 meters of the artworks and activated an introductory message inviting users to get closer and ask questions. Figure 3 shows a typical "The Voice of Art" screen and figure 4 shows an example of a real session of a visitor with the system, translated by the authors from the original dialogue in the Brazilian Portuguese language. At the end of the experience, visitors answered a satisfaction evaluation with a rating scale from 1 (low) to 5 (high).

4 EXPERIMENTAL PROCEDURES

We explore now how our research questions, what kind of topics visitors want to ask artworks, and whether the distribution of those topics is different according to the artwork (1), art movement style (2), feedback rating (3), and artwork location (4). "The Voice of Art" logged the visitors' interactions from April 05, 2017, when the exhibition started, to the last day, December 31, 2017. This dataset contains 5,242 unique sessions and 142,463 questions for the seven different artworks. A session is the collection of all utterances made by a unique visitor from picking up the device to returning it. The data was made available to the authors by the developers of "The Voice of Art," and no personal information was included

[visit starts]
[visitor in front of Lindonéia artwork]

01. **Visitor** [meaning]: who is Linda Léa
02. Guide: Lindonéia is a fictionous character, created by the author's imagination.
03. **Visitor** [meaning]: what is the color of this yellow
04. Guide: The planar area of the yellow color and the other colors associate the image to the advertisement language, a feature very used in Pop Art.
05. **Visitor** [meaning]: what is the meaning of the picture
06. Guide: We can imagine that the author wanted to criticize the military regime [...].
07. **Visitor** [fact]: who is the author of this artwork
08. Guide: The author of this artwork is Rubens Gerchman.
09. **Visitor** [author]: what is the history of the author
10. Guide: Rubens Gerchman was born in 1942 in the city of Rio de Janeiro. He studied [...]
11. **Visitor** [meaning]: what is the meaning of the brown in the picture
12. Guide: The planar area of the yellow color and the other colors associate the image to the advertisement language, a feature very used in Pop Art.
13. **Visitor** [meaning]: what is the meaning of an impossible love
14. Guide: We do not know who loved Lindonéia or who she loved by observing [...].
15. **Visitor** [meaning]: because is beautiful, yeah, died at 18 years-old

[visitor in front of O Porco (The Pig) artwork]

16. **Visitor** [outside]: is this a real pig?
17. Guide: This is a real pig! It was stuffed and put inside a wooden crate.
18. **Visitor [**meaning]: what does this artwork represent
19. Guide: Originally, the pig had a chain tied to its neck, and it had in the other extreme a smoked shank. [...] Afterall, who defines what is art?
20. **Visitor** [meaning]: why the pig is jailed
21. Guide: Nelson Leirner bought the stuffed animal and put it in a wooden crate as part of the artistic series Matter and Form.
22. **Visitor [**author]: what is the history of the author
23. Guide: Nelson Leirner was born in São Paulo in 1932 and comes from a family of [...].
24. **Visitor** [meaning]: it may feel sad in jail
25. Guide: That is hard to know. But nobody would be happy inside such a tigh crate.
26. **Visitor** [meaning]: he is a fat pig
27. Guide: A live pig, adult, may reach a weight up to 500 kg. But this one weights around 25 kg because it is filled with expanded polyurethane, a very light material.
28. **Visitor** [author]: what is the history of the author of the work
29. Guide: Nelson Leirner was born in São Paulo in 1932 and comes from a family of [...].
30. **Visitor** [meaning]: is it a male pig female pig
31. Guide: This is a male pig.
32. **Visitor** [meaning]: why the author chose a pig
33. Guide: Nelson Leirner was one of the founders of the Rex Group, which created polemics around the relationship between art and the art market. Since 1961 [...]

[visitor in front of São Paulo artwork]
[visit continues]

Figure 4: Real example of visitor session, including labels for the content types of her utterances (translated from original in Portuguese).

in the data. A speech-to-text API processed the visitors' voices in real-time and the data only contained the transcriptions generated by the system.

To start, we performed standard statistics on this data considering how many artworks each visitor made questions to in her session, the distribution of the duration of the sessions, and the distribution of the total number of questions per session. On average, the session duration was 104 minutes with visitors asking an average of 27.2 questions and exploring 4.1 artworks. The sessions tended to be long, and 50% of the users explored 4 or more artworks, and 21.4% of them explored all seven artworks. Of all sessions, 50% of them had a duration of at least 81 minutes and 50% of the visitors asked at least 18 questions.

To better analyze this data, we consolidated this data into a dataset with 11 attributes and 142,463 observations which correspond to every utterance posed by a visitor to the "The Voice of Art". The attributes in this dataset are (i) session id, a unique identifier for each session; (ii) the date and time when the session was started; (iii) the date and time when the visitor asked each question (iv) the title of the artwork the question referred to (v) the input text resulting from a spoken question transcribed using the speech to text component; (vi) the intent recognized by the conversational system based on the input text; (vii) the answer given by the app; (viii) the rating (from 1 to 5 starts, 5 being the best rate) provided optionally by the visitor at the end of the tour; (ix) the date and time when the session was finished; and (x) the total duration in minutes of the session.

4.1 Categorization of Visitor Utterances

To simplify the analysis of this data, we decided to group the identified intents of the visitor utterances which covered similar aspects of the artwork, such as whether the question was about the author, facts, or the historical context. We explored different ways to group the intents into meaningful groups, such as the terms used to describe artworks for museum viewers [9], or used in the design of audio guides for museums [11]. However, the kind of questions the visitors asked included issues related to the meaning of the artwork, the intention of the author, and even playful content.

We found a better guidance to define the 8 categories described next in a text advising how students should write essays about artwork [19]. Based on that text, we decided to use the following eight *content types* to classify the 388 different intents related to the seven artworks:

- **fact:** questions related to who is the author, when it was made, its size, or where it has been exhibited;
- **author:** visitor utterances about the author's life, which art movement she was part of, or stylistic influences;
- **visual:** questions about colors and materials used, brushing techniques, etc.;
- **style:** questions about the style of the artwork, which school it belonged to and its characteristics, or artworks with style.
- **context:** inquiries about the historical, political, or social context where the artwork was produced;
- **meaning:** questions related to intentions, meanings, or whys, and the stories possibly behind the people and elements depicted in the artwork;

- **play:** utterances of playful engagement with the artwork, asking questions beyond the scope of the work, such as which soccer a character roots for;
- **outside:** groups questions related to the conversational guide itself, its technology, or unrecognized utterances.

In particular, we found useful, in the analysis of this data, the distinction between stylistic, contextual, and meanings of the artwork, which provided, in our view, a suitable framework to distinguish among the kind of questions the visitors asked.

We ran a blind manual classification with four judges. The four judges were the authors of this paper. Of all intents, 255 (65.72%) had full agreement by the four judges; 84 (21.90%) were classified under the same category by three judges and were assigned by the majority vote. The remaining 48 intents, which had significant disagreement among the judges, were classified in a conciliation session with the four judges. In the end, the four judges agreed in the categories of all the 388 intents.

Figure 4 shows part of a real session with a visitor asking questions about two of the *pop art* artworks. Due to space constraints, some of the answers by the conversational system were shortened in the figure. We include the content type of each question (in brackets), which was computed by mapping the *intent* of the question, as determined in real-time by the system, into the final, mutually-agreed categorization of content types. Some questions were not correctly mapped in real time by the system (such as the one about the brown color of the first artwork), but nonetheless the visitor continued to work with the system. Notice also that many visitor questions in fig. 4 are categorized as *meaning*. As we will see in the next section, that was fairly common.

5 RESULTS

In this section, we address the five research questions based on the dataset described before and the distributions of the eight content type categories described. *Chi-square* and *Contingency Coefficient* tests were used to perform categorical tests. A *Python Notebook*[1] with the package *Pandas*[2] was used to pre-process and run the descriptive analyses. The package *scipy.stats.contingency*[3] was used to execute the Chi-square and Contigency Coefficient tests. The categorical tests were performed to address questions [RQ2], [RQ3], and [RQ5]. Linear regression was executed using the package *statsmodels.api*[4] to address [RQ4].

5.1 Distribution of Content Types in the Artworks

Considering all the 142,463 user utterances, and their computed intents, we computed their distribution according to the content types. The results are depicted as the black bars in fig. 5, with the corresponding percentages as labels. With those results, we are ready to address [RQ1]:

[RQ1] What kind of content does people ask to artworks using conversational AI systems?

As shown in fig. 5, we found that more than the absolute majority (about 60%) of all questions from visitors were related to the
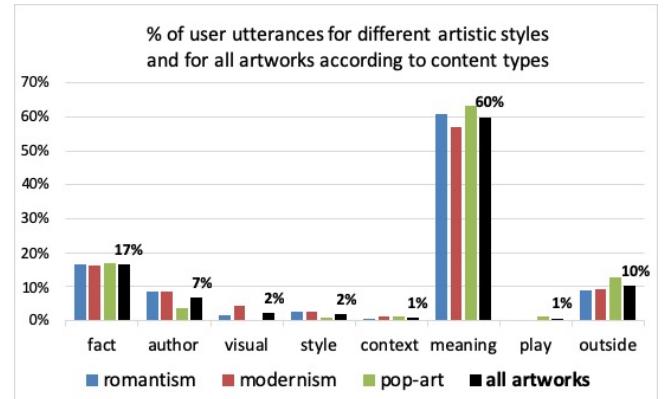
Figure 5: Distribution of user utterances for different artistic styles and for all artworks according to content types.

meaning of the artwork. It is followed by factual questions about the artwork (17%), and about the author's biography (7%). About 10% of the questions were not understood or were outside the scope of the artwork. The other 4 content types, together, corresponded to under 7% of the questions.

5.2 Distribution of Content Types by Style

Figure 5 also shows the distributions of the content types according to the artistic style of the artworks, what addresses our second research question:

[RQ2] Do different artwork styles affect the distribution of content type asked using conversational AI systems?

As we can see in fig. 5, the distributions of content styles by each of the artwork style, Romanticism, Modernism, or Pop-art, seems to be reasonably similar. Indeed, we performed a Chi-Square test and a contingency coefficient test to determine if artwork styles and content types are independent or not. We found no association between artwork style and content type (χ^2=5,397.49, N=142,463 C=0.19). This Chi-Square test was applied considering a contingency table created from all questions in our dataset.

Those results strongly suggest that there is no significant correlation between artwork style and the content type of the visitor questions; therefore, the answer to **[RQ2] is NO**.

5.3 Distribution of Content Types by Artwork

Figure 6 shows the distribution of the content types of the visitor questions for each of the seven artworks, allowing us to investigate our third question:

[RQ3] Do different artworks induce questions with the same content type distribution when visitors are using conversational AI systems?

Figure 6 shows little variation in the distribution of content styles according to the artwork, albeit some variations. For instance, the *meaning* category fluctuates from about 52% (in the *Bananal* piece) to the maximum of 72% in the case of the narrative *Lindonéia* piece.

To test [RQ3], we performed a Chi-Square test and a contingency coefficient test to determine whether the artwork and content type
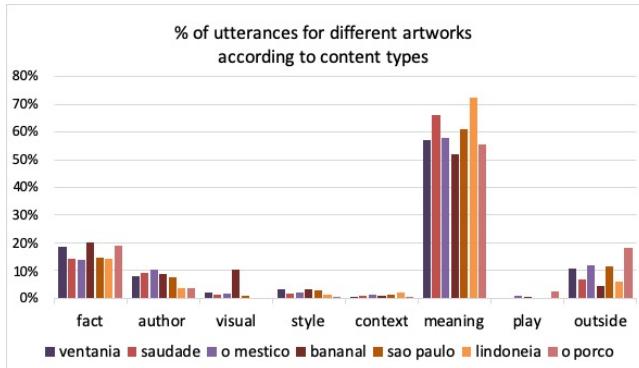
Figure 6: Distribution of user utterances for different artworks in the exhibition according to content types.

Figure 7: Distribution of user utterances for different levels of visitor experience perception (self-reported) according to content types.

are independent. We found a modest association between artwork and content type (χ^2=16,560.079, N=142,463 C=0.32). This Chi-Square test was applied considering a contingency table created from all questions in our dataset. Since the results suggest no significant correlation between the artwork and content type, the answer to **[RQ3] is NO**.

5.4 Distribution of Content Types by Location

Next, we address the influence of the physical location of the artwork.

[RQ4] Does the physical location of the artworks affect the distribution of content type distribution among visitors?

The seven artworks were placed in 4 different locations, grouped by similar artistic style, except that *o mestico* was not together with the other modernist artworks. The distributions of content types were thus very similar to the ones shown in fig. 5, with no apparent differences. We run statistical tests identical to the ones used for artwork style and found no evidence of significant differences. Therefore, the answer to **[RQ4] is NO**.

5.5 Distribution of Content Types by Experience

To understand whether the visitor' experience was related to types of content the visitor asked and the answers she received, we compared the distributions of content types according to the self-reported evaluations of visitors who provided it at the end of the session. We grouped the 1 and 2 ratings as *poor*, the 4 and 5 ratings as *good*, and considered the 3 ratings as *neutral*. Considering those groupings, we then address

[RQ5] Does the visitor experience perception change according to content type when visitors are using conversational AI systems?

First, it is important to consider that, in general, the visitors evaluated their experience quite well. Of all the 4,051 visitors (73% of all) who provided a final evaluation, 90% classified the experience as *good*, 6.7% as *neutral*, and 3.3% as *poor*.

Figure 7 shows the distribution of content types according to the different levels of visitor experience, which shows very small
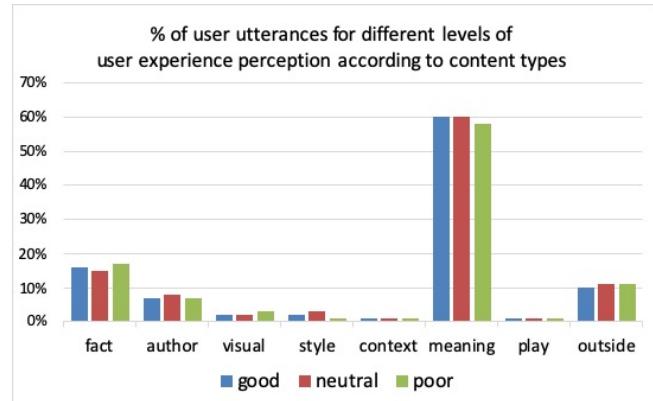
variations among visitors who evaluated their experiences as *good*, *neutral*, or *poor*.

In addition to the visual inspection, a linear regression was performed to identify the significance between each content type and the overall experience perception (represented by the attribute rating). This analysis was done using a transformed dataset from the initial one where the attributes are the number of questions asked in each session related to content type: *author*, *fact*, *context*, *meaning*, *outside*, *play*, *style* or *visual*. The dependent variable is *rating* and the examples are all sessions.

In table 1 we show the results of the linear regression. The model created by the linear regression can not fully explain the rating (*R-squared* = 0.463). However, there were some variables that, in combination, can affect the rating result. Here, positive coefficients indicate that the predictor tends to increase the category's change, and negative ones show the opposite effect. Considering table 1, the content types with statistical significant results ($P(>|t|) < 0.05$) were *author* ($P(>|t|) = 0.000$), *meaning* ($P(>|t|) = 0.000$), *outside* ($P(>|t|) = 0.003$) and *visual* ($P(>|t|) = 0.000$). For the other content types, *fact*, *context*, *play* and *style*, we did not find any statistical significance.

We found the following coefficients for the content types: *meaning* (*coef* = 0.0753), *author* (*coef* = 0.0872), and *visual* (*coef* = 0.1255). They may indicate that when the visitor asks more questions in those categories, it increases the tendency for the visitors to evaluate their experience with a higher score. We can see that *visual* has the highest coefficient, resulting in more influence in a positive visitor evaluation.

We can also observe that *outside* (*coef* = -0.04) increases the chance of a negative visitor evaluation, being the only significant result which was negative. It is coherent because questions categorized as *outside* are often connected to situations where the visitors asked which the system did not have an answer. The influence of breakdowns in visitor experience has been discussed before [6, 18], and it suggests that in this context, they may also influence the overall experience. We plan to investigate further the breakdown situations of this exhibit in a future paper.

Table 1: Resulting coefficients and *P*-values of the linear regression model, where the content types with statistical significance have their respective *P*-values underlined

content type	coefficient	$P > |t|$
fact	-0.0178	0.051
author	0.0872	<u>0.000</u>
visual	0.1255	<u>0.000</u>
style	0.0142	0.624
context	0.0384	0.348
meaning	0.0753	<u>0.000</u>
play	0.0322	0.593
outside	-0.0400	<u>0.003</u>

In summary, the answer for **[RQ5] is YES**, for content types *author*, *meaning*, *visual*, and *outside*. We also tested the impact on the visitor evaluation in combinations of content types. We only found that a session with a high number of *visual* questions and a low number of *outside* questions increased the chance of having a high rating value.

6 DISCUSSION

Our results show that 60% of visitors' questions were related to author intentions, the meaning of the artworks or their elements, and narratives around them. This majority interest in meanings proved to be relatively independent of the artwork itself and independent of artistic movement or the location in the exhibit of the artwork. Visitors with different evaluations of their experience also seem to ask questions following the same distribution of content types. However, we saw some correlations of the content type with the visitor experience rating.

The second most usual content type is related to facts, 17% of questions, also without significant variations across the conditions detailed in [RQ2] to [RQ4] and, in some ways, in the [RQ5] context. Similarly, about 10% was outside the scope of the artwork, including system mistakes, followed by author-related questions (7%), while the other types of content fared basically below 2%.

It may be true that for most visitors, it was the first time they have used a conversational guide in a museum, and perhaps, the first time using a chatbot at all. However, discounting that, and considering the results, it seems fair to assume, for the rest of this discussion, that when people ask questions freely to an artwork using a conversational system, they ask mostly about the author's intentions and its meanings. The interesting question is why visitors may be behaving like that.

If the issue were difficulties in understanding what an artwork is showing, what is being depicted, and so forth, we would probably have a much smaller percentage of meanings-related questions for the romanticism artworks, which depict subjects and situations in fairly common contexts. That was not the case, as seen in fig. 5 and in the negation of [RQ2].

A possible contributor to a large number of meaning-related questions is the 2nd-person paradigm employed by"The Voice of Art." As described before, and shown in the example dialogue of fig. 4, the visitor asks questions to the artwork itself as if an artwork could talk, thus establishing a direct conversation with it. Notice that this direct conversation with the artwork is in starch contrast to most museum audio guides, which often provide commentary in a 3rd-person perspective as if a museum guide was talking about the artwork.

We hypothesize that the visitor-to-artwork conversational framework of "The Voice of Art" may provide a stronger sense of intimacy which may be strengthened by the private nature of the experience allowed by the use of headphones. Also, the use of voice frees the eyes to explore further the artwork, without sensorial conflicts, furthering the disencumbering of the interaction. Moreover, possibly, from the increased sense of contact with the artwork, and expanded detachment from the surroundings, the visitor may feel more comfortable to ask more difficult, interpretative, and open-ended questions.

Unfortunately, the data collected does not allow us to understand the reasons of so many *whys* from the visitors. However, it points towards a general direction of shaping content for conversational systems around meanings and intentions, at least for the ones structured around the 2nd-person framework of unmediated conversation with the artwork. Moreover, we performed this study in Brazil, and in a traditional museum, so those results might vary in other contexts. For other contexts involving conversational systems, our findings should be considered mostly as advice of possible content types visitors expect to learn from exhibits.

Finally, it was interesting to see that the proportion of questions of specific content types, namely *author, meaning, visual, outside*, influences the quality of the visitor experience. In particular, an increase in the number of breakdowns (signaled by *outside* questions) tends to decrease the experience by the visitor. We plan to explore the issue of breakdowns in a future paper.

7 IMPLICATIONS FOR CONTENT DESIGNERS

Those findings, as discussed previously, suggest the following design implications which might assist when defining, producing, and curating content for conversational systems in museums:

DI1: Curators and exhibition designers should consider tailoring the exhibition content of conversational systems to questions related to intentions, meanings, whys, and the stories of the characters portrayed, followed by facts related to the artwork. Our results show that the distribution of the content of the questions asked was independent of artistic styles and the physical location of artworks.

DI2: Curators and exhibition designers should consider the visitors' perceptions of the experience when choosing the content. Conversational system in museums should have an evaluation feedback tool to assist in content curation.

DI3: Curators and exhibition designers should use voice-based conversational interfaces to offer a tailored, personalized, and engaging experience. Up to 75% of the visitors indeed asked at least 33 questions, interacted at least with six artworks, and 50% with four artworks or more, asking at least 18 questions in a little more than 1 hour. This user-initiative conversational mode can also help curators measure visitors' interest and validate whether visitors' interests matched the exhibition's designed aim.

DI4: Curators and exhibition designers should consider giving conversational systems in museums as an "appropriation tool" for the artwork. Giving a voice to art objects may open a direct channel

to visitors to ask questions related to *meaning* using 2nd-person pronouns, allowing the exploration of the meanings behind the exhibits in a more approachable and personalized way.

8 FINAL REMARKS AND FURTHER WORK

This study helped us to explore the main types of content visitors may ask directly to artworks using a voice-based conversational system. We determined that most of the content the visitors asked was related to the *meaning* category and it is clear that the distribution of content types was not affected by the artwork or the movement style, neither the artwork location.

We also observed that the overall perception of the experience can increase according to certain types of content (*meanings, author, visual*). The opposite effect might occur if *outside* of scope content belongs to the tour. We also discovered that a few content categories are not so relevant to increase visitors' satisfaction (*fact, context, play, style*).

We plan next to analyze the conversation flow of individual experiences [7], such as *repair strategies* [2], *progressivity* [10], and other interaction patterns which might emerge from the future conversation analysis, but also considering technological limitations [27]. We did not have access to visitors and curators to interview them while the exhibition was on display. The museum expects to make the exhibition available for the public again in 2021 which will give us a chance to run qualitative studies and access visitors. We plan to create dataset with concrete types of questions from the data collected which may be included in initial corpora for future curators use.

Finally, this paper provides real-world experience and implications for the design of the content of voice-based conversational systems backed up by a large-scale visitor data which is rare to find in the literature. We believe making this discussion available to our community is very important in the current context of lack of design guidelines for those systems and, particularly, in public experience scenarios with embedded bots.

REFERENCES

[1] [n.d.]. *Art's Most Popular: exhibition and museum visitor figures | The Art Newspaper*. Technical Report. https://www.theartnewspaper.com/
[2] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 254.
[3] Roberta Beccaceci, Francesca Fallucchi, Cristina Giannone, Francesca Spagnoulo, and Fabio Massimo Zanzotto. 2009. Education with" Living Artworks" in Museums.. In *CSEDU (1)*. 346–349.
[4] Benjamin B Bederson. 1995. Audio augmented reality: a prototype automated tour guide. In *Conference companion on Human factors in computing systems*. ACM, 210–211.
[5] Stefania Boiano, Ann Borda, Guiliano Gaia, Stefania Rossi, and Pietro Cuomo. 2018. Chatbots and New Audience Opportunities for Museums and Heritage Organisations. In *EVA London 2018*. https://doi.org/10.14236/ewic/eva2018.33
[6] Heloisa Candello and Claudio Pinhanez. 2018. Recovering from Dialogue Failures Using Multiple Agents in Wealth Management Advice. In *Studies in Conversational UX Design*. Springer, 139–157.
[7] Heloisa Candello, Claudio Pinhanez, Mauro Pichiliani, Paulo Cavalin, Flavio Figueiredo, Marisa Vasconcelos, and Haylla Do Carmo. 2019. The effect of

[8] audiences on the user experience with conversational interfaces in physical spaces. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300320
[8] Candello. Heloisa and Fabrício Barth. 2018. Visitor experiences with intelligent voice-based exhibitions. (2018). https://www.researchgate.net/publication/323750166_Visitor_experiences_with_intelligent_voice-based_exhibitions
[9] Margareta Ekarv. 1999. Combating redundancy: writing texts for exhibitions. *The educational role of the museum* (1999), 201–204.
[10] Joel E Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for voice interface design. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. ACM, 26.
[11] Lou Giansante. 2015. Writing verbal description audio tours.
[12] Eva Hornecker and Luigina Ciolfi. 2019. Human-Computer Interactions in Museums. *Synthesis Lectures on Human-Centered Informatics* 12, 2 (apr 2019), i–153. https://doi.org/10.2200/S00901ED1V01Y201902HCI042
[13] Sherry Hsi and Holly Fait. 2005. RFID enhances visitors' museum experience at the Exploratorium. *Commun. ACM* 48, 9 (2005), 60–65.
[14] Stefan Kopp, Lars Gesellensetter, Nicole C. Krämer, and Ipke Wachsmuth. 2005. A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In *Lecture Notes in Computer Science*. Springer-Verlag, 329–343. https://doi.org/10.1007/11550617_28
[15] H. Chad Lane, Clara Cahill, Susan Foutz, Daniel Auerbach, Dan Noren, Catherine Lussenhop, and William Swartout. 2013. The Effects of a Pedagogical Agent for Informal Science Education on Learner Behaviors and Self-efficacy. Springer, Berlin, Heidelberg, 309–318. https://doi.org/10.1007/978-3-642-39112-5_32
[16] Seung Jae Lee. 2017. A review of audio guides in the era of smart tourism. *Information Systems Frontiers* 19, 4 (2017), 705–715.
[17] Gaea. Leinhardt, Kevin. Crowley, and Karen. Knutson. [n.d.]. *Learning conversations in museums*. Lawrence Erlbaum. 461 pages.
[18] Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5286–5297.
[19] Marjorie Munsterberg. 2009. *Writing about art*. http://writingaboutart.org/
[20] Susan Robinson, David Traum, Midhun Ittycheriah, and Joe Henderer. 2008. What would you ask a conversational agent? Observations of human-agent dialogues in a Museum setting. In *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*. European Language Resources Association (ELRA), 1125–1131.
[21] Maria Roussou, Sara Perry, Akrivi Katifori, Stavros Vassos, Angeliki Tzouganatou, and Sierra McKinney. 2019. Transformation Through Provocation?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 627, 13 pages. https://doi.org/10.1145/3290605.3300857
[22] Michelle Kelly Schultz. 2013. A case study on the appropriateness of using quick response (QR) codes in libraries and museums. *Library & Information Science Research* 35, 3 (2013), 207–215.
[23] Jetze Schuurmans and Flavius Frasincar. 2019. Intent Classification for Dialogue Utterances. *IEEE Intelligent Systems* (2019).
[24] Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2006. Interactive humanoid robots for a science museum. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 305–312.
[25] William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, Chad Lane, Jacquelyn Morie, Priti Aggarwal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selina Chu, and Kyle White. 2010. Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. Springer, Berlin, Heidelberg, 286–300. https://doi.org/10.1007/978-3-642-15892-6_30
[26] Angeliki Tzouganatou. 2018. Can Heritage Bots Thrive? Toward Future Engagement in Cultural Heritage. *Advances in Archaeological Practice* (2018). https://doi.org/10.1017/aap.2018.32
[27] Peter Wallis and Bruce Edmonds. 2019. How language works & what machines can do about it. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. ACM, 11.
[28] Rafal Wojciechowski, Krzysztof Walczak, Martin White, and Wojciech Cellary. 2004. Building virtual and augmented reality museum exhibitions. In *Proceedings of the ninth international conference on 3D Web technology*. ACM, 135–144.
[29] Victor W Zue and James R Glass. 2000. Conversational interfaces: Advances and challenges. *Proc. IEEE* 88, 8 (2000), 1166–1180.