TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY-NC-SA

Citation (APA)
Love, J. S., Gu, E. H., & Vandommele, J. (2026). Artificial Curators - Design Considerations for Interactive Chatbot Experiences in Cultural Heritage Settings. *Digital Culture and Education*, *16*(3), 6-40. https://www.digitalcultureandeducation.com/volume-163

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

DIGITAL CULTURE & EDUCATION, 16(3) 2026, ISSN 1836-8301.



Artificial Curators - Design Considerations for Interactive Chatbot Experiences in Cultural Heritage Settings

Heng Gu, Jeroen Vandommele and Jeffery Scott Love

Online Publication Date: 22.6.2026

To cite this Article: H.Gu, J.Vandommele, and, J.S.Love (2026). Artificial Curators - Design Considerations for Interactive Chatbot Experiences in Cultural Heritage Settings. *Digital Culture & Education*, 16(3), 6-40. Available at: https://www.digitalcultureandeducation.com/volume-163

PLEASE SCROLL DOWN FOR ARTICLE

Artificial Curators - Design Considerations for Interactive Chatbot Experiences in Cultural Heritage Settings

Heng Gu (TU Delft)[1]
Jeroen Vandommele (National Library of the Netherlands)[2]
Jeffrey Scott Love (TU Delft)[3]

Abstract: *This paper explores the design and implementation of CuratorBot, an interactive chatbot developed to enhance visitor engagement in cultural heritage institutions using generative large language models. Positioned as a conversational guide, CuratorBot simulates a docent's role of enabling personalised, dynamic dialogues about historical artefacts, specifically focusing on the Dutch National Library's Visboeck manuscript. CuratorBot offers visitors the opportunity to actively interact with cultural heritage by having a conversation with a computer partner. Through iterative design experiments, user interaction observations and surveys, we assess the chatbot's capacity to foster curiosity, support free-choice learning and facilitate human-like conversations, thereby making cultural experiences more accessible and engaging for a broader audience. The development process involved continuous user feedback which highlighted key areas of improvement and adaptation to better serve visitor needs. The paper addresses opportunities and challenges of using LLM-based conversational agents in heritage settings and offers insights for future development of visitor engagement tools with embedded machine learning. We also discuss broader implications of incorporating generative conversational interfaces in GLAMs by examining how such tools can contribute to the evolving landscape of digital cultural heritage.*

Keywords:

LLM Chatbots, GLAMs and Free-choice Learning

The advent of large language models (LLMs) with the capacity to imitate human speech patterns has precipitated an increase in novel, experimental public services featuring written or verbal exchanges between people and machines. As applications which incorporate this type of text-based machine learning (ML) become more varied and more pervasive in our daily interactions with digital systems, we have a growing need for appropriate spaces to explore how these technologies could or should be used. Systems with the ability to mimic speech have stirred an emergent debate on the merits and ills they could bring to individuals and to society, a 'social shock' (Berry 2023), and cultural heritage institutions such as galleries, libraries, archives and museums (GLAMs) with pedigrees as loci of public debate and negotiation have a leading role to play as venues for discussions on the role of ML and society. GLAMs maintain a high level of public trust on account of their perceived integrity and benevolence (cf. Mayer, Davis and Schoorman 1995), and as trustworthy public bodies they can help navigate social and individual acceptance of the transformational nature of LLMs as they have done with of other technologies in the past.[1] This is part of a larger discussion on perceived (lack of) trustworthiness of ML-embedded systems being proliferated throughout several domains (cf. Kleizen et al. 2023). Debates on machine learning applications are already many and varied, and amid the most prominent at present is the use of conversation-based systems supported by generative machine-learning, one of several software applications given the obfuscating and

[1] ORCiD: 0000-0001-7679-698X
[2] ORCiD: 0000-0001-5441-0791
[3] Corresponding author: j.s.love@tudelft.nl

overly general label of 'Artificial Intelligence'. In this study we present how generative conversational systems could function within GLAM spaces by focusing on the design process of building a prototype informed by the opinions, judgements and perceived values of such a system contributed by diverse sets of expertise (Padilla 2020). Discussions of this system are intended to inform others and ultimately to gain more holistic perspectives on possible and (un)preferable futures (cf. Voros 2017) involving machine-generated language.

Dialogic conversations with or via generative LLMs present one path towards constructivist narratives in heritage spaces (Hornecker and Ciolfi 2019, 11-12) in that they offer a potential means for open-ended journeys unique to individual visitors because they adapt on the fly according to how each visitor interacts with them. In order to query how - or whether - this kind of protean human-machine dialog is suitable in a public space such as a library, we developed a prototype conversational agent (CA), or chatbot, dubbed the CuratorBot, as it is designed to be a digital adaptation of a docent or curator, cultural heritage specialists who possess the challenging roles of interpreters and tour guides (Giltinan 2008, 104), as well as the tasks of introducing and discussing complex historical materials to members of the public. The CuratorBot is one realization of recent speculations (Cordell 2020, Chen 2023) on possible directions for machine learning in libraries. It serves as a conversational technology probe (Hutchinson et al. 2003) to spark public discussions on the utility of machine-mediated conversations relating to culture and history. By actively participating in the discussion of ML applications, we collectively negotiate how this technology can be understood within the context of GLAMs as a potential benefit to visitors and staff rather than a nebulous, solutionist panacea (Coleman 2020). The open-endedness afforded by this type of system is intended to support free-choice learning (Heimlich and Falk 2009, Anderson et al. 2007) by creating a space in which individuals can explore ambiguous historical materials and ideas by posing any questions they wish on their own terms while mitigating information overload through curation of topic and system behaviour.

To reify an artificial docent we ran a design experiment at the National Library of the Netherlands (KB) aimed at identifying and translating the experience of having a dialogue about an historical artefact with a machine in lieu of a person. We acknowledge that the experience is not a direct translation, nor is it intended to be, but there remains a certain optimism that artificial partners with potentially unlimited memory and recall, coupled with a kind disposition, may serve as intermediate, temporary guides or teachers during some of our more brief sojourns through libraries and museums. Here we recognize heightened expectations of what digital experiences of heritage can be and attempt to determine whether human-computer dialogue has the capacity to re-enchant the mundane (Baggensen 2019) that is so often promised by technological interventions in cultural heritage settings by invoking curiosity with its speaking partners. Insights come primarily from extensive discussions over the course of two years as well as three datasets collected between March - August 2023. User-submitted questions and the responses provided by the CuratorBot were captured and collected as a form of unobtrusive observation of the system's usage, as were 'like' and 'dislike' judgements of the responses. Visitors were also invited to fill in a survey following their session with the CuratorBot.

This study of how a conversational agent might function within GLAM spaces marks an attempt at the fruitful and humane (Padilla 2020) creation of a CA which could in the future serve as a recognized institutional representative. We approached the challenge of shaping a machine with a curator's voice by collaborating with as many people as possible during the system's creation, maintenance and use, as well as employing continuous, deliberate reflection. Where possible we employ good practices for co-design and for systems involving large-scale machine learning (Yildirim et al. 2022). Our assumption is that the high expectations demanded by GLAM stakeholders will provoke conscious decisions in the development and evaluation of ML-based systems that can be transferred to other domains where traditions of debate and consensus are not as robust. The timing for this and related research at the intersection of ML & GLAMs

is particularly critical as an increasing number of cultural heritage (CH) institutions incorporate Artificial Intelligence and ML into their strategic frameworks.[2]

Use of CAs in libraries has been suggested at least as early as 2010 (Rubin et al. 2010), but up to now there have been very few documented experimental setups for developing or testing them. The CuratorBot charts some of the complexity involved in using generative text ML while serving as a discussion point for multiple perspectives around the emerging technology. At the time of writing there are numerous and disparate voices contributing to the debate of generative text machine-learning. They range from ebullient to catastrophic (Kirschenbaum 2023) to cautious (Jo and Gebru 2020).[3] As we try to listen to as many as we can, there remains a lack of mechanisms to engage with these debates, to make them tangible and allow people to contribute, especially those outside of academic or large tech circles. The CuratorBot serves to fill this need to stimulate conversation about LLMs and other kinds of systems which include them. With the larger debate surrounding LLM-derived services in mind, we use the CuratorBot to explore two questions on CAs within knowledge institutions such as libraries and how developing in these contexts might positively impact the design of interfaces infused with generative ML in the interest of the public good:

1. How feasible and trustworthy are conversational agents as a medium for interacting with history?
2. To what extent can CAs facilitate curiosity as part of the process of learning within cultural heritage institutions?

As a baseline the CuratorBot is designed to uphold values held by GLAM institutions and their members. Mandates regarding GLAM values vary,[4] but in general it is envisioned as part of spaces which enable GLAM learning outcomes (Hooper-Greenhill 2004, Giusti 2008) such as informal learning, enjoyment, inspiration and creativity. For the second question, we approach curiosity as complex, and we limit ourselves to a general understanding of it as a desire to seek new knowledge or otherwise experience the unknown (cf. Grossnickle 2016). Our foray into the phenomenon here is by necessity limited to exploring whether (private) conversations with machines might be worth further investigation in relation to curiosity-based learning.

Chatbots as a Subset of LLM Services

Chatbots are currently one common implementation of LLMs using text interfaces for information retrieval. Some of the more experimental or playful chatbots more closely resemble those created in GLAM contexts, e.g. the Ronald Reagan bot (Qian et al. 2022) and the Star Trek bot (Grishma et al. 2017), two examples among many one-off experiments. Elsewhere in commercial sectors, the exploitation of LLM-based technology is in full swing, where chatbots frequently appear as the occasionally useful but often vexing web-based chat box appearing in the corner of one's screen. These systems are built primarily to limit and/or influence user choices, actions, and decision-making processes (Khan et al. 2024, Tiwari et al. 2022). The domain of health and wellbeing in particular has witnessed a trend in which chatbots have been instrumental in nudging patients towards healthier behaviours and choices (Gentner et al. 2020, Li et al. 2023).

Despite the rapid proliferation of chat-based systems, we remain in the early stages of constructing a collective understanding of how to design them (cf. Jain et al. 2018). For example, we are still discovering the capacity of the technology to be personalized or align with user preferences in order to achieve specific goals. In the educational arena, the deployment of personalized chatbots has not only enhanced learner engagement (Bhanuse 2021) but also positively impacted learning outcomes (Urdaneta-Ponte et al. 2021).

These educational chatbots also restrict questions and responses to content designated in an information corpus. However, recent commercial services incorporating LLMs such as OpenAI's ChatGPT and Google's Bard AI present the possibility, for good or ill, of unscripted, open-ended scenarios for learning and beyond. Chatbots constructed using these and other generative language models are now being used as foundations for text-based interfaces that can look very similar to earlier chatbots but with vast information backgrounds behind them.

Now that development access to these kinds of seemingly boundless generative text systems has been released to the public, we are witnessing an explosion of experimental applications (cf. Sadek et al. 2023, Moore 2023) of varying sophistication and detail. Yet amid the rapid rise and fall of generative chatbot innovations we need to start looking more deeply into how such systems perform and more specifically at how individual LLMs operate within certain contexts on specific tasks. Initial explorations of services such as ChatGPT demonstrate that systems which incorporate it inherit significant merits, including a capacity to field a slew of questions on disparate topics, of varied complexity and in multiple languages (Brown 2020). However, the oft-mentioned promise of completely free discourse remains unfulfilled despite the large quantities of diverse textual materials supplied as training sources for their underlying proprietary models. Some of the shortcomings are immediately visible, not least being superficial/generic responses (Kim et al. 2024) and outright falsehoods to which some currently apply the contested appellation 'hallucinations' (Hicks et al. 2024). One can assume that these technologies will continue to develop, but even with known limitations it is already possible to seek out affordances presented by the flexible, interactive language tools fueled by LLMs.

A Brief History of GLAM Chatbots

Over the past decade or so there have been numerous experiments employing chatbots for a variety of purposes in library, museum and related fields such as tourism.[5] However, there are relatively few robust studies focused specifically on the design and evaluation of CA systems in GLAM settings.[6] Recent lists of notable applications have been furnished by Vartimidias et al. (2021) and Gaia et al. (2019), but the hype surrounding the promise of novel ML technologies, attempts to exploit them and the relative ease of disseminating work online has resulted in an ever-expanding corpus of case studies rich in diversity and difficult to synthesize meaningfully. Parallel to this we see an increasing urgency among GLAM stakeholders to gain perspective on how these new technologies will affect their organizations and professionals working within them (Cox 2021).

A few examples of earlier GLAM chatbots are worth highlighting here on account of salient features which could be beneficial for most, if not all CAs operating in cultural heritage spaces. Máximo the Titanosaur at the Chicago Field Museum (Field Museum 2024) represents an attempt to build a chatbot with an audience disposed toward curiosity (children). The Anne Frank Messenger Bot (Anne Frank House 2017) marks a foray into using a CA to engage with thematically difficult material. And the 'Dialogue with Artwork' installation created by Bruno Barbosa (2016) with the MAMBA in Buenos Aires demonstrates how individual works in a museum could be imbued with their own conversational personality. These all serve as direct inspirations for the CuratorBot.

A brief word on technologies is warranted here. Both Máximo and the Anne Frank bot are built using the Messenger service provided by Facebook/Meta,[7] and this is already more information that we have on most. Practically all CAs hitherto created for GLAMs are experimental and evanescent in nature. As such few details about their design process or technical specifications are immediately available, and in almost all cases the CAs themselves are no longer accessible. At the time of writing, and to the best of the authors' knowledge, there is no GLAM-based, publicly accessible CA built on top of an generative LLM service

besides the prototype CuratorBot, which itself ceased to be maintained in late 2023. By the time of publication we anticipate that the CuratorBot will exist only as static images and recordings due to the difficulties inherent in maintaining prototypes as well as issues stemming from incorporating commercial services into digital interfaces.[8]

Within the complex milieu of rapid technological change the authors' best efforts have been made to identify and incorporate lessons learned from current and recent activities relating to machine-generated conversations within heritage settings. Design of the CuratorBot draws on these earlier efforts as well as work from multiple research domains, including human-computer interaction, museology, library science, pedagogy and history. For system design we relied heavily upon Sanders & Stappers (2014) and Leslie (2019). For adaptation of chatbots to GLAMs we have attempted to embody our interpretations of findings presented in Damala et al. (2019), IFLA (2019) and Proctor & Tellis (2003).

Conversations with the CuratorBot

Falk & Dierking (2008, 20) describe their Contextual Model of Learning in museums as 'a never-ending dialogue between the individual and his or her physical and sociocultural environment'. A conversational agent has the capacity to embody a more or less literal interpretation of their model through an interaction featuring endless dialogue about any type of museum object if given the appropriate instructions for doing so. The CuratorBot is a step in the direction of defining what might constitute appropriate instructions for limitless dialogue and projects the storyteller aspect of curators who create informed narratives tailored to their audiences (Ceyhan 2018). Ideally the CuratorBot should be able to field questions that would be addressed to a curator or person with a similar role, such as a tour guide, who is knowledgeable about objects within a GLAM collection. Choices of how to represent a curator's mentality and the salient features of their personality were iteratively assembled from a series of discussions among the authors, university colleagues, members of library staff and members of the public during demonstrations. The person or people speaking to the CA is envisioned as anyone curious about, or potentially interested in, an object in the collections. This carries an assumption that there are numerous possible underlying motivations for wanting to know about a piece of heritage. The CuratorBot is, like a curator or tour guide, intended to foster peoples' personal interests through a Ciceronian rhetorical voice that aims to teach, to delight and to move. At present it is a given that in-person interactions with an enthusiastic curator are in numerous respects superior to a scripted experience (Eardley et al. 2017), but there may be capacity for a semi- or unscripted machine to inspire similar enthusiasm or else foster it in different terms. In other words, our dialogues with machines need not be the same as those we have with other people.

Early CuratorBot Concept

In November 2022, we introduced the first version of CuratorBot in the 'Speak with Art' demo, featured as an interactive installation at the 'Extending Realities' workshop held at TU Delft, where around 100 members of the Dutch cultural heritage community were invited to interact with it. It featured Edward Hopper's 'Nighthawks' (Figure 1), and the initial concept was to create an experience where users could 'interrogate' an artwork enhanced with comprehensive information about itself. This version utilised the WikiArt API to provide background details about the artist and the artwork's context. An improved version was built using Python and HTML and expanded to include a curated collection of 400 popular paintings.

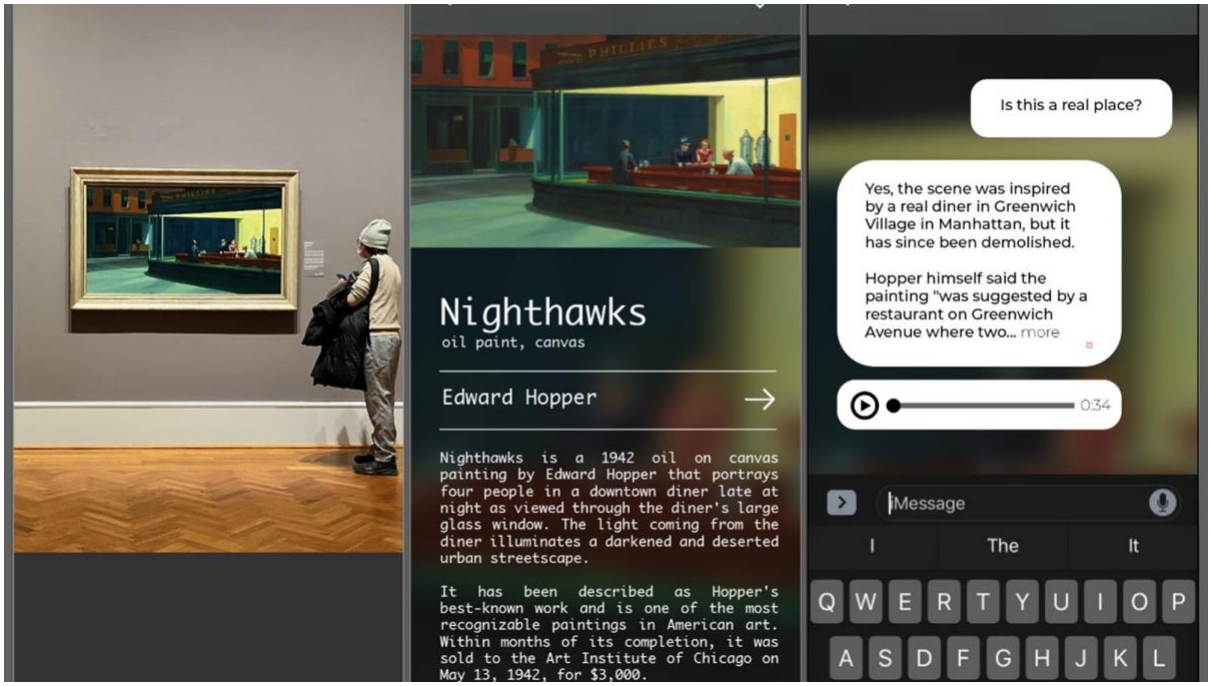


Figure 1. Storyboard for 'Speak with Art'

Conversing about Manuscripts

Through a series of discussions among the authors, 'Speak with Art' was re-designed to become the CuratorBot with the features and design goals discussed above. For the sake of simplicity, the CuratorBot was re-built to facilitate conversations about a single object. Adriaen Coenen's *Visboeck*, one of the treasures of the KB collection, forms the centrepiece of the experience. The *Visboeck* is an illustrated manuscript completed in 1577 and represents a unique piece of Dutch literary history. While the CuratorBot technology is currently exclusively text-based,[9] it was deemed important to have a strong visual presence for the prototype. The volume features numerous eye-catching illustrations which viewers can quickly browse and easily formulate questions to engage with the creatures and monsters presented in the book. Eighteen pages of fantastic beings form a showcase from the *Visboeck*, a deliberate selection of materials which a curator might select to introduce the volume to new audiences. However, the abundance of text posed a technical challenge that CuratorBot system design effectively addresses, albeit imperfectly. One further advantage of the *Visboeck* lies in the extensive knowledge corpus around it, part of which includes themed tours (*rondleidingen*) produced by curators at the KB that guide online visitors through a sequence of pages within the *Visboeck*.[10] Our hypothesis was that the chatbot could heighten engagement compared to the web portal version of the same tour due to visitors' ability to actively participate by asking questions. Based on the types of questions posed, we can be reasonably certain that the selected pages were sufficiently provocative, although the presence of images also surfaced an unforeseen confusion by users about how the system works (see Observations below).

Content Adaptation and Enrichment

The initial CuratorBot prototype of the *Visboeck* relied solely on information derived from the Dutch Wikipedia article on the work[11] on top OpenAI's GPT-3 davinci model. Responses to questions posed using solely this information were deemed too inaccurate too frequently in early tests, and so based on discussions among project partners and assessment of the types of questions users posed to the CuratorBot during pilot studies, we added details from the themed tours mentioned above to extend the background information in the system's knowledge base. When the architecture was subsequently redesigned to use

GPT 3.5 around February 2023, the practice of fine-tuning became an early exercise in prompt crafting. During the final months of the experiment (June-Aug 2023) we had intended to extend the knowledge base further. At this point we experimented with adding two further sections of text: two chapters from a published exhibition catalogue which had featured the *Visboeck* along with hand-edited transcriptions of pages of the *Visboeck* featured in the CuratorBot generously provided by Florike Egmond. However, at the time even this relatively small amount of text (c. 15000 words) proved far too large for the context window available through the GPT 3.5 API, and in the end these technical limitations prevented further expansion of the knowledge base. In short: we possess a large body of materials - transcriptions of the manuscript itself, secondary literature, etc. which could (greatly?) improve accuracy of the CuratorBot and the breadth of questions it can answer well, but performance began to suffer (slow-downs, forced restarts), and it is too expensive to have these passed through OpenAI's API each time the system is queried. This will be the case for all manuscripts and other heritage objects which have been studied in any depth. Some possibilities to address the lack of control over the knowledge base are discussed below.

Nevertheless the impetus for adding to the knowledge base deserves some explanation. This intervention had multiple goals: The first, more specific objective was to address what was perceived to be a fundamental need for interacting with an historical book; people repeatedly requested to know what the text said, as many found the handwritten, early modern Dutch orthography difficult to decipher. Since they were available, edited transcripts of the full text were the logical choice to fill this gap, though this is a rare luxury for any written or printed heritage artifact. The project team also felt that the CuratorBot might benefit from having more thorough background knowledge of the *Visboeck* beyond what is provided in the Wikipedia page and digital tour guide. The final version of the CuratorBot performed quite well in terms of positively surprising speakers with its behaviour and responses, but subject experts and people delving deeper through a series of questions place higher demands on the accuracy and precision of responses. Future audiences will undoubtedly demand the same as the novelty of interactive text models become more normalized.

System Design

To evaluate the efficacy of a chatbot in delivering improved accessibility and enhanced learning outcomes for visitors, we developed a conversational user interface to deliver a specially curated tour titled 'Wonderlijke wezens in het Visboeck' ('Miraculous creatures in the *Visboeck*').[12] It is designed to address any inquiries visitors might have as they navigate through the pages of the *Visboeck*. OpenAI's general GPT 3.5 model, including its own set of proprietary adaptations, was fine-tuned to give it a 'curator personality'. Final instructions provided for the creation of the curator persona were fairly simple yet targeted to capture part of their personality as educational facilitators. We employed a core prompt with a programmable slot that refreshes the information corresponding to the currently displayed page. Grounded in advanced prompt design methodologies (White et al. 2023), this foundational prompt comprised:

1. **Task**: *Act like a curator Chatbot for the Dutch Royal Library.*
2. **Goal:** *Explaining the pages of the Visboeck to visitors.*
3. **Rules**: *Speak in Dutch. Ask questions to probe interest.*
4. **Page Slot:** *The current page features the creature:* [expert-written animal description from the *rondleiding*]

The rationale behind the programmable slot is twofold: 1. Given the token limit of the GPT-3.5 model we utilized, it was impractical to cover the entirety of the tour in one go; thus, segmenting it based on the current creature or activity featured on the page aids in minimizing tokens passed to the API for generation.

2. As the dialogue progresses, the LLM tends to overlook earlier prompt components, hence, regular updates with the most pertinent information assist in keeping the chatbot's responses on track and relevant.

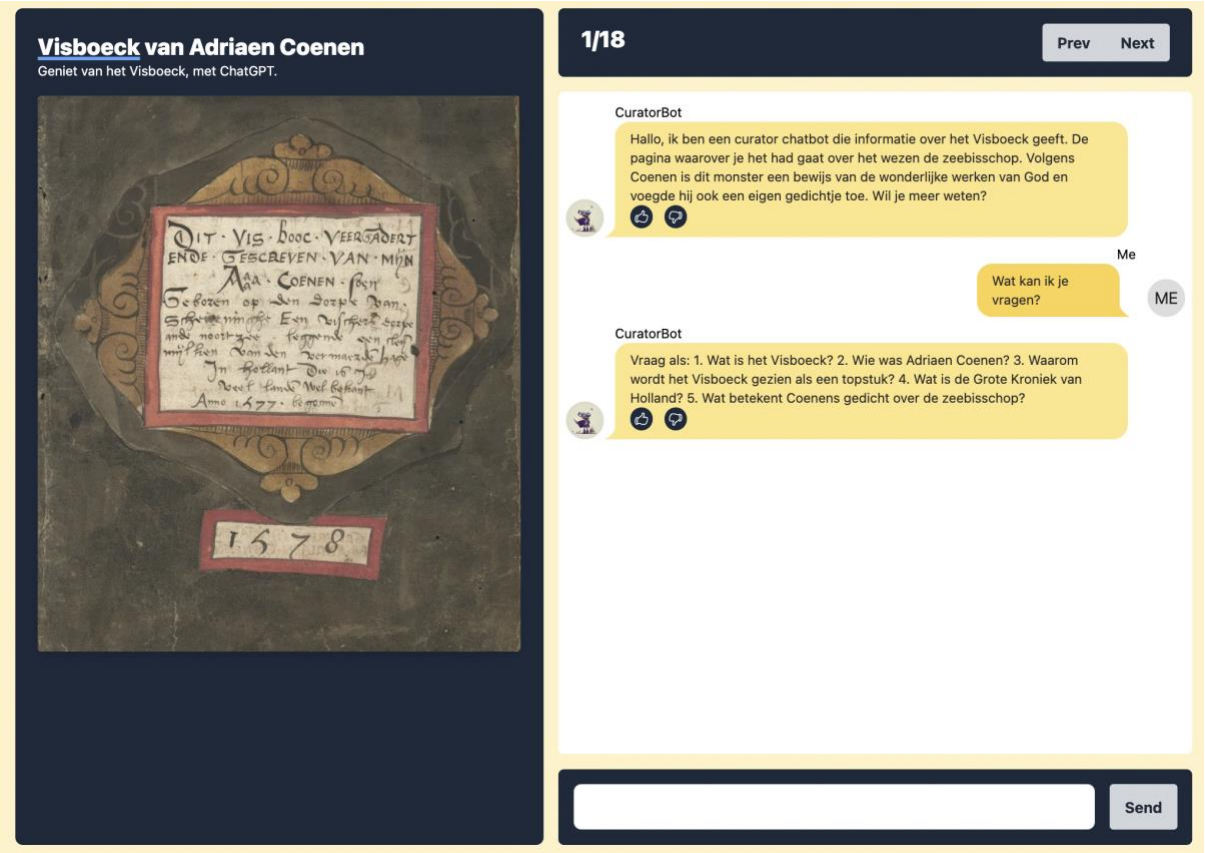


Figure 2. Screenshot of the chat interface showing a visitor interaction with the CuratorBot. The visitor types their question into the text input field at the bottom right. The left panel features one page of the 'Wonderlijke wezens in het Visboeck' tour. Visitors can navigate all of the pages through the buttons above the chat field.

Iterative refinements were made to modify the CuratorBot's perceived behaviour. One major drawback to earlier versions was the authoritative voice used by the CuratorBot in response to queries. Curators and other experts who serve as guides and teachers to the past, tend to be more circumspect with how definitive an answer is. They can explain how confident they are about a choice of explanations or simply say that they do not know, both behaviours which are not default to the general LLM models available for developing CAs. As a result we included prompt instructions for the machine interlocutor to be more circumspect.[13] This was done as a direct response to a general perception among testers that earlier versions of the CuratorBot were speaking too authoritatively, even in situations where it gave incorrect information, the aforementioned 'hallucinations'. Adaptations here were only moderately successful, on which see the 'Observations' section below. As one of the interaction goals was to stimulate curiosity, we also prompted the system to explicitly ask more questions. Given the context of the National Library of the Netherlands, we also changed it to converse in Dutch rather than the default setting of English. A straightforward fix but important for setting tone and expectations.

Data Collection Experiments

Insights presented in this paper primarily derive from conversation, building and testing the CuratorBot prototype, but we also captured three datastreams between March and August 2023 to serve as metrics for the interactions people have with it. Together these assist in addressing the questions introduced above on

feasibility/trustworthiness of CAs for interactions with historical materials and as instigators of curiosity. Collected data include:

- Questions posed by visitors to the CuratorBot (n=884) and machine-generated responses to those questions (n=866)
- Visitor reactions in the form of clickable thumbs-up/-down icons (n=96)
- A survey to gather a self-report on curiosity and feedback on general experience with the system (n= 11)[14]

Questions that Needed Answering

Visitor-submitted questions and answers provided by the CuratorBot were anonymous by default with minimal information collected besides the text submitted by the visitor. Each conversation was given a unique identifier, and each statement within a conversation was labelled as coming from the visitor or the chatbot and assigned a turn number. While the CuratorBot was freely available online, it remained a fragile prototype, and the authors were deliberately selective about where to share the URL. Thus while it is not possible to speak in detail about the anonymous users, we assume that the majority were KB visitors and attendees of events where the CuratorBot was demonstrated. Several visitors tried out the CuratorBot while it was installed as a kiosk in the KB reading room during the period March - June 2023 (Figure 3). Two further large audiences for the CuratorBot include the aforementioned technology showcase event held in Delft in November 2022 and as a poster demonstration during the IFLA WLIC conference in Rotterdam in late August 2023.



Figure 3. CuratorBot kiosk installed within the KB Reading Room

Questions posed to the CuratorBot give us a sense of visitors' encounters with the *Visboeck*. Naturally we anticipated basic questions about the current page on display or the work as a whole, and there were indeed several which fell into this category. Numerous questions were submitted about the birds, fish, unicorns, 'sea bishops' and other beasts depicted in the *Visboeck* pages on

display, as well as a number of queries about the book's author, where the physical book lies, about the colours of the pages, etc. These types of questions tend to appear in the early turns of a dialogue. Beyond these are questions where visitors start to probe the world beyond inspired by materials in the *Visboeck*. There are several queries about whether certain creatures are real, whether there were wild elephants in the Netherlands and what various types of fish ate. Moreover we see attempts to push the system beyond the *Visboeck* to ask about other Dutch historical figures such as Vermeer and Mondriaan. Many questions could be answered directly and well by the CuratorBot, though there were many others it could not handle. In the latter category, visitors frequently asked visual questions about spatial relationships (e.g. what is below an image) or asked the system to perform actions beyond its scope such as flipping to another page. Neither of these were anticipated in building the system, but they provide fruitful possibilities for future features as parts of a conversational system or other type of digital interaction.

Given the anonymous nature of the system we had also expected a modicum of humorous queries, trolling or otherwise surprising questions. In this we were not disappointed, although unexpected questions were relatively few and tame in content. Several queries represent visitors testing the limits of the system. One of the most common questions was whether the CuratorBot spoke a certain language. Multilingualism is one of the strengths of the GPT 3.5 model, and during the demonstration sessions this emerged as one of the most positively surprising aspects mentioned by visitors. Other queries pushed the limits of the system's knowledge base by asking about everything from Harry Potter to whether one can geocache in Rotterdam to what one should bring to an Easter brunch. On the last the CuratorBot performed well in responding that it couldn't help and that its expertise was on collections from the KB, but it did not always respond in this way, and the CuratorBot frequently drew on its general information bank to answer queries outside of its domain. Here it is also worth mentioning the several instances where people spoke to the CuratorBot as if it were human, or at least asked about human qualities. They would ask what the CuratorBot itself thought of the book or how it felt. One asked whether it dreamed (and then naturally followed up to ask if such dreams were of electric sheep).

Many lessons may be taken away from this brief interpretation of visitor questions. To start we might draw a distinction between the kinds of information visitors seek. On the one hand they expect it to possess extensive details on the work itself, including spatio-visual information and full text, as well as meta-details about both the work presented and the system itself. On the other hand people push the boundaries, and it is good to account for this by either moving the boundaries (e.g. expanding the types of information the system can discuss) or else clearly marking when they are met or crossed. We also take away information about the types of interaction which do not yet exist but that visitors would like or at least can imagine: flipping pages, being routed to other systems and contesting/correcting information to name a few. We also tentatively conclude here that the CuratorBot was indeed effective in sparking curiosity, possibly only for a short while and possibly solely due to the novelty of the interaction, but this is a promising start.

Quick Responses: Thumbs-up/-down feedback

Within the web interface we installed a quick and dirty metric to allow the anonymous web users to indicate (dis)satisfaction with a response from the CuratorBot (Figure 4). This was presented in the form of a small thumbs-up and thumbs-down icon below each response generated by the system. The quick response feedback was used to get a rough sense of the types of machine responses about which CuratorBot users had stronger perceived reaction. Visitors made fair use of this feedback mechanism, 96 times out of the 866 captured responses, 11 of which came from a single conversation. The 'like' button was used more frequently than 'dislike' (66 vs. 30 respectively). Given the relative paucity of information we restrict ourselves here to more preliminary or speculative analysis of why certain responses were 'liked' or 'disliked'.

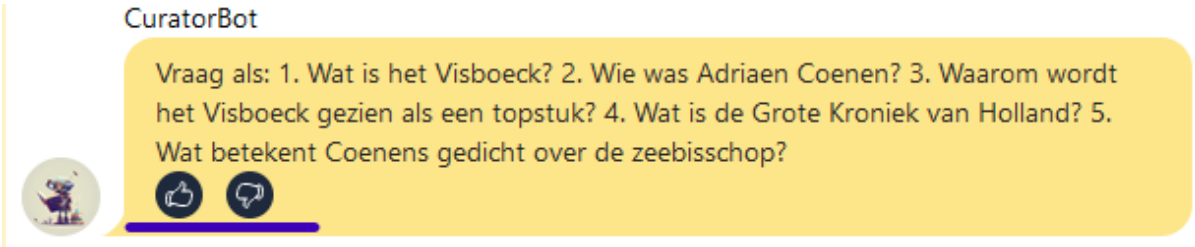


Figure 4. Quick response feedback icons in the CuratorBot interface

Within this small set we can see multiple favourable indications from visitors marking approval for responses in which the CuratorBot explains its behaviour by describing itself as a chatbot and pointing out its limitations.[15] People were less impressed when provided with a demonstrably or factually false answer,[16] but they seemed to appreciate when the system acknowledged corrections to false information given.[17]

Quantitative Analysis of Questions and Responses

In total we collected 884 visitor questions and 866 responses from the CuratorBot. The discrepancy in figures is an artefact of how frequently we were able to capture responses on our web-server. Visitor questions were captured immediately, but it was possible for a session to end before capturing the final machine response in a dialogue. Given the relatively large number of questions and answers passed through the system, we elected to see what types of insights might be visible through a selection of current natural language processing (NLP) statistical tools.

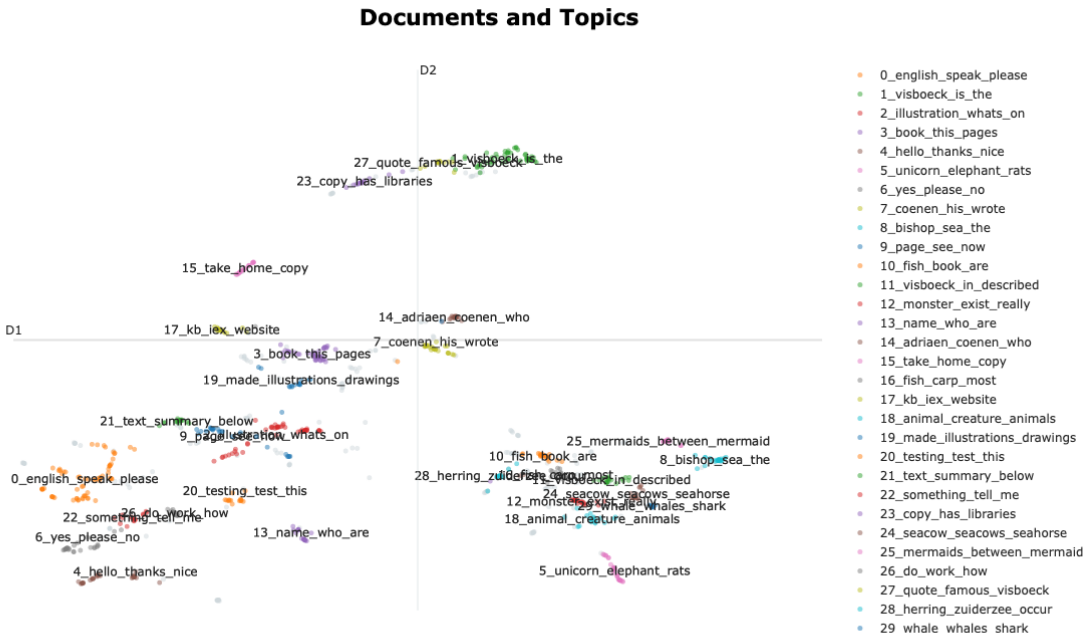


Figure 5. BERTopic clusters of topic presence in the documents

BERTopic, performed with the default model 'all-MiniLM-L6-v2' and parameters of n_gram_range=(1, 2), min_topic_size=10, and top_n_words=10, uncovered 30 distinct topics (see Figure 5). Visualization of the topic semantic space reveals a well-defined cluster related to Visboeck monsters, including themes such as mermaids, herring, and the sea bishop, while another, more loosely connected cluster centres around conversational phrases and instructional language. Notably, despite users describing the conversational bot (CuratorBot) as feeling inhuman, a sizable collection of polite conversational phrases, like 'hello_thanks_nice', emerged. Interestingly, the *Visboeck* monster topics are ranked in order of their appearance in the book, with the most common topic, *unicorn_elephant_rats*, corresponding to the second page in the interface on 'De eenhoorn en olifant'. The prevalence of the *English_speak_please* topic may be attributed to the fact that the CuratorBot was defaulted to Dutch, and without a toggle for language, users instinctively asked the bot to speak in English, reflecting a natural adaptation to the system's limitations. This demonstrates the way user behaviour is shaped by system constraints, as seen in the spontaneous linguistic adjustment.

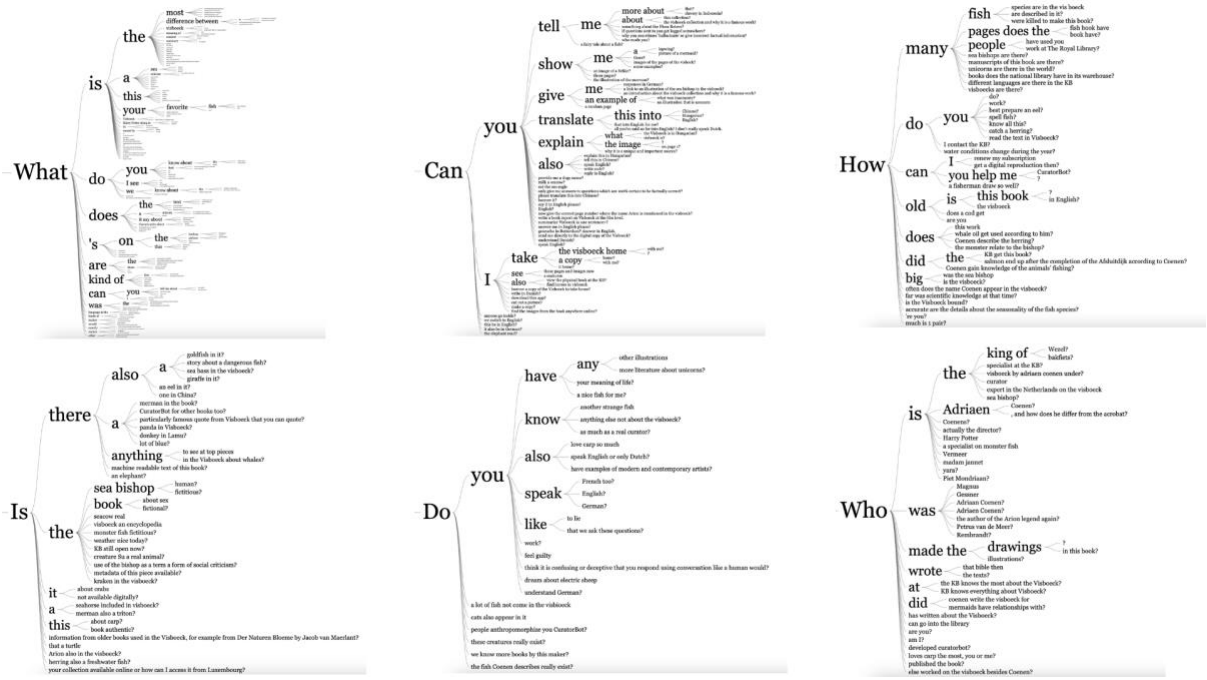


Figure 6. WordTree of top six question syntaxes

The WordTree reveals the most common syntactical patterns users employ when asking questions and reflects their engagement with the system's content and capabilities. '*What*' and '*How*' questions are the most frequent, typically seeking definitions or descriptive explanations related to the Visboeck (e.g., '*What is a sea bishop?*' or '*How big is the fish in Visboeck?*'). '*Who*' and '*Is*' questions tend to focus on verifying facts and confirming the existence or accuracy of specific details. (e.g. '*Who was Adriaen Coenen?*', '*Is the sea bishop real?*'). Interestingly, questions beginning with '*Can*' explore the system's abilities, with users asking if it can translate, show, or tell specific things (e.g., '*Can you show me an example?*').

Curiosity and GLAM Chatbots

One key attribute of docents and curators is their ability to inspire curiosity when discussing collections with visitors of various kinds (Ceyhan 2018, 16). In order to see whether, or to what extent, chatbots of this type have the capacity to facilitate curiosity we built two ways of measuring into the CuratorBot design. In addition to the examination of the questions above, we elected for a survey to allow visitors to self-report their own perceived curiosity while using the system. Eighteen survey questions were derived from Naylor's 1981 State-Trait Curiosity Inventory, which provides a quantifiable metric for a small number of facets relating to curiosity. Naylor's instrument was chosen for its specific focus on curiosity at a given moment, in this case while speaking to the CuratorBot, rather than a more general or diachronic measure of someone's curiosity such as Kashdan et al.'s (2009) CEI-II. It was also hoped that this figure could serve as a means for comparison with other studies which use Naylor's measure. We also began to explore the possibility of incorporating quantitative linguistic metrics, in particular lexical diversity using e.g. MTLD (McCarthy and Jarvis 2010), as proxies for measuring curiosity, but comparison between these and other design interventions require further study to assess meaningfully.

Given the relatively low number of responses to the survey (n=11; fig. 7), we cannot draw any generalizable conclusions about whether people feel this type of interaction with a virtual curator inspires feelings of curiosity nor with confidence compare the numbers against other studies. Instead we highlight a few categories where respondents gave relatively high scores (mean scores on a scale of 1-4): 'I want things to make sense' (3.55), 'I am intrigued by what is happening' (2.91), 'I feel interested in things' (3.09), 'Things feel incomplete' (3.36) and one noticeably low score 'I feel absorbed in what I am doing' (1.80). These may be viewed as hints at further directions to investigate interactions with collections through conversational agents. On lack of absorption it could be beneficial to study how to facilitate 'flow' (Csikszentmihalyi 2014) as part of a conversational experience with collections. Another further question is the extent to which people were curious about the materials on display (the *Visboeck*), the CuratorBot itself or other factors.

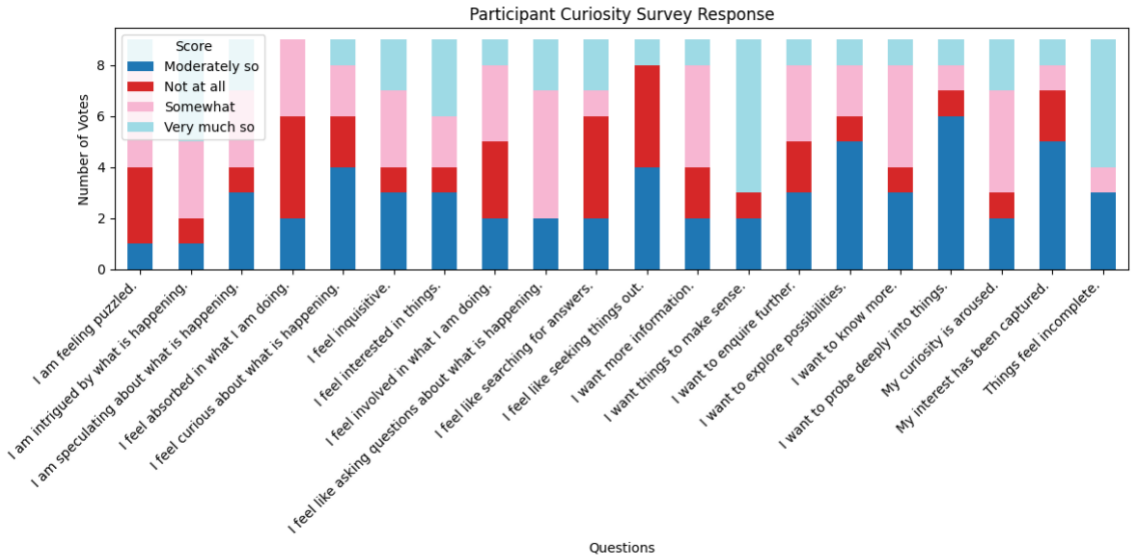


Figure 7: Curiosity Survey Results

Usability & Satisfaction

As mentioned above, most insights into visitors' experiences of the CuratorBot were derived through demonstration sessions and informal conversations with approximately 200 different people over the course of development. These included historians, librarians and other members of library staff, designers, artists and professionals working in other cultural heritage organizations such as museums. Information gathered during these sessions was integrated into subsequent iterations of the CuratorBot, but one final assessment of its overall usability and desirability was conducted through eighteen questions in the optional survey offered to visitors to the website.

Determining the feasibility of CAs as appropriate media for historical inquiry requires a mixed-methods approach. This was partly assessed through analysis of the questions above. A more quantitative view was sought in the second half of the survey (cf. Appendix 2) which includes items adapted from the questionnaire produced by Pallud & Monod (2010) for holistic museum experiences with additional questions relating to the historical content and the chatbot's library environment. The limited number of survey responses indicate mixed perception of CuratorBot.

The user experience section of the survey suggests that while participants were open to future interactions with the CuratorBot, fewer felt that the experience was personal or comparable to interacting with a real person. Participants' open-ended responses further highlighted a mixed reaction. On the positive side, users appreciated the bot's speed and detail, with comments like '*It was much faster - I was impressed*' and '*Detailed information.*' Some also noted that it sparked their curiosity, saying *'My interest has been aroused.*' However, several issues were raised. Many found the bot's responses overwhelming and repetitive, stating '*The CB gives too much information at once...*' and '*I was a bit annoyed by the repetition.*' Some also pointed out inaccuracies, such as '*The curator should not tell incorrect statements*,' and others felt the interaction lacked a personal touch, remarking '*Did not feel like talking to a real person*'.

Participants offered suggestions for improvement, including shorter responses, better accuracy and more contextual understanding. One noted, '*It is not very good at keeping the answers short,*' while another suggested, '*Contextual awareness, perception of time, specifically centuries*.' Some also recommended clearer communication about the bot's role, saying '*Make it clear to the reader from the start what this is exactly*'.

The survey results, while very limited in number, highlight a few important considerations for improving the CuratorBot. We extract the following design requirements:

- Concise and focused responses to further mitigate information overload
- Enhance personification to make the conversation more human-like
- Clear communication of the CuratorBot's capabilities and limitations to set expectations

Observations, Limitations & Future Work

This initial exploration of generative ML text in GLAM contexts suggests multiple research avenues as the technology evolves. Below, we outline positive features for dialogue-based systems and current services for conversational interfaces, followed by promising directions for further inquiry, design, and testing.

CuratorBot's accessibility stands out, notably its ability to communicate in various languages using natural speech. Seamless switching between languages adds comfort, making interactions more inclusive. Understanding the limits of this affordance and how visitors perceive it could be advantageous. Language acumen is one of several features which seem to positively surprise users of the CuratorBot system. It has an air of near-futurity and stands as an embodied example of potential helps and harms of a system incorporating parts of 'artificial intelligence'. There is high value in a setup such as this where people are invited to explore and then individually challenged when their expectations of a seemingly intelligent computer are confirmed, exceeded or disappointed. The CuratorBot has demonstrated the capacity to let large numbers of people interact directly with a new technology and provide an opportunity to develop their own questions about its utility, functions and in-context value.[18] Most insights on this point were derived from observations and discussions with users during in-person demonstrations, and this finding could

be more strongly supported through a study dedicated to observing surprise and otherwise examining expectations of the technology.

The finite knowledge base of systems like CuratorBot is a significant limitation. Users expect near-perfect recall, which remains far from being achieved. Expanding the knowledge base to include both directly relevant details and broader contextual information is essential. Promising approaches, such as using vector databases through Retrieval Augmented Generation (RAG) (cf. Lewis et al. 2020) present one possible approach for doing so. A key discussion for GLAMs is defining the knowledge such systems should have to effectively serve as docents. It may be useful to distinguish between generalist and specialist bots and consider how a machine curator should differ from a human one, perhaps by having perfect recall of past exhibitions and diverse interpretations. We should also discuss how to supplement chatbot knowledge over time and who would be responsible for this. These considerations are crucial for future conversational agents designed for tasks like seeking materials, navigating spaces or creative purposes.

For the moment we seem to be collectively invested in building on current LLM models. The critical issue of fabricated facts might be addressed through better understanding and control of LLM attention mechanism on the information provided to it[19] or perhaps via more complex and detailed behaviour instructions as mentioned above.[20] Anyone who uses a CA should have a reasonable level of confidence in the answers they receive as well as a means of validating them from independent sources. One further caveat here is that we do not understand well what happens during extended conversations between a person and a system such as the CuratorBot. We hypothesize that a dialog with an extended number of speaking turns will eventually result in the chatbot drifting away from its instructions. How long this takes and how impactful it is on a conversation are unknown, but we saw fit to pre-empt this issue by regularly refreshing the CuratorBot system with our relatively minimal set of instructions to maintain precision. Experiments with larger bodies of information and more complex personas will have to reckon with how to stay on topic.

Connected to the information sources made available to conversational systems are the mechanisms whereby the information is conveyed. The tendency of chatbots to give inaccurate or outright false information is a direct result of their fundamental reliance on statistical inference to produce language in human-readable form. While conversational systems like the CuratorBot lack human qualities like emotion or intentionality,[21] they can simulate these traits through language, and potentially other modalities, to create convincing user experiences. One open challenge is how to design these simulations to align with user expectations.

A number of questions also remain open regarding the presentation of CAs. That they should be adapted to local contexts is obvious, but it remains unclear to what extent they should signal their intended use and limitations. People who interact with the CuratorBot seem to appreciate occasional reminders that they are speaking to a machine without omniscience. The medium for language exchange - keyboards, pens, voice, etc. - in a variety of contexts also needs further investigation. The CuratorBot lived briefly as a booth, a kiosk and as a free-standing webpage for mobiles and desktops, all of which made sense for a library prototype and were accepted as such,

but flexibility for input is necessary, not least for purposes of accessibility, and each combination of input + context will provide unique challenges. One wonders further which alternatives to text and typing would be more appropriate to speaking with cultural systems and under what circumstances. Voice and audio are an obvious first foray, but one might also consider embedding such dialogues into features such as learning exercises or games. In the traditionally quiet spaces of libraries it would be strange to speak and be spoken to, but perhaps in the future we could ask our chatbots to whisper or only write for a moment until we move to a place where one might chat aloud freely.

From a holistic design perspective it would be beneficial to know more about how CAs fit into the broader experiences people have when interacting with heritage materials. This could build on existing insights into how pre- and/post-visit studies of experiences in GLAM settings. One can imagine crafting non-invasive ways for someone to 'meet' a CA near the beginning of a visit, and it could likewise contribute to reminder mechanisms such as a closing message to a visit or in some other fashion to stick in the memory as a special type of activity associated with visiting museums (Falk and Dierking 2012, 201ff.). These visitor journeys, as well as the interaction media mentioned just above, speak further to the extent of customizability or personalization we might include in heritage-based CA systems, which may serve to foster diversity and a sense of belonging among GLAM visitors (Antoniou et al. 2016, Price and Applebaum 2022). It would be fairly simple to allow, for instance, organizations or visitors to reconfigure their CA to mimic a certain personality, speak in a different language or exhibit certain quirks of speaking style. It would be worth co-creating some speculative scenarios from these to explore other possible and preferred futures involving talking machines.

Last but not least, it is worth considering how we teach people about computer systems with the capacity to generate speech (and images). There is a common medium of language for exchange, but a more thorough understanding of how LLM systems function and where they come from could enable people to approach grokking them on their own terms. Tied to understanding of the tools are very pragmatic concerns for long-term viability of CA systems. GLAMs and other institutions underpinned by values of independence and education require comprehensive understanding and control over systems - physical and digital - wherever possible in order for them to persist over the medium- and long-term and assert a sense of digital sovereignty (cf. Floridi 2020). Current commercial solutions do not allow this, though it is hoped that challenges to create open alternatives[22] to generative ML monopolies will yield feasible infrastructures as well as a sufficient number of people with the expertise to build upon them.

Notes

1. Cf. e.g. Hong et al. 2002 on the exploration of Web 2.0 in libraries
2. Cf. DE 2021 and BnF 2021 as well as cultural heritage-themed working groups such as Europeana's 2019-2021 'AI Task Force' (https://pro.europeana.eu/project/ai-in-relation-to-glams) and the Turing Institute's ongoing interest group for 'AI & Arts' https://www.turing.ac.uk/research/interest-groups/ai-arts
3. For brief summaries of expert opinions on threats posed by LLMs, see the poll taken by Strickland and Zorpette (2023).

4. For an extensive look at GLAM missions and strategies see Anderson 2019.
5. E.g. Clarizia et al. (2019) on the creation of a 'Digital Cicero' guide.
6. Two notable early exceptions on designing chatbots in GLAMs are Machidon et al. (2020) and Mckie and Narayan (2019).
7. On design of chatbots with Messenger see Smutny and Schreiberova (2020) and Jain et al. (2018), although both surveys focus more on existing usages rather than examinations of why or how the chatbots function in their contexts.
8. A brief demo video of the CuratorBot featuring the authors is available at www.doi.org/10.5281/zenodo.13918317
9. At the time of writing OpenAI's GPT-4o model was recently made publicly available. It allows for both text and visual inputs and outputs, and there will likely be more complex models released in the coming months and years.
10. https://galerij.kb.nl/kb.html#/nl/visboek/
11. https://nl.wikipedia.org/wiki/Visboeck
12. See Appendix 1 for contents. Code for the CuratorBot is available at https://doi.org/10.5281/zenodo.13889523
13. This intervention also marks an attempt to incorporate Hornecker & Ciolfi's (2019, 28) exhortation to remove 'scientific voice' from presentations in museums.
14. A total of 26 surveys were submitted to the system (Qualtrics), but over half (15) were submitted empty. Datasets of the questions, reactions and survey responses are available at: www.doi.org/10.5281/zenodo.13918583
15. This evidenced by responses such as: 'As an AI chatbot, I don't work in the traditional sense, but I am programmed to assist and provide information about the collection of the Dutch Royal Library.' and 'Helaas kan ik je als chatbot geen afbeeldingen laten zien, maar ik kan je wel beschrijven hoe de zeebaars in het Visboeck wordt afgebeeld' (Unfortunately, as a chatbot I cannot show you any images, but I can describe how the *zeebars* was depicted in the Visboeck.).
16. E.g. 'Interessant! Helaas staat er geen afbeelding van een eenhoorn in het Visboek.' (Interesting! Unfortunately there are no pictures of a unicorn in the Visboeck) when there is indeed an image of one, and the false statement 'Ja, Coenen schreef naast de Visboeck ook nog andere werken. Hij was namelijk niet alleen geïnteresseerd in de natuur, maar ook in de geneeskunde en de volkskunde.' (Yes, besides the Visboeck Coenen wrote other works. He was not only interested in nature but also medicine and anthropology).
17. E.g. 'Mijn excuses, je hebt gelijk! Onder de afbeelding van de eenhoorn en olifant zie je inderdaad ratten.' (My apologies, you are correct! Beneath the image of the unicorn and elephant one can indeed see rats). This followed a 'dislike' indication of a response including 'Op deze pagina hebben we het niet over ratten, maar over de eenhoorn en olifant.' (This page isn't about rats but rather the unicorn and the elephant).
18. Cf. Hsi (2008, 139) who puts forth the idea of deliberately educating users of a system about its purpose through interacting with it.
19. On the possibility to more rigorously adjust behaviour of an underlying model and its attention cf. e.g. work on DeepMind's Sparrow by Glaese et al. (2022).
20. Commands, tricks and workarounds to allow LLMs to respond with uncertainty, such as the 'basic' and 'advanced' strategies for Claude (Anthropic 2024), are emerging from various communities.
21. Even these uniquely human attributes are contested, cf. e.g. Terzidis et al. (2023).
22. Current promising competing models include Stanford's Alpaca based on the open-source LLaMA model: https://github.com/tatsu-lab/stanford_alpaca and BigScience's BLOOM: https://arxiv.org/abs/2211.05100

References

- Anderson, D., Storksdieck, M., and Spock, M. (2007). 'Understanding the Long-term Impacts of Museum Experiences'. In Falk, J., Dierking, L., and Foutz, S. (eds.) *Principle, in Practice - New Perspectives on Museums as Learning Institutions*. Lanham: Rowman & Littlefield, pp. 197-215.
- Anderson, G. (2019). *Mission Matters: Relevance and Museums in the Twenty-First Century*. Lanham: Rowman & Littlefield.
- Anne Frank House. (2017). *Anne Frank House Launches Bot for Messenger*. https://web.archive.org/web/20240920080030/https://www.annefrank.org/en/about-us/news-and-press/news/2017/3/21/anne-frank-house-launches-bot-messenger/ (Accessed 6 February 2026)
- Anthropic. (2024). *Strengthen Guardrails: Reduce Hallucinations*. https://web.archive.org/web/20240829143847/https://docs.anthropic.com/en/docs/test-and-evaluate/strengthen-guardrails/reduce-hallucinations (Accessed 6 February 2026)
- Antoniou, A., et al. (2016). 'Capturing the Visitor Profile for a Personalized Mobile Museum Experience: An Indirect Approach' in *24th ACM Conference on User Modeling, Adaptation and Personalisation (UMAP 2016), Workshop on Human Aspects in Adaptive and Personalized Interactive Environments (HAAPIE), Halifax, Canada, 13-17 July 2016*. http://eprints.gla.ac.uk/143234/
- Baggesen, R.H. (2019) 'Mobile Media, Mobility and Mobilisation in the Current Museum Field' in K. Drotner et al. (eds.) *The Routledge Handbook of Museums, Media and Communication*. Milton Park: Routledge, pp. 115–127.
- Barbosa, B. (2016) *MAMBA - Dialogue with the Artwork*. https://brunobarbosa.work/archive (Accessed 6 February 2026)
- Bhanuse, R.S. and Mal, S. (2024) 'Optimal E-learning Course Recommendation with Sentiment Analysis Using Hybrid Similarity Framework', *Multimedia Tools and Applications*, 83(6), pp. 16417–16446.
- Ceyhan, P. (2018) *Facilitated Museum Experience: How Mediated Interactions with Artworks Delivered via Mobile Technology Foster Intellectual Curiosity among College Students*. PhD diss., NC State. http://www.lib.ncsu.edu/resolver/1840.20/35754
- Brown, T. et al. (2020) 'Language Models are Few-shot Learners' in *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates, Inc., pp. 1877–1901.
- Berry, D. M. (2023) 'The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot', *Weizenbaum Journal of the Digital Society*, 3(3). https://doi.org/10.34669/WI.WJDS/3.3.2
- BnF 2021 = Bibliothèque nationale de France. (2021) *BnF Artificial Intelligence Roadmap 2021-2026*. https://web.archive.org/web/20240920083133/https://www.bnf.fr/sites/default/files/2022-01/Poster_AI%20Roadmap_BnF_202112.pdf (Accessed 6 February 2026)
- Chen, X. (2023) 'ChatGPT and Its Possible Impact on Library Reference Services', *Internet Reference Services Quarterly*, 27(2), pp. 121–129. https://doi.org/10.1080/10875301.2023.2181262

- Clarizia, F. et al. (2019) 'A Context-Aware Chatbot for Tourist Destinations' in *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. pp. 348–354. www.doi.org/10.1109/SITIS.2019.00063
- Coleman, C. (2020) 'Managing Bias When Library Collections Become Data', *International Journal Of Librarianship*, 5(1), pp. 8–19.https://doi.org/10.23974/ijol.2020.vol5.1.162
- Cordell, R. (2020) 'Machine Learning + Libraries'. LC Labs. https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf (Accessed 6 February 2026).
- Cox, A.M. (2021) 'Research report: The impact of AI, machine learning, automation and robotics on the information professions', *CILIP White Paper*. https://www.cilip.org.uk/page/researchreport (Accessed 6 February 2026).
- Csikszentmihalyi, M. (2014) *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. New York: Springer.
- Damala, A., Ruthven, I. and Hornecker, E. (2019) 'The MUSETECH model: A comprehensive evaluation framework for museum technology', *Journal on Computing and Cultural Heritage (JOCCH)*, 12(1), pp. 1–22.
- DE 2021 = Digitaal Erfgoed. (2021) 'National Digital Heritage Strategy'. https://web.archive.org/web/20240920082856/https://netwerkdigitaalerfgoed.nl/wp-content/uploads/2022/05/National-Digital-Heritage-Strategy-2021-2024.pdf (Accessed: 6 February 2026).
- Falk, J.H. and Dierking, L.D. (2008) 'Enhancing visitor interaction and learning with mobile technologies' in *Digital Technologies and the Museum Experience: Handheld Guides and Other Media*. Lanham: Rowan & Littlefield, pp. 19–33.
- Falk, J.H. and Dierking, L.D. (2012) *The Museum Experience Revisited*. Walnut Creek: Left Coast Press.
- Field Museum. (2024) *Máximo the Titanosaur*. https://web.archive.org/web/20240920075457/https://www.fieldmuseum.org/exhibitions/maximo-titanosaur (Accessed 6 February 2026).
- Floridi, L. (2020) 'The Fight for Digital Sovereignty: What it is, and Why it Matters, Especially for the EU', *Philosophy & Technology*, 33, pp. 369–378.
- Gaia, G., Boiano, S. and Borda, A. (2019) 'Engaging museum visitors with AI: The case of chatbots' in *Museums and Digital Culture: New Perspectives and Research*. Cham: Springer, pp. 309–329.
- Gentner, T. et al. (2020) 'A systematic literature review of medical chatbot research from a behavior change perspective' in *Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, pp. 735–740.
- Giltinan, K. (2008) 'The Early History of Docents in American Art Museums: 1890-1930', *Museum History Journal*, 1(1), pp. 103–128. https://doi.org/10.1179/mhj.2008.1.1.103
- Giusti, E. (2008) 'Improving Visitor Access' in Loic Tallon and Kevin Walker (eds.) *Digital Technologies and the Museum Experience*. Lanham: Rowan & Littlefield, pp. 97–108.
- Glaese, A. et al. (2022) 'Improving alignment of dialogue agents via targeted human judgements'. https://arxiv.org/abs/2209.14375
- Grishma, J. et al. (2017) 'Enterprise to Computer: Star Trek Chatbot'. https://arxiv.org/abs/1708.00818

- Grossnickle, E.M. (2016) 'Disentangling Curiosity: Dimensionality, Definitions, and Distinctions from interest in educational contexts', *Educational Psychology Review*, 28(1), pp. 23–60.
- Heimlich, J.E. and Falk, J.H. (2009) 'Free-choice learning and the environment' in John H. Falk, Joe E. Heimlich and Susan Foutz (eds.) *Free-choice learning and the environment*. Plymouth: AltaMira Press, pp. 11–21.
- Hicks, M.T., Humphries, J. and Slater, J. (2024) 'ChatGPT is bullshit', *Ethics and Information Technology*, 26(2), p. 38. https://doi.org/10.1007/s10676-024-09775-5
- Hong, W. et al. (2002) 'Determinants of User Acceptance of Digital Libraries: An Empirical Examination of Individual Differences and System Characteristics', *Journal of Management Information Systems*, 18(3), pp. 97–124.
- Hooper-Greenhill, E. (2004) 'Measuring learning outcomes in museums, archives and libraries: The Learning Impact Research Project (LIRP)', *International Journal of Heritage Studies*, 10(2), pp. 151–174.
- Hornecker, E. and Ciolfi, L. (2019) *Human-Computer Interactions in Museums*. [s.l.]: Springer. www.doi.org/10.1007/978-3-031-02225-8
- Hsi, S. (2008) 'Designing for mobile visitor engagement' in Loic Tallon (ed.) *Digital Technologies and the Museum Experience: Handheld Guides and Other Media*. Plymouth: AltaMira Press, pp. 125–146.
- Hutchinson, H. et al. (2003) 'Technology probes: inspiring design for and with families' in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM Press, pp. 17-24.
- IFLA 2019 = International Federation of Library Associations and Institutions (IFLA). (2019) 'IFLA Statement on Libraries and Artificial Intelligence'. White paper. https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla_statement_on_libraries_and_artificial_intelligence.pdf (Accessed 6 February 2026).
- Jain, M. et al. (2018) 'Evaluating and informing the design of chatbots' in *Proceedings of the 2018 Designing Interactive Systems Conference*. New York: ACM Press, pp. 895–906. https://www.doi.org/10.1145/3196709.3196735
- Jo, E.S. and Gebru, T. (2020) 'Lessons from archives: strategies for collecting sociocultural data in machine learning' in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. New York: Association for Computing Machinery, pp. 306–316. www.doi.org/10.1145/3351095.3372829
- Kashdan, T.B. et al. (2009) 'The Curiosity and Exploration Inventory-II: Development, Factor Structure, and Psychometrics', *Journal of Research in Personality*, 43(6), pp. 987–998. https://doi.org/10.1016/j.jrp.2009.04.011
- Khan, M. et al. (2024) 'ChatGPT and the future of impact assessment', *Environmental Science & Policy*, 157, p. 103779. https://doi.org/10.1016/j.envsci.2024.103779
- Kim, Y. et al. (2024) 'Understanding users' dissatisfaction with ChatGPT responses: Types, resolving tactics, and the effect of knowledge level', *Proceedings of the 29th International Conference on Intelligent User Interfaces*. https://www.doi.org/10.1145/3640543.3645148

- Kirschenbaum, M. (2023) 'Prepare for the Textpocalypse', *The Atlantic*, 8. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/ (Accessed 6 February 2026)
- Kleizen, B. et al. (2023) 'Do Citizens Trust Trustworthy Artificial Intelligence?', *Government Information Quarterly*, 40(4). https://doi.org/10.1016/j.giq.2023.101834
- Leslie, D. (2019) 'Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector'. White Paper. The Alan Turing Institute. https://zenodo.org/doi/10.5281/zenodo.3240528
- Lewis, P. et al. (2020) 'Retrieval-augmented generation for knowledge-intensive NLP tasks', *Advances in Neural Information Processing Systems*, 33, pp. 9459–9474.
- Li, B. et al. (2023) 'Users' intention to adopt artificial intelligence-based chatbot: a meta-analysis', *The Service Industries Journal*, 43(15–16), pp. 1117–1139.
- Machidon, O. et al. (2020) 'CulturalERICA: A conversational agent improving the exploration of European cultural heritage', *Journal of Cultural Heritage*, 41, pp. 152–165. https://doi.org/10.1016/j.culher.2019.07.010
- Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995) 'An integrative model of organizational trust', *Academy of Management Review*, 20(3), pp. 709–734.
- McCarthy, P.M. and Jarvis, S. (2010) 'MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment', *Behavior Research Methods*, 42(2), pp. 381–392.
- Mckie, I.A.S. and Narayan, B.. (2019) 'Enhancing the Academic Library Experience with Chatbots: An Exploration of Research and Implications for Practice', *Journal of the Australian Library and Information Association*, 68(3), pp. 268–277. https://doi.org/10.1080/24750158.2019.1611694
- Moore, O. (2023) 'How Are Consumers Using Generative AI?', *Andreessen Horowitz*. https://a16z.com/how-are-consumers-using-generative-ai/ (Accessed 6 February 2026).
- Naylor, F.D. (1981) 'A State-trait Curiosity Inventory', *Australian Psychologist*, 16(2), pp. 172–183.
- Not, E. et al. (2018) 'Writing postcards from the museum: composing personalised tangible souvenirs' in *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*. Article 5, pp. 1–9. https://www.doi.org/10.1145/3125571.3125583
- Padilla, T. (2020) 'Responsible Operations: Data Science, Machine Learning, and AI in Libraries'. *OCLC Research Position Paper*, 26 August 2020. https://www.oclc.org/research/publications/2019/oclcresearch-responsible-operations-data-science-machine-learning-ai.html (Accessed 6 February 2026).
- Pallud, J. and Monod, E. (2010) 'User experience of museum technologies: the phenomenological scales', *European Journal of Information Systems*, 19(5), pp. 562–580.
- Price, C.A. and Applebaum, L. (2022) 'Measuring a Sense of Belonging at Museums and Cultural Centers', *Curator: The Museum Journal*, 65(1), pp. 135–160.
- Proctor, N. and Tellis, C. (2003) 'The State of the Art in Museum Handhelds in 2003', *Papers: Museums and the Web 2003*.https://eric.ed.gov/?id=ED482157 (Accessed 6 February 2026).
- Qian, X., Oard, D.W. and Chan, J. (2022) 'Conversational Interaction with Historical Figures: What's It Good For?' in Smits, M. (ed.) *Information for a Better World: Shaping the*

Global Future. iConference 2022. Lecture Notes in Computer Science, vol. 13193. Springer, pp. 38–48. https://doi.org/10.1007/978-3-030-96960-8_3

- Rubin, V.L., Chen, Y. and Thorimbert, L.M. (2010) 'Artificially intelligent conversational agents in libraries', *Library Hi Tech*, 28(4), pp. 496–522.
- Sadek, M., Calvo, R.A. and Mougenot, C. (2023) 'Co-designing Conversational Agents: A comprehensive review and recommendations for best practices', *Design Studies*, 89, p. 101230. https://doi.org/10.1016/j.destud.2023.101230
- Sanders, L. and Stappers, P.J. (2014) 'From designing to co-designing to collective dreaming: Three slices in time', *Interactions*, 21(6), pp. 24–33.
- Smutny, P. and Schreiberova, P. (2020) 'Chatbots for Learning: A Review of Educational Chatbots for the Facebook Messenger', *Computers & Education*, 151, p. 103862. https://doi.org/10.1016/j.compedu.2020.103862
- Strickland, E. and Zorpette, G. (2023) 'The AI Apocalypse: A Scorecard', *IEEE Spectrum*. https://spectrum.ieee.org/artificial-general-intelligence (Accessed 6 February 2026).
- Terzidis, K., Fabrocini, F. and Lee, H. (2023) 'Unintentional Intentionality: Art and Design in the Age of Artificial Intelligence', *AI & SOCIETY*, 38(4), pp. 1715–1724. https://doi.org/10.1007/s00146-021-01378-8
- Tiwari, C.K. et al. (2024) 'What drives students toward ChatGPT? An investigation of the factors influencing adoption and usage of ChatGPT', *Interactive Technology and Smart Education*, 21(3), pp. 333–355.
- Urdaneta-Ponte, M.C., Mendez-Zorrilla, A. and Oleagordia-Ruiz, I. (2021) 'Recommendation systems for education: Systematic review', *Electronics*, 10(14), p. 1611.
- Vartimiadis, S. et al. (2021) 'Graph-Based Conversational AI: Towards a Distributed and Collaborative Multi-Chatbot Approach for Museums', *Advanced Technologies in Digitizing Cultural Heritage*, 11(19), p. 9160. https://doi.org/10.3390/app11199160
- Voros, J. (2017) 'Big History and Anticipation: Using Big History as a Framework for Global Foresight', in Poli, R. (ed.) *Handbook of Anticipation: Theoretical and Applied Aspects of the Use of Future in Decision Making*. pp. 1–40.
- White, J. et al. (2023) 'A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT', *arXiv preprint arXiv:2302.11382*. https://arxiv.org/abs/2302.11382
- Yildirim, N. et al. (2022) 'How Experienced Designers of Enterprise Applications Engage AI as a Design Material', *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 483, pp. 1–13. https://www.doi.org/10.1145/3491102.3517491

CRediT Statement (https://credit.niso.org/)

Conceptualization	H. Gu, J.S. Love, J. Vandommele
Investigation	H. Gu, J.S. Love
Software	H. Gu
Validation	J. Vandommele
Writing - original draft	J.S. Love
Writing - edits and final draft	H. Gu, J. Vandommele, J.S. Love

Appendix 1: CuratorBot Contents

The following pages form the corpus of the CuratorBot prototype (source: Het Geheugen):[4]

1. fol. 11r (title page)
2. fol. 13r(unicorn and elephant)
3. fol. 102r (Monstervis [Zeekoe])
4. fol. 104r (Monstervis)
5. fol. 112r (Boomgansjes)
6. fol. 185v (Zeebischop)
7. fol. 187v (Zeemonick)
8. fol. 190v (Meermin en meerman)
9. fol. 191r (meermin)
10. fol. 193v (wilde vrouw)
11. fol. 200v (zeearend)
12. fol. 226r (Su beest)
13. fol. 19v (Plinische rassen)
14. fol. 230r (Zeedraken)
15. fol. 228v (Zeepaard bij Delft)
16. fol. 239v (Zevenkoppig monster)
17. fol. 276v (Zeebezem)
18. fol. 326v (Krakovische monster)

Appendix 2: CuratorBot Survey

You are being invited to participate in a research study on Interactive Chatbots for Cultural Heritage Experiences. This study is being conducted by J.S. Love and H. Gu from the TU Delft in collaboration with the National Library of the Netherlands (KB) as part of the Future Libraries Lab.

The purpose of this research study is to learn more about how conversational agents - which some might call AI Chatbots - could or should be used in libraries and museums. It will take you approximately 15 minutes to complete. The data will be studied and used in a series of scholarly academic publications. We will be asking you to rate a series of statements about your interaction with the CuratorBot as well as get some feedback from you on your overall experience with the prototype.

This survey is designed to be completely anonymous, so please answer as honestly as you can. As with any online activity the risk of a breach is always possible, but we will work to the best of our ability to ensure that there will be no way to connect your answers in this study specifically to

4 https://geheugen.delpher.nl/nl/geheugen/view/visboeck?coll=ngvn&maxperpage=36&page=1&query=visboeck&identifier=TOPSTUK%3A0071

you. We believe in open science, and the scores & comments from this survey may eventually be made public as authentic evidence of how people interact with this technology in a library setting.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any questions. However once you submit the form there will be no way to separate your responses from those given by others.

For further details about the project and the prototype, visit [https://visboeck.vercel.app/about]. If you have questions or concerns, you are welcome to get in touch at j.s.love@tudelft.nl

By clicking the arrow below you agree to the terms above and will proceed to the survey.

Q16 We'd like to know how you're feeling right now after using the CuratorBot to explore the Visboeck. Please rank the following statements on the following scale: 1 - not at all, 2 - somewhat, 3 - moderately so, 4 - very much so

	Not at all (1)	Somewhat (2)	Moderately So (3)	Very much so (4)
I want to know more. (1)	o	o	o	o
I feel curious about what is happening. (4)	o	o	o	o
I am feeling puzzled. (5)	o	o	o	o
I want things to make sense. (6)	o	o	o	o

I am intrigued by what is happening. (7)	o	o	o	o
I want to probe deeply into things. (8)	o	o	o	o
I am speculating about what is happening. (9)	o	o	o	o
My curiosity is aroused. (10)	o	o	o	o
I feel interested in things. (11)	o	o	o	o
I feel inquisitive. (12)	o	o	o	o
I feel like asking questions about what is happening. (13)	o	o	o	o
Things feel incomplete. (14)	o	o	o	o
I feel like seeking things out. (15)	o	o	o	o

I feel like searching for answers. (16)	o	o	o	o	
I feel absorbed in what I am doing. (17)	o	o	o	o	
I want to explore possibilities. (18)	o	o	o	o	
My interest has been captured. (19)	o	o	o	o	
I feel involved in what I am doing. (20)	o	o	o	o	
I want more information. (21)	o	o	o	o	
I want to enquire further. (22)	o	o	o	o	

Q20 How was your overall experience speaking with CuratorBot?

	Definitely not (14)	Probably not (15)	Might or might not (16)	Probably yes (17)	Definitely yes (18)

Would you be interested in talking to CuratorBot again about the Visboeck or other books? (1)	o	o	o	o	o
Do you think people you know would enjoy this? (2)	o	o	o	o	o
Did you feel like your conversation with CuratorBot was a personal experience? (3)	o	o	o	o	o
Did talking to CuratorBot feel like talking to a real person, e.g. a tour guide or a curator? (4)	o	o	o	o	o
Was there anything you thought was missing when talking to CuratorBot instead of a person? (5)	o	o	o	o	o

Was there anything you thought was better when talking to CuratorBot instead of a person? (6)	o	o	o	o	o
Is there anything you would like to see CuratorBot show you after the Visboeck? This might be another book, several books on a subject or any other situation where talking to CuratorBot might be interesting to you. (7)	o	o	o	o	o
Did CuratorBot provide enough details about the historical context of the Visboeck? (8)	o	o	o	o	o

Did you find that CuratorBot explained well the cultural, geographical, and historical context of the Visboeck? (9)	o	o	o	o	o
Did you manage to gain a better understanding of historical topics while talking to CuratorBot? (10)	o	o	o	o	o
Did talking to CuratorBot about historical characters or topics provoke emotional reactions such as joy, sadness, surprise, or anger? (11)	o	o	o	o	o
Did the conversation with CuratorBot contribute to your self-projection into the past? (12)	o	o	o	o	o

Did talking to CuratorBot help you understand the history of the Visboeck and its connections to Dutch Culture? (13)	o	o	o	o	o
Did talking to CuratorBot about historical topics make you imagine yourself living in the past? (14)	o	o	o	o	o
Did talking to CuratorBot about the Visboeck offer you an opportunity to connect with your own heritage? (15)	o	o	o	o	o
Did talking to CuratorBot provide you with an opportunity to understand history better? (16)	o	o	o	o	o

Did you feel like the conversation with CuratorBot was personalized for you? (17)	o	o	o	o	o
Did you feel like you were able to have a meaningful conversation with CuratorBot? (18)	o	o	o	o	o

Q22 Is there anything else you would like us to know about your experience with the Visboeck CuratorBot? Tell us here!

Image Descriptions

1. Caption: Storyboard for ‘Speak with Art’
   Desc: three panels: 1. a man looking at Hopper's 'Nighthawks' painting. 2. A mobile phone display with Hopper's painting and a museum label. 3. a mobile chat interface
2. Caption: Screenshot of the chat interface showing a visitor interaction with the CuratorBot. The visitor types their question into the text input field at the bottom right. The left panel features one page of the ‘Wonderlijke wezens in het Visboeck’ tour. Visitors can navigate all of the pages through the buttons above the chat field.
   Desc: A webpage with the cover of the Visboeck on the left and a chat interface on the right
3. Caption: CuratorBot kiosk installed within the KB Reading Room
   Desc: A computer screen and keyboard on top of a wooden cabinet containing three shelves of plants and flanked by a curtain
4. Caption: Quick response feedback icons in the CuratorBot interface
   Desc: A detail screenshot containing the robot mascot of the CuratorBot, a response generated by the system and small thumbs-up and thumbs-down icons
5. Caption: BERTopic clusters of topic presence in the documents
   Desc: A graph containing thirty word clusters corresponding to topics classified by BERT
6. Caption: WordTree of top six question syntaxes
   Desc: Six word trees showing the sentences beginning with the most common words: what, can, how, is, do and who
7. Caption: Usability and Satisfaction Survey Results
   Desc: A stacked bar chart showing the ratings given by respondents to survey questions on participant curiosity