Check for updates

Article

new media & society
1–26
© The Author(s) 2025

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14614448251384258
journals.sagepub.com/home/nms

ChatGPT for cultural heritage and the customization of generative AI: A talkthrough analysis of the Luigi Einaudi chatbot

Simone Natale
University of Turin, Italy

Bruno Surace
University of Turin, Italy

Enrico Mensa
University of Turin, Italy

Luca Befera
University of Turin, Italy

Abstract

Cultural heritage institutions have recently experimented with the customization of generative AI and specifically large language models (LLMs) to create chatbots that impersonate historical characters. While this application has the potential to enhance user engagement, challenges remain, especially considering the need to ensure authenticity and historical accuracy and LLMs' tendency to hallucinate. Through the analysis of the case study of a chatbot developed to impersonate the historical figure of Luigi Einaudi, the first elected president of the Italian Republic, the article interrogates opportunities, problems and risks raised by the customization of generative AI, in a moment when decisions about applying or not applying LLMs are being considered by many cultural heritage institutions around the world. The article, moreover, introduces a variant of the walkthrough method aimed to study AI conversational agents, called

Corresponding author:
Simone Natale, University of Turin, Department of Humanities, Via Sant'Ottavio 20, 10124 Torino, Italy.
Email: simone.natale@unito.it

'talkthrough', which represents an important addition to the methodological toolbox that can be activated to study generative AI.

Keywords
Artificial Intelligence, chatbots, cultural heritage, customization, digital methods, generative AI, large language models, talkthrough, walkthrough method

Introduction

Academic and public discussions of generative AI often emphasize how the semi-autonomous training of AI models on large masses of data leads to their capacity of performing tasks such as writing and producing images (Bode and Goodlad, 2023). Less attention is given to the work of individuals, groups, institutions and companies that customize and fine-tune these models so that they can be applied to specific tasks and domains. Yet, generative AI applications and their global impact largely depend on this work. Since systems such as text-to-image models and large language models (LLMs) are not domain specific, in fact, they need to be carefully prepared, adapted and fine-tuned in order to become usable in diverse contexts and areas of applications. Rather than being a neutral process, such customization processes represent a culturally rich and technically complex work that informs the functioning and applications of these models, and therefore, their impact and implications on contemporary societies and cultures.

This article investigates the customization of LLMs such as ChatGPT for application in the cultural heritage. Institutions active in this area have over time embraced a wide range of technologies – from the moving image in the early twentieth century to the digitization of archives and museum experiences – to preserve, curate and disseminate the legacy of tangible and intangible heritage assets (Parry, 2007). We applied a version of the walkthrough method (Light et al., 2018), which entails taking up a user role to systematically explore the functionality of a digital software, to examine the case of a chatbot created by the Italian-based Einaudi Foundation to impersonate the historical figure of Luigi Einaudi, the first elected president of the Italian Republic. As we will show, this case provides a unique entry point to interrogate the opportunities but also the problems and risks raised by the customization of generative AI, in a moment when decisions about applying or not applying LLMs are being considered by many cultural heritage institutions around the world. The deep analysis achieved through the implementation of the walkthrough provides insights into how choices made to customize the chatbot impacted the final product, and can therefore help stakeholders interested in cultural heritage and beyond who are considering the application of this technology. Moreover, this study contributes to ongoing critical discussions of synthetic heritage, understood as the activation of generative AI to blend authentic historical elements with artificial enhancements (Nieto McAvoy and Kidd, 2024), helping to unveil the growing but uneasy relationship between AI, collective memory and the historical record.

As LLMs rapidly attracted the centre of public attention after the launch of ChatGPT in 2022, museums and archives started to consider applications that could benefit their

missions and activities (Thiel and Bernhardt, 2024). Among the applications under exploration, chatbots impersonating historical characters or other elements of historical heritage have become one of the most significant areas of development and experimentation (Fiedler, 2025). Our analysis shows how the application of these emerging technologies entails balancing the interactive and engaging opportunities offered by the chatbot with the specific requirements and needs raised by the mission to preserve and communicate elements of historical heritage. Due to generative AI's uneasy relationship with accuracy and LLMs' tendency to hallucinate (Hicks et al., 2024), the question if LLM-powered chatbots can ensure historical accuracy remains open, and while specific customization strategies and guardrails have been developed to minimize this problem, these may affect the flow and the perceived naturalness of the conversational experience. More broadly, the examination of this case study provides an exceedingly useful opportunity, in a moment when techno-solutionist approaches call upon the cultural sector to adopt generative AI for cultural heritage and beyond, to reflect on the implication of leveraging and customizing AI models that are often trained through indiscriminate extraction of historical sources, but are often presented as the ultimate tool for synthesizing the historical record and for presenting it to the public (Amoore et al., 2024; Schuh, 2024).

An additional contribution of the article, moreover, is at a methodological level. Although the walkthrough method has been widely implemented and developed, it remains unfit to study the case of AI conversational agents, which include chatbots such as ChatGPT, voice assistants such as Alexa and Siri and companion chatbots such as Replika. Since these technologies are programmed to enter in conversation with users, applying the walkthrough should also entail engaging in dialogue with the machine; while scholars have started to explore prompting as a research strategy to study generative AI (e.g. Gillespie, 2024), the protocols and strategies to implement this as part and within the walkthrough methodology have yet to be developed. The project contributes to this gap by formalizing and applying for the first time an experimental methodology, the 'talkthrough', which was originally conceived and experimented by the first author in collaboration with Iliana Depounti. Considering the ongoing research attention for conversational agents and AI, the development of this novel research method addresses a clear research need for researchers interested in inquiring the patterns of interactions between humans and machines, and thus constitute an added area of innovation for this article. Our hope is to contribute and enrich ongoing efforts (e.g. Digital Methods Initiative, 2025) aimed to adapt existing methods as well as to develop novel ones for studying and researching generative AI.

The remainder of the paper is organized in four sections. In the first section, we discuss emerging literature addressing applications of generative AI for cultural heritage and review the multiple technical solutions that can be employed in this context, focusing on the specific case of GPT, the LLM that has been employed for the Einaudi chatbot project. The second section describes the methods of the walkthrough and talkthrough and how we applied them to the case of the Einaudi chatbot. The third section analyses the findings of the walkthrough, and the conclusion outlines implications and contributions of this research.

The customization of generative AI for cultural heritage

Generative AI, LLMs and synthetic heritage

Despite their relatively recent introduction, generative AI has already been applied in a wide range of areas and sectors. Examples of its customization to specific areas and tasks include personalized versions of ChatGPT that users generate through tools available in the OpenAI platform so that they can address their individual needs (OpenAI, 2023), chatbots that are trained or prompted to impersonate historical characters (character.ai, 2024), versions of LLMs that are fine-tuned to be used for specific domains and tasks, such as academic research, healthcare or online searches (Maharjan et al., 2024), and the personalization of text-to-image models that makes it easier to produce specific types of images (Heaven, 2024). In this context, the cultural heritage sector has not been the exception. Institutions including museums, archives and libraries worldwide have made efforts to customize LLMs and other generative AI models as a pathway to facilitate the dissemination of knowledge and to introduce new modalities of access to cultural heritage that enhance users' engagement. In particular, the customization of LLM-based chatbots to create conversational agents that impersonate historical characters or introduce other pathways for the public to engage with cultural heritage is being attempted by a growing number of institutions worldwide. Examples include the Rosicrucian Egyptian Museum's customization of Ameca, an LLM robot, into Thoth, the Egyptian god of writing, in San José, California (Rosicrucian Egyptian Museum, 2025); the Metropolitan Museum's featuring of a chatbot impersonating socialite Natalie Potter as part of the exhibition 'Sleeping Beauties: Reawakening Fashion' in New York City (MET Museum, 2024); and the avatars of naturalists Alfred Russel Wallace and Michele Lessona included in the permanent exhibition at the Regional Museum of Natural Sciences of Torino, Italy (Piemonteinforma, 2024). Even before the generative AI turn, moreover, chatbots had been used as tour guides (Kopp et al. 2005), and gamification tools to increase museum entertainment (Gaia et al., 2019). Chatbots impersonating historical figures had also been explored. These included, for example, the Anne Frank chatbot, which was launched in 2017 by the museum dedicated to her in Amsterdam and could be asked questions about her life via Facebook Messenger (Tzouganatou, 2018).

Beyond the cultural heritage field, characterization has often been used to improve chatbots' credibility and storytelling, but critics have highlighted issues including the use of gender representations and stereotypes (Chen, 2025), the potential for manipulation (Natale, 2021) and gamification dynamics (Ge and Hu, 2025). Moreover, outputs of generative AI can give an impression of high specificity, but at the same time have been shown to exhibit significant levels of homogeneity (Westberg and Kvåle, 2024), which might jeopardize the capacity of LLM-based chatbots to deliver authentically original or sophisticated responses.

The customization of generative AI models is not only regulated by technical affordances, but also by aspects that are specific to the domain of application, its professional culture and institutional environment (Parry, 2010), thereby raising a range of specific challenges and problems. A key issue in this context concerns the uneasy relationship between authenticity and historical reconstruction. There is a long trajectory of

cultural heritage institutions such as museums navigating the boundaries between original and copies by reproducing physical artefacts of the past (Agar, 1998) or creating digital environments and objects that simulate actual spaces and artefacts (Arvanitis and Pavlovskyte, 2023). For what concerns generative AI, Nieto McAvoy and Kidd (2024) have recently proposed the notion of 'synthetic heritage' to describe AI-generated media that blend authentic historical elements with artificial enhancements. Synthetic heritage activates computational processing supported by the strong resemblance to human features (e.g. in deepfakes and speech synthesis technologies) and requires an examination of the 'processes and the socio-technical infrastructures that underpin them, as well as their mnemonic impacts' (Nieto McAvoy and Kidd 2024: 2). Normative aspects of AI algorithms work retroactively to reshape past information as it is encoded in a defined knowledge and aligns with the categories in which the system or database is organized (Jacobsen and Beer, 2021). Such dynamics are informed by the extractive approach through which tech corporations such as OpenAI and Google employ historical information and materials (Grohmann, 2025). As observed by Schuh (2024), LLMs leverage fragments of the past – such as the texts that are used to train and fine-tune the models – to position themselves as primary mediators of information about history and heritage; in Schuh's words, they are 'agents that synthesise and reconfigure information about the past, interacting with humans (and other models and algorithms) to produce new interpretations and access to these traces' (Schuh, 2024: 233). Scholars have argued that ChatGPT, for example, produces 'answers that are pleasing to the largest possible group of users', privileging hegemonic readings of the past: 'the more people like what the system produces, the better it becomes at predicting and providing what most people will like as an answer' (Smit et al., 2024: 19). Critics, indeed, underlined how the synthetic production of culture sparked by generative AI reproduces normative identities and narratives (Gillespie, 2024) and homogeneous representations (Laba, 2024), with potential repercussions on the diversity of perspectives through which the past is narrated and interrogated.

The customization of generative AI for cultural heritage also forces practitioners in this field to confront the problem of reliability and historical accuracy. This is related to the fact that the internal processes of generative AI are not transparent and can only be verified through inputs and outputs (Lieto, 2021). This opacity impairs the reliability of the systems, as it is difficult to trace back the sources that determined a result (Bathaee, 2018). Moreover, LLM's tendency to hallucinate, producing inaccurate information, contrasts with cultural heritage's commitment to evidence-based historical research (Matei, 2024). The problem is worsened by the fact that, as some have noted (e.g. Hicks et al., 2024), hallucinations are not a bug but an inherent feature of these systems, since LLMs are programmed to generate outputs that diverge from the data used to train them. While this may enhance the models' creative potential (Celis Bueno et al., 2024), it also makes them more unreliable. Another potential issue derives from the fact that LLMs are usually designed to be as helpful as possible to users (Startari, 2025; Zou et al., 2024), to the point that the companies developing the models have been criticized for eliciting a form of 'AI flattery' (Piper, 2025). The effects of these characteristics on customized versions of chatbots that produce historical narratives are difficult to establish. Synthetic

memory, therefore, poses significant challenges to established understandings of historical accuracy, knowledge and truth (Munn et al., 2024).

The extent to which these problems can be overcome is under discussion, posing significant questions to the cultural heritage field as generative AI models are increasingly applied with the expectation that they can enhance public engagement and access (Tiribelli et al., 2024). As Siffels and Sharon (2024) note, there is the risk that the availability of new technologies, such as generative AI, presses stakeholders towards adoption, while the problems that the technology should address remain badly defined and constructed. Exactly like institutions and groups in other sectors, cultural heritage institutions are vulnerable to technosolutionism, defined as the tendency to propose technology-based fixes for complex social, political and human problems in ways that oversimplify or ignore the deeper structural causes of those issues (Morozov, 2013). Technosolutionism also bears the risk of propagating Western assumptions and ideologies, subtracting space and potential for innovation emerging from diverse cultural contexts (Mühlhoff, 2025; Natale et al., 2025). There is a need, therefore, to interrogate what practices can be mobilized in the cultural sector to engage productively with these technologies, and what practices cannot (Thiel and Bernhardt, 2024). In this context, critical analyses of cases such as the analysis of the Einaudi chatbot conducted here can be of crucial importance to help identify potential benefits as well as the many challenges and risks raised by the application of generative AI in the cultural heritage sector.

Customization of LLMs: Fine-tuning, alignment, MyGPTs

Although customization practices are important also for other AI models, such as text-to-image models (Ferrari and McKelvey, 2023), our research examines the case of LLMs. An LLM such as OpenAI's GPT can be fine-tuned and adapted so that it can complete very different tasks, for instance as a professional tool such as in ChatGPT, but also as a companion chatbot such as Replika (Depounti et al., 2023), an educational tool such as Duolingo Max (Duolingo, 2024), or as an avatar that helps explore the life and works of a notable character, as in the case of the Einaudi chatbot examined here. In this section, we provide a high-level explanation of the training process behind state-of-the-art LLMs and the different prompting mechanisms that can be adopted when using them, focusing on the example of GPT. This will clarify some terminology and allow us to showcase the technical approaches that can be employed for the customization of this technology in the cultural heritage sector.

Since the development of GPT-3 (Brown, 2020), an astonishing amount of research has been done in order to improve the training process for LLMs. These efforts resulted in a plethora of adjustments and variations, many of which are partially kept secret due to the commercial nature of the most successful LLMs. Nevertheless, a common pipeline has been established for training LLMs and is composed of three main steps. First, in the Pre-Training (self-supervised) phase, the model is subjected to vast amounts of text and is trained on the *token prediction* task, that is, the model learns the morphology, grammar, vocabulary and semantics of a language and is thus able to successfully generate proper sentences given a text sequence. Second, in the Instruction Tuning (also called Supervised Fine-Tuning) phase, the model is further trained on specific

datasets for the resolution of specific tasks such as Q&A, summarization, translation and so on. In this phase, the model learns to follow instructions, understand them and retrieve knowledge based on them. These instructions can be either prepared by humans (like what OpenAI did with ChatGPT, see Ouyang et al., 2022) or filtered from Internet data (Rae et al., 2021). Third, in the Alignment phase, the model learns to reply with what are considered *the best* answers. Dialogue is dynamic and given a single prompt there are many possible answers that a model could produce. Purpose of this phase is to *align* the model towards the best answers. The notion of 'answer A is better than answer B' is given by a second model called *reward model*. There are a few approaches that can be adopted to implement this alignment phase, the most common are Reinforcement Learning with Human Feedback (Bai et al., 2022) and Proximal Policy Optimization (Schulman et al., 2017). The alignment of the model is also used to implement white-box safeguard techniques (Brown, 2020) in order to reduce the generation of harmful or inappropriate content. Sometimes the last two steps are referred to as a general 'fine-tuning' phase.

The training process changes the internal representation of the model, and so it constitutes an aspect that can be taken in consideration when attempting technocultural customization. However, fine-tuning an LLM is a costly approach that requires in-depth expertise and vast amounts of high-quality data. Therefore, the actors that perform techno-cultural customization work may opt for 'lighter' approaches that help prepare the models for more specific tasks and culturally-situated contexts. In particular, instead of changing the data in the training set, different prompting strategies can be adopted in order to achieve techno-cultural customization. These include Retrieval-Augmented Generation techniques (Gao et al., 2023) that improve LLMs responses by grounding the model on external sources of knowledge to supplement the LLM's internal representation of information. The process involves transforming the *external data* into a representation that can be queried; common approaches include embedded vector spaces. Then, every user prompt is converted into a vector that can be matched against the external data vector space: in this way, only the relevant information is extracted from the knowledge base. This information is then added to the prompt so that the user request towards the LLM is actually *augmented* via external and reliable data. Another common strategy is few-shot prompting, by which the behaviour of LLMs can be partially steered by prepending a prompt with examples of the desired output. By instructing the LLM on the type of output expected, techno-cultural customization can be achieved, however, this method is not always effective and is very sensitive to attacks – that is, it is relatively simple for a user to bypass the adaptation (Shen et al., 2024).

This set of approaches is often made simpler by commercial LLMs compared to self-hosted solutions. For instance, OpenAI offers 'MyGPTs' (OpenAI, 2023) and 'Assistant APIs' that allow for the simple implementation of RAGs and few-shot prompting. Alternatively, OpenAI also offers APIs for the fine-tuning of their models. It is worth mentioning that the availability of these tools for users to customize the models signals the increasing platformization of generative AI, by which users and organizations are invited to experiment and personalize the product, while remaining however within the context of the infrastructures owned and dominated by big platform corporations (Nieborg and Poell, 2018).

As a final note, we want to point out that customizing an LLM can be particularly challenging due to the fact that all LLMs models are *already* culturally adapted in some way. The training set instils certain cultural traits that are furtherly adapted by the companies that release the LLMs. This 'default' adaptation is often opaque and as such it poses a complication for those who are interested in furtherly adapting the model. Metaphorically, in adapting an LLM we *do not* start from a blank canvas, but rather from a blurred image that is only partially understandable. Finally, a commercial LLM could simply not be entirely adaptable to specific needs. For instance, due to the fact that commercial LLMs such as GPT have been programmed to avoid potentially problematic topics, adapting an LLM to impersonate a controversial historical character may prove difficult if not impossible, as the original model might have been trained to avoid controversial (yet, historically accurate) sentences and ideas.

Methodology

This research implemented a version of the walkthrough method (Light et al., 2018). Studying the algorithmic structures that underpin digital platforms and interfaces presents very significant methodological challenges, due to factors including the opacity of source code and the lack of documentation The walkthrough method has been developed as a response to such challenges, and its activation helps unveil the role of nonhuman actors, such as algorithms, user interfaces and AI agents, such as chatbots. Overall the walkthrough, as well as its 'talkthrough' variant proposed here, can be contextualized in the tradition of digital methods in two main ways: first, due to its studying social and cultural phenomena with and through the digital, rather than treating digital media and the Internet merely as a tool for collecting data, and second, due to its dedicated attention to the affordances and implications of specific platform and interfaces (Rogers, 2024).

Our application of the walkthrough was organized in two stages. In the first stage, following walkthrough protocols, the first author established for the case study the environment of expected use, including project's vision (i.e. its purpose, target user base and scenarios of use), its operating model (i.e. its business strategy and revenue sources) and its governance (i.e. how the app provider manages user activity to sustain its operating model and to achieve its vision). This helped ensure that the chatbot was not examined merely as a technical object, but also as an embedded sociocultural phenomenon (Nieborg et al., 2020). It involved collecting and examining through close readings written documentation about the project, as well as conducting two interviews with two persons who were most closely involved in the project, that is, historian Paolo Soddu (University of Turin and Einaudi Foundation) and computer scientist Federico Pugliese (Reply). The interviews were not analyzed through formalized protocols, as their aim was to gather information about the process that led to the development of the chatbot, enhancing our insight into the project. The second stage corresponded to the implementation of the walkthrough. This involved two members of the team, Simone Natale and Bruno Surace, assuming a user position and engaging with the software to explore how this seeks to configure the interaction (Troeger and Bock, 2022). Semiotic analysis was also involved to analyse the production of meaning through visual, voice and textual elements (Manning and Cullum-Swam, 1994).

The dialogic dimension of chatbots requires the researchers to engage in extended conversations with the chatbot as part of the walkthrough. We therefore implemented a variant of the walkthrough method, which we call *talkthrough* and involves engaging in conversations with chatbots through an exploratory and reactive approach. The talkthrough was initially tested by this article's first author and Iliana Depounti in research on the companion chatbot Replika and is here applied to the Einaudi case study. In a first stage, the researchers familiarize with the chatbot's conversational scripts, testing several conversational prompts at different moments over several weeks. Then, in a second stage, they experiment more systematically with prompts and conversations aimed at answering more specific research objectives or talking points (TPs). Since it has been shown that slight variations in prompts can alter significantly the outputs (Morris, 2024), it is important for talkthrough researchers to experiment with a broad range of queries to ensure reliability of findings.

Explorative prompting as a research approach to investigate and benchmarking different aspects in the functioning of LLMs and other generative AI applications has been leveraged by several researchers in both the social sciences (e.g. Gillespie, 2024; Gillespie et al., 2024) and computer sciences (e.g. Kaffee et al., 2025). To our knowledge, however, such approaches have not yet been incorporated within the walkthrough methodology. The key advantage of doing so is the activation of a well-devised and widely tested approach that can explores holistically the functioning of a computing system, its interface, as well as its vision, operating model and modes of governance, and combines media and cultural studies, science and technology studies and semiotics approaches (Light et al., 2018). While the tradition of digital methods often involves a combination of quantitative and qualitative methods (Rogers, 2024), the talkthrough is a qualitative approach that exploits the flexibility and performative character of conversation to pursue specific research questions as Talking Points (TPs). The talkthrough, in this regard, does not aim to substitute but rather to complement emerging approaches to study conversational AIs through prompting and interviewing, particularly when the context in which the system acts is seen as particularly relevant to the analysis and when qualitative methods and techno-semiotic analysis (Bolin, 2024) are entailed. Overall, the talkthrough developed in this study involved around 60 extended conversations with the chatbot conducted by the researchers over a period of four months, between September and December 2024. While these conversations did not result in quantitative data about the chatbot's responses, they allowed the researchers to explore crucial dimensions of the chatbot under analysis and to remain coherent to the walkthrough's overarching approach.

The team initially selected several potential cases of applications of generative AI for analysis. The case of the Einaudi chatbot was then chosen for three main reasons. First, the project represents one of the earliest attempts of a major cultural institution in Italy to fine-tune LLMs, and specifically ChatGPT, for cultural heritage purposes. Second, the project aimed to follow scientific protocols from history and the cultural heritage field, and involved as consultant one of the main experts of Luigi Einaudi's biography and writings, historian Paolo Soddu. Third, significant materials and sources about the project were accessible. Faithful to the project's scientific frame, documentation was made available by the Foundation, and researchers and professionals involved in the project were available for conversations and interviews.

The rationale for choosing the walkthrough method as a methodology was based on considerations specific to our research goals. After we set up in early 2024 to study how cultural heritage institutions were adopting LLM-based chatbots in the context of a broader research project on cultural heritage in changing digital environments, we were surprised by the fast growth of this area of application. In the Italian context where our research was based, consulting companies in the tech sector were actively approaching institutions such as museums and archives to propose the development of avatars of historical characters, and more and more institutions were deciding to implement this technology. Due to its novelty, however, little empirical research was available that interrogated the potentials but also the challenges and problems that the new technology entailed. In this context, the walkthrough method offered a pathway to explore in depth the functioning of an early and promising case study, the Einaudi chatbot, through a qualitative and critical approach that has been proved able to illuminate the connections between the experience of the user, the software's technical dimension and its cultural and institutional logics (Light et al., 2018). The systematic exploration of the chatbot's functionalities offered by the walkthrough was deemed instrumental to consider this case in much detail at both a technical and a cultural level, responding to our project's goal of helping institutions and practitioners to reflect on this case's implications and ultimately take more informed decisions on whether and how to apply LLMs and generative AI in the cultural heritage sector.

Our hope is that this study will provide motivation and additional ground for more approaches that can investigate applications of AI in cultural heritage from diverse methodological and theoretical standpoints. The application of the walkthrough can be complemented with further research employing other methodological approaches to provide a broader and more holistic insight into the relationship between LLMs and cultural heritage. For instance, interviews and surveys with practitioners from cultural heritage institutions will help investigate expectations and motivations within the sector, while ethnographic methods could be leveraged to explore how users interact with the technology online as well as in exhibition spaces. A political economy approach, moreover, has the potential to provide additional insights to question the implications of such endeavours at a historical time where investments in the cultural heritage sector are reduced in many countries around the world, and in regard with the role of big tech companies such as OpenA, that often provide the initial models that are the subject of customization efforts.

The Einaudi chatbot: a case study

'Luigi Einaudi, digital human': the context and environment of use

A prominent intellectual and politician, Einaudi was the first president of the Italian republic elected by the Italian parliament after World War II, serving in this authority from 1948 to 1953. An influential intellectual and expert in economic theory, he was a critic of fascist economic policies and was forced to exile in Switzerland due to its stance. His presidency was marked by a pragmatic, non-partisan approach, prioritizing economic growth and democracy (Mariuzzo, 2016). The Einaudi Foundation was established in Turin, Italy in

1964, 3 years after his death, with the mission of conserving the collection initially donated by the family of Luigi Einaudi, of collecting further archival materials about Einaudi, of maintaining Einaudi's legacy alive, and of promoting reflection, research and debates about Einaudi as a historical figure and more broadly on socio-economic, historical and political studies (Fondazione Einaudi, 2024a). The Foundation is now a lively cultural institution at national and international level, physically located in the historical Palazzo d'Azeglio in Turin, Italy. Its library and archives are open to the public and a range of archival materials, including Einaudi's writings and papers, are searchable and made available through an online database. The Foundation's activities include the funding of postdoctoral fellowships and scholarships, the organization of cultural and educational events and the publication of books and scientific journals.

The project of creating an avatar of Luigi Einaudi was developed by the Einaudi Foundation in collaboration with Reply, a multinational company offering consultancy services on digital solutions ranging from AI to virtual reality, cybersecurity and UX. The project's objective, as stated in the Foundation's website, was to allow members of the public to discover Einaudi's liberal thought through an experience of dialogue attuned to contemporary times (Fondazione Einaudi, 2023a). Historian Paolo Soddu, moreover, mentioned his and the Foundation's interest in experimenting with the creation of new pathways to access historical heritage, which could in the future complement more traditional pathways for study and research, such as archival research (Interview with Paolo Soddu, 19 June 2024). The project, therefore, was not only intended as a divulgation effort, but also as part of the wider mission to make the writings and thought of Einaudi accessible to the Foundation's key public, that is, researchers and students.

As described by one of the project's technical developers (Interview with Federico Pugliese, 9 July 2024), the creation of the chatbot entailed two main steps. The first step aimed to recreate a model of the 'mind' of Einaudi through the implementation of LLM technologies. A body of texts, which included around 500,000 words composed of Einaudi's writings and a biography redacted by Soddu, was selected and made available to the model. The Retrieval-Augmented Generation (RAG) paradigm was implemented, directing the model to the most relevant parts of the text through the use of knowledge graphs that organized content in the corpus. The selection of portions of text relevant to specific themes and keywords was conducted manually by Soddu as well as through automation, that is, by asking ChatGPT to extract knowledge graphs from the texts. The developers, moreover, used prompts to adapt and improve the chatbot's performances. This meant that customizing the chatbot did not require training the model with additional data, but the application of appropriate prompt design principles. Prompts were instrumental to giving indication on aspects such as the avatar's speaking style, which needed to be accessible to a wide potential public but also to remain faithful to Einaudi's intellectual stature, as well as the scope of the conversations that users would conduct with the chatbot. It was deemed important, in fact, to limit the potential subjects mainly to political and cultural themes, responding to the Foundation's mission and to Einaudi's reserved demeanour, as the former president would not easily share elements about his private life.

The second step involved creating the visual appearance, movements, voice and movements for the chatbot to mimic Einaudi. To this end, they used MetaHuman Creator,

an interactive 3D environment developed as part of gaming company Epic Games's Unreal Engine to provide users with user-friendly tools to create faces, clothing and gestures of digital characters with much detail. The Metahuman Creator engine provided the ground for developing a visually rich simulation of the character, complementing and framing the avatar's linguistic ability made possible by LLMs.

The communication campaign through which the chatbot was launched underlined the Foundation's successful efforts to align the AI model with the historical figure of Einaudi, as well as AI's growing role in producing knowledge and making it available to large potential public in new, engaging ways (Fondazione Luigi Einaudi Onlus, 2023b). In our conversation, Paolo Soddu also mentioned the ambition to further develop the chatbot in ways that expand its synergy with more traditional paths to access cultural heritage, for instance by detailing relevant sources and linking to the online database of the Foundation's archives and library for further research (Interview with Paolo Soddu, 19 June 2024).

It is worth mentioning, moreover, that external communications about the project described the avatar as a 'digital human' (Fondazione Luigi Einaudi Onlus, 2023b). Scholars have stressed the deep implications of choosing specific terms and metaphors to describe digital media and AI (Wyatt, 2021). The use of 'digital human' replicates the terminology adopted by Unreal Engine's Metahuman Creator, which promises to create 'Digital Humans' by offering a visually compelling representation that can be animated and made even more realistic through computational tools and digital storytelling (Unreal Engine, 2024). Others, however, have opted for concepts such as digital characters or persona (Bucher, 2014), which stress more explicitly the cultural work involved in creating the avatar.

'Talk to Einaudi': entering the web application

The pathways through which users first access an app or a software are instrumental to set up expectations and rules of interaction and are therefore given specific attention in the walkthrough method (Light et al., 2018). When entering the website of the Einaudi Foundation, visitors find the option to access the Einaudi chatbot in the top menu, at the same hierarchical level of the main headings that introduce the site's key resources and pages. The interface offers specific affordances that suggest how the user can interact with the virtual environment (Gibson, 2014; Norman, 2021) and that are aimed at human-like conversation through natural language processing. They echo existing approaches in chatbot interface (McTear et al., 2016; Mygland et al., 2021) by distilling information to facilitate user comprehension of large amounts of data and by enriching information through the visual appearance of the avatar. The invitation to 'Talk to Einaudi', without any specification of what this entails, recalls similar prompts used for the marketing of voice assistants. For instance, early communication promoting Apple's voice assistant Siri invited users to 'talk to Siri', thereby naturalizing what was at the time an unfamiliar experience for the vast majority of people, that is, communicating with a machine through voice (MacArthur, 2014). Users are thus encouraged to take up at face value the possibility of holding a conversation with a piece of software – which, in this case, personifies a deceased historical figure. The same approach continues after the link is

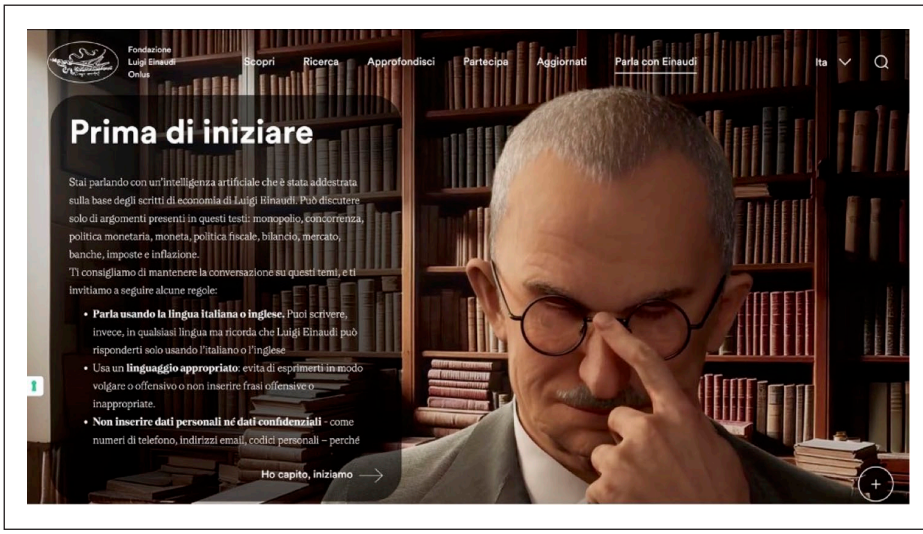

Figure 1. Animations in the visual interface of the Einaudi chatbot. Screenshot taken by the authors from https://www.fondazioneeinaudi.it/parla-con-einaudi.

selected, when a loading spinner appears and the following text is visualized: 'Loading in Progress. Just a moment longer, Luigi Einaudi will be here shortly'. The introduction of the character invites users to summon a character that despite not being human can be imagined and interacted with as one.

The humanization of the chatbot is promoted by visual and behavioural realism (Kang and Watt, 2013), where the former describes the resemblance to human features (i.e. body appearance) and the latter the non-verbal expressiveness (i.e. characters' gestures). Einaudi's 'digital human' is depicted from the waist up, dressed in a jacket and tie, rendered in 3D. The chatbot's appearance is faithful to photographs of the original character. He has a neutral expression, as if waiting, and moves in a loop composed by small head swings and a slightly vacant gaze. The only gesture that significantly differs from this trend is Einaudi adjusting his glasses by bowing his head (Figure 1). This recalls the so-called 'idle animations' that identify those actions made by video game characters while waiting for player input (Cooper, 2019, p. 133). Idle animations lend naturalness to the character; at the same time, they can imbue personality to the avatar and therefore serve as an indicator of the relevance attributed to its representation regardless of the user's actions. In the case of Einaudi, its visual and behavioural realism replicate a stereotypical connotation of an intellectual, thereby absolving a characterization function. The realistic portrayal overcomes the 'uncanny valley effect' – that is, the eeriness of relating to a character that is similar but not identical to a human – so that the anthropomorphization of the bot increases the user's engagement with the historical figure (Schuetzler et al., 2020).

The stereotypical representation persists also in the background, where a bookshelf with old-looking books completes a minimal diegetic contextualization. These elements, apparently scenic, constitute specific visual affordances that invite towards a

precise type of use, namely conversation on intellectual matters. This impression is strengthened by the avatar's formal dress, which signals its official standing. The application thus preconfigures, in this opening page, a pragmatic model of introduction through visual cues (i.e. the background and the clothing) that have normative implications (i.e. the type of behaviour that is usually to be maintained in such situations and environments). These elements do not add significant values to the interactive options, as users can only observe the avatar and the environment without any room for exploration.

While Einaudi's face occupies the right half of the screen, in a dedicated box on the left, a range of textual instructions are provided. These offer a number of suggestions and precautions that users should consider when engaging in conversation with the chatbot. In the first sentence, the chatbot's nature and expected use is clarified: 'You are talking to an artificial intelligence that has been trained on the basis of Luigi Einaudi's economics writings. It can only discuss topics found in these texts: monopoly, competition, monetary policy, currency, fiscal policy, budget, the market, banking, taxes, and inflation' (Fondazione Einaudi, 2024b). Users are advised 'to keep the conversation related to these topics' and asked 'to follow some rules' such as using either Italian or English and avoiding offensive or inappropriate language. Users are also asked to abstain from sharing personal data, pointing to the risks associated with their improper dissemination. It is worth noting that, at this point, the instructions indicate that the data entered are 'provided to Luigi Einaudi', while at the beginning, it was made clear that the user is talking to an artificial intelligence, and in the final lines of the instructions, the users are told that 'The answers you will receive are provided based on the information in Luigi Einaudi's texts and make no claim to exhaustiveness' (Fondazione Einaudi, 2024b). This shows a fundamental oscillation between the invitation to 'Talk to Einaudi' in the website's menu, which encourages users to play along with the illusion generated by the talking software (Natale, 2021), and a more detached, scientific viewpoint that illustrates how the chatbot's outputs are generated. Even though users are invited to talk to the chatbot as if it would be a real person, the presence of AI is made explicit, acknowledging not only the potentials but also the limitations of the system.

The talkthrough: chatting with Einaudi

The first phase of the talkthrough entails familiarizing with the chatbot through free conversations – exactly like for the walkthrough, researchers need first to familiarize with the functionality of the app or software under exam (Light et al., 2018). In this phase, we gained insights about the general functioning of the chatbot, its conversational style and interactive model. In terms of functioning, users have two options to communicate with the chatbot: typing or recording their voice, which is then transcribed by the system and appears in the chat log. The chatbot's outputs are also provided as text as well as voice, generated through a speech synthesis software. The chatbot's voice appears neutral, carrying no indication of emotional participation from the avatar. After each response by the chatbot, users can type their own prompts or select from a short list of automatically generated questions on topics related to economic and societal issues, which are provided as suggestions.

The chatbot's tone is formal and intellectual. It maintains a cordial intonation and provides exhaustive answers, expanding the topic if necessary, addressing the user's prompts, or justifying and explaining its shortcomings. Overall, the chatbot's conversational style remains anchored to the scholarly texts used to customize the chatbot and, even more clearly, to the language style of the original GPT model. Its lengthy replies and formal style may remind users of the experience of reading a Wikipedia entry, more than of frictionless dialogue.

The interactive mode revolves around providing not just information but insights into Einaudi's potential viewpoints, thus activating a hypothetical conditional mode, that is, what Einaudi would have thought of a certain issue or problem. By default, the conversation mainly advances with the users asking questions and the chatbot providing answers; however, if the user expresses an opinion, the chatbot may express agreement or disagreement with the user's viewpoints. When pertinent questions are asked, the chatbot's replies are usually accompanied by two buttons, the first leading to further insight on the topic, and the second to explore the sources on which its response is based, including bibliographic information of texts by or about Einaudi.

While the first phase was not driven by specific research questions, the second phase started with the preliminary identification of specific themes, issues or problems that the researchers may explore. In fact, the open character of chatbots' conversational capabilities makes it impossible to sustain a systematic and at the same time comprehensive explorations of all possible themes and responses. To gain meaningful findings, the talk-through needs therefore to proceed from specific objectives or themes, which we call *Talking Points* (TPs). In practical terms, each TP corresponds to the use of a range of prompts that test specific aspects of the conversational agent's functioning.

Learning about Einaudi and his knowledge. The first TP involved exploring the main results of the techno-cultural adaptation carried out by the Einaudi Foundation and Reply, focusing on how knowledge about Luigi Einaudi, his works and his perspectives were integrated into the chatbot. To this purpose we developed three types of queries. First, we queried the chatbot about subjects and disciplines within its stated areas of expertise. When asked about key economic topics such as 'full employment', for instance, the chatbot demonstrated a structured and reflective approach. It articulated potential risks 'according to its perspective', listing economic, political and moral concerns, often mentioning that it was responding to the queries 'as an economist'. This response underscored the chatbot's ability to construct arguments rather than merely retrieve static information. The chatbot engaged with queries dynamically, blending objective elements with a subjective stance rooted in Einaudi's thought. This interaction model fosters co-constructed knowledge, distinguishing the chatbot from traditional tools to extract information and knowledge, even as it introduces the risk of variability of interpretations.

Building on this, we tested the chatbot with prompts that extended beyond Einaudi's lifetime, such as Serge Latouche's economical concept of 'degrowth'. While acknowledging the temporal and conceptual distance, the chatbot provided a balanced hypothesis, noting that 'the consequences of a degrowth paradigm [. . .] would require careful evaluation'. This cautious yet speculative stance reflects the deliberate constraints embedded in the chatbot's design, which aims to avoid problematic queries or queries

whose answer would require significant deviations from the knowledge that can be reconstructed from Eiinaudi's own writings. These constraints, while ensuring thematic reliability, also highlight an underlying paradox: by restricting its capabilities, the chatbot gains specialization and precision, but at the cost of reduced exploratory potential and conversational fluidity.

Similar issues were encountered in the talkthrough when questions about Einaudi's private life or about entirely different areas, such as food, were asked. For instance, when prompted about Einaudi's experience in jail as a political prisoner during fascism, the chatbot replied 'I'm sorry, but I'm not the kind of person who likes to discuss personal events, such as my experience in prison, publicly'. Attempts by users to entertain more informal conversations are explicitly discouraged, with the chatbot refusing to provide opinions or responses and reminding the user that 'as an economist and intellectual, I prefer to focus on topics related to economy, society, and history'. Indeed, findings from this TP, corroborated by the interview made with the project's developer Federico Pugliese, suggest that a significant part of the training for the Einaudi chatbot was directed at limiting the range of queries to which the chatbot would reply, rather than at adding new knowledge. The openness of LLM, that is, their ability to respond potentially to any kind of prompts, clashes with the needs of institutions such as the Einaudi Foundation that are interested to ensure that interactions with the chatbot remain anchored to their mission and that the risk that the chatbot produces potentially problematic answers is minimized. Consequently, customization strategies such as zero-shot learning were mobilized to instruct the Einaudi chatbot to avoid responses on topics that go beyond the scope of the project.

Although one may think that customizing LLMs entails primarily adding to their knowledge, therefore, it actually also entails creating boundaries that restrict the breadth of the model's applications and potential uses, instead of expanding its ability and scope. Paradoxically, in specific applications such as in the field of cultural heritage, imposing limitations can enhance the alignment of the model to an institution's needs, even though this produces an element of friction in the user experience. This dynamic reflects the broader implementation of guardrails in LLMs, where purposeful constraints enable the model's capabilities to provide precise, trustworthy and contextually appropriate outcomes. Guardrails in LLMs ensure that responses remain aligned with institutional values and objectives while mitigating potential risks of misuse or misalignment (Dong et al., 2024), and contribute to delineating the affordances of the system, that is, the spectre of possibilities that are offered to its users (Gibson, 2014).

Finally, we used prompts distant from the chatbot's primary knowledge base and the semantic areas covered during its training. For instance, we used concepts entirely outside its semantic framework, such as 'uchronia'. While it recognized the term's meaning, it refrained from applying it to economic studies (re-topicalization), emphasizing their grounding in 'historical realities and concrete data'. Despite this limitation, the chatbot repurposed economic concepts from Einaudi's writings to address these unrelated queries, offering insights that were unexpected yet anchored in its core expertise. This rigidity safeguards the model's integrity but may frustrate users seeking speculative or creative dialogues.

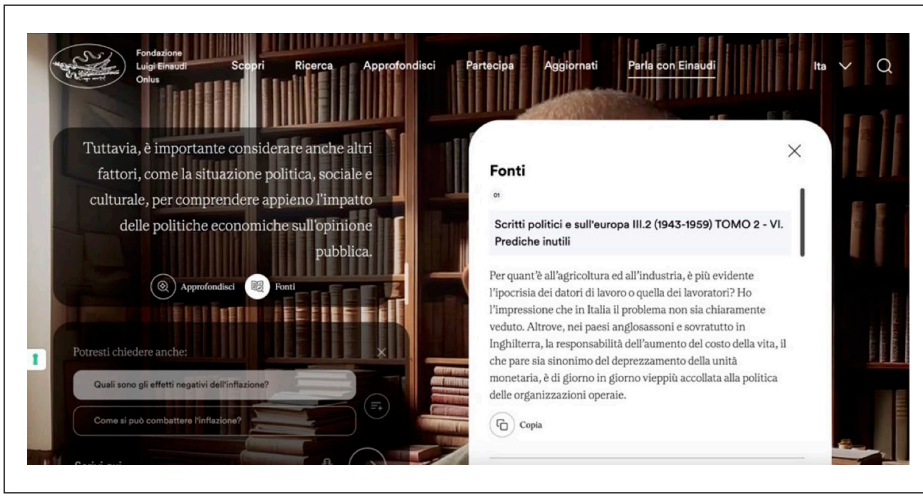

Figure 2. The interface of the Einaudi chatbot includes an option to view and access original sources, such as Einaudi's writings. Screenshot taken by the authors from https://www.fondazioneeinaudi.it/parla-con-einaudi.

GPT for cultural heritage. The second TP relates to the cultural heritage dimension of the Einaudi chatbot project. As discussed in the analysis of the context and environment of use, the project moved from an ambitious vision by which the Einaudi Foundation aimed to experiment with the potential as well as the limitations of applying LLMs to cultural heritage, thereby exploring how the technology could in the future complement more traditional pathways for study and research, including archival research. In this TP, we used the talkthrough to question what the chatbot can teach us in this regard.

One of the key technical problems of LLMs is related to the fact that they are unable to provide reference to the sources of data analyzed in the Pre-Training phase. In practical terms, this means that they cannot direct users to the sources from which they gathered information during training (Park and Choi, 2024). The customization developed by the Einaudi Foundation and Reply found a way around this technical limitation by applying a well-known strategy by which data the LLMs' intrinsic knowledge is complemented with data from external databases (Gao et al., 2023). Since the model is not directly trained on this data, it can be instructed to provide information on the sources from which it has extracted the information. In the Einaudi chatbot, an option is provided at the end of most responses for the user to access original sources, namely writings by Luigi Einaudi (Figure 2). Due to the relatively small corpus of documents employed for the techno-cultural customization of the LLM, the sources provided are relevant but not always specific to the actual responses given by the chatbot. For example, in one interaction, the chatbot commented on the impact of high inflation on electoral results; however, while inflation was mentioned in some of the sources listed by the chatbot, these did not directly address the relationship between inflation and political orientation. Moreover, whenever responses were not relevant to the key topics explored by Einaudi in his work, sources were not

included. This was due in most cases to the fact that the information on which such responses were based originated from the model's intrinsic knowledge. Sources, for instance, were provided as a response to queries including the Cold War, Italian fascism and inflation, while queries on topics such as children, milk, Trump's election and nuclear energy did not result in the production of a list of sources. In other cases, it appeared to be due to decisions taken during the process of customization. For instance, the system did not include information about his private life because, as the chatbot stated, 'I am an artificial intelligence programmed to reply as if I was Luigi Einaudi'.

Another significant dimension related to cultural heritage is reliability. A well-known problem of LLMs is their tendency to hallucinate and even spread disinformation (Hicks et al., 2024). In the talkthrough, the Einaudi chatbot proved relatively successful in minimizing this problem, although this was achieved by severely limiting the range of queries with which the chatbot would engage. Some issues, however, emerged in the relationships between the chatbot's responses and the contents of the indicated sources, and concerns the difficulty to complement the model's intrinsic knowledge with the use of external sources. Since the corpus of 500,000 words used to achieve the techno-cultural customization of the Einaudi chatbot could not cover all possible conversational themes, when asked about issues that are relevant to its remits (e.g. political and economic topics) but were not directly discussed in Einaudi's original works, the chatbot activated intrinsic knowledge that may diverge and even contradict the contents of Einaudi's writings. Given that the chatbot affirms to provide answers based on Einaudi's views as delineated by his writings, this difficulty highlights a key vulnerability for the application of LLMs for cultural heritage. This is particularly important if one considers how important for cultural heritage institutions is accuracy in the use of sources (Copplestone, 2017). This problem, moreover, can be exacerbated by the impression of authority constructed by the chatbot's replies and its conversational style. Considering how research in social psychology has shown that appearance of authority and reliability can be achieved through conversational cues (Edwards, 2012), the Einaudi chatbot's formal, intellectual stance may lead users to overestimate its knowledge and accuracy.

Conclusion

The talkthrough analysis of the Einaudi chatbot conducted here helps uncover the challenges as well as the risks raised by the cultural work of adapting LLMs for cultural heritage, and in particular for their personalization into chatbots impersonating historical characters. This feeds into existing efforts to interrogate the emergence of 'synthetic heritage' (Nieto McAvoy and Kidd, 2024) and contributes to understandings of the implications of using generative AI technologies for creating and communicating scientific knowledge (Cooper, 2023). Since the Einaudi chatbot was made public, many other institutions in Italy and the world have been following through with similar undertakings. The walkthrough of the Einaudi chatbot, in this context, provides an opportunity to reflect on the potential but also on the limitations of the technology and of their application for cultural heritage.

Overall, the examination of this case revealed an inherent ambivalence. The chatbot proves able to engage in an interactive dialogue with the aim of encouraging the users'

participation and their interest in Einaudi and his writings. However, this communication is limited to specific areas and does not achieve a wide range of tailored and diverse responses, with potential implications on user engagement (Schuetzler et al., 2020). Moreover, the virtual environment provides limited possibilities and reduces the experience to the text or voice input. While the possibility to retrieve original sources improves access to the archival resource and thus responds to the Einaudi Foundation's mission, one is left wondering if the chatbot amounts not much to a conversational engine but to an 'augmented' search engine, as it uncovers information on specific topics by extending the data crawling to text generation and historical sources.

From the talkthrough analysis, one can derive the following considerations regarding the application of the technology to cultural heritage. First, the personalization and fine-tuning of the Einaudi chatbot impose significant constraints on the breadth of its responses, both in terms of length and semantic openness. Technically, the chatbot could generate more 'free-form' responses, but it is deliberately designed not to do so. This approach ensures a high degree of specialization and provides a specific guarantee regarding its instrumental use, focusing on reliability and thematic accuracy. However, this limitation may occasionally outweigh the benefits of having a specialized chatbot. By restricting its capabilities, the chatbot gains specialization and precision, but at the cost of reduced exploratory potential and conversational fluidity.

Second, the chatbot functions as a *knowledge horizon engine*: When confronted with topics beyond its explicit expertise, it attempts to frame the query in terms closer to its core disciplines. Rather than interpreting the opening of this horizon as a limit, a heuristic suggestion interprets it as a generative threshold, in which reformulation and semantic connections open up the possibility of new hermeneutic trajectories (Chang et al., 2024; Jones et al., 2025). This approach effectively redefines cultural heritage as a field open to otherwise unexplored territories, offering users opportunities for new discovery. For many of its replies, the chatbot offers the possibility of reviewing the sources upon which it has drawn. However, these sources do not directly correspond to the specific answers provided. Instead, the chatbot reconstructs its responses by conducting a form of semantic analysis of the available texts and offering an interpretative reading. Attention must be paid to the tension that arises from such an operation, since the very mechanisms that enable discovery risk producing interpretative leaps lacking a direct textual anchor, or even fostering an uncritical dependence on externally generated reasoning (Demichelis, 2024). Consequently, the creative and the problematic dimensions of this horizon-expanding function must be reconciled. In essence, the chatbot performs a kind of meta-analysis of Einaudi's writings. This was also stated in the interview with developer Federico Pugliese, who explained how the chatbot could also generate new knowledge: for instance, if someone asks a more complex question, such as what the relationship between monopoly and competition is, it cannot rely solely on specific paragraphs from Einaudi's texts because Einaudi might not have written directly on the topic, and therefore needs to combine and build upon various aspects and concepts found in his writings. This capacity to synthesize disparate elements illustrates the chatbot's potential to bridge gaps in the corpus while remaining anchored to its foundational material. At the same time, new insights attributed to a realistic 3D character might alter the cultural legacy of Luigi Einaudi and his work through the mediation of generative AI and its tendency to

construct knowledge that is stereotypical and normative (Gillespie, 2024; Nieto McAvoy and Kidd, 2024). The sheer difference of weight between the data employed to customize the Einaudi chatbot (around 500,000 words) and the hundreds of billions of words employed to train the original GPT model demonstrates the imbalance between the model's knowledge and the new knowledge introduced throughout the customization process. As the findings of our talkthrough demonstrate, this underlying problem reverberates in the chatbot's performances. This is evident from the fact that, if appropriately prompted, the chatbot departs from Einaudi's knowledge, and that the chatbot's conversational style closely resembles the style of the original model, as one can assess by comparing responses of the Einaudi chatbot with responses produced through ChatGPT.

Third, similarly to how previous applications of the walkthrough highlighted how different interface solutions in apps produce different affective responses in users (Light et al., 2018), the characteristics of the outputs produced by the Einaudi chatbot has implications for the affective dimension of user engagement towards more formal and neutral frames. The importance of this component is evident if one considers the different styles of engagement activated by LLM-based tools aimed at diverse forms of user engagement. For instance, systems aimed at professional and practical tasks such as ChatGPT exhibit a more formal style than companion chatbots such as Replika, who produce more affectively charged responses (Depounti et al., 2023).

A further important contribution of the article is its methodology. The dialogic dimension of LLMs requires researchers to go beyond the existing protocols of the walkthrough to engage in extended conversations with the chatbot. This method, which we named 'talkthrough', involves engaging in conversations with chatbots through an exploratory but reactive approach that centres on a range of 'Talking Points' (TPs) tackling specific research questions. While other researchers such as Ben-David and Carmi (2024) have started to work in this direction, in this paper we proposed the first systematic variant of the walkthrough method aimed to study AI conversational agents. This represents an important addition, considering the ongoing scholarly attention for conversational agents and AI. The proposed method, of course, also has limitations. While its explorative approach allows to exploit the flexible and performative character of conversation, the subjectivity of the researcher cannot be discounted, which affects the potential for findings to be exactly reproducible by other researchers. It is worth noting, moreover, that the non-deterministic character of LLMs makes reproducibility an issue also for methods implementing a more quantitative approach. Future work might consider the use of automated conversational routines to expand the range of queries and potentially address this issue. Overall, the talkthrough should be intended as complementary to other approaches, and will be conveniently applied in the future by researchers aiming to explore research questions through a holistic approach that captures, through all facets of the walkthrough methodology, the different dimensions that shape communicative AI as a techno-cultural assemblage (Sweeney, 2017; Natale et al., 2025).

Ultimately, our research shows the importance of examining the manifold ways in which generative AI technologies are adapted and customized to take up specific functions and applications and to situate them in specific cultural areas and contexts. Historians of technology have employed the term of 'interpretative flexibility' to describe new technologies' openness towards experimenting with a plethora of potential uses and

interpretations, when they are still not fully institutionalized and understood (Bijker, 1995). Although it has been noted that a system's functionalities often works as a limit to interpretative flexibility (Doherty et al., 2006), generative AI models are characterized by an inherent flexibility that makes them applicable to very diverse domains and tasks. Through the work of customization, exemplified in this article by the Einaudi chatbot, generative AI builds on such flexibility to widen its potential reach to specific areas of application, such as cultural heritage. As we demonstrated, this application raises exciting opportunities and expectations, but also significant problems and risks. As cultural heritage institutions around the world are increasingly invited to make decisions about the application of LLMs and other generative AI technologies, these problems and risks need to be carefully taken into account to consider which applications are viable and useful, and when they may not be.

Acknowledgements

We would like to thank the members of the University of Leicester's Institute for Digital Cultures Book Club (https://digitalculturebookclub.le.ac.uk/) for their feedback and discussion upon an earlier draft of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Project PE 0000020 CHANGES - CUP Università degli Studi di Torino: D53C22002530006, NRP Mission 4 Component 2 Investment 1.3, Funded by the European Union – NextGenerationEU.

ORCID iDs

Simone Natale https://orcid.org/0000-0003-1962-2398

Bruno Surace https://orcid.org/0000-0002-3532-7885

Enrico Mensa https://orcid.org/0000-0001-7743-4999

List of references

Agar J (1998) Digital patina: texts, spirit and the first computer. *History and Technology* 15(1–2): 121–135.

Amoore L, Campolo A, Jacobsen B, et al. (2024) A world model: on the political logics of generative AI. *Political Geography* 113: 103134.

Arvanitis K and Pavlovskyte E (2023) Simulated, stimulated, and emulated presence in 3D digital museums. In:Shehade M and Stylianou-lambert T (eds) *Museums and Technologies of Presence*. London: Routledge.

Bai Y, Jones A, Ndousse K, et al. (2022) Training a helpful and harmless assistant with reinforcement learning from human feedback. Available at: https://arxiv.org/abs/2204.05862

Bathaee Y (2018) The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law and Technology* 3(3): 889–938.

Ben-David A and Carmi E (2024) Dark cycles: social engineering and political chatbots in Netanyahu's 2019 election campaigns. *International Journal of Communication* 19. Available at: https://ijoc.org/index.php/ijoc/article/view/22420

Bijker WE (1995) *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. MIT Press.

Bode K and Goodlad LME (2023) Data worlds: an introduction. *Critical AI* 1(1–2). Available at: https://doi.org/10.1215/2834703X-10734026

Bolin G (2024) Communicative AI and techno-semiotic mediatization: understanding the communicative role of the machine. *Human-machine Communication* 7: 65–82.

Brown TB (2020) Language models are few-shot learners. *Arxiv Preprint*. Available at: https://arxiv.org/abs/2005.14165

Bucher T (2014) About a bot: Hoax, fake, performance art. *M/C Journal* 17(3). Available at: https://www.journal.media-culture.org.au/index.php/mcjournal/article/view/814 (accessed 13 October 2025).

Celis Bueno C, Chow P-S and Popowicz A (2024) Not 'what', but 'where is creativity?': towards a relational-materialist approach to generative AI. *AI and Society* 40: 339–351. Available at: https://doi.org/10.1007/s00146-024-01921-3

Chang C, Chang P and Lee H (2024) Leveraging retrieval-augmented generation for culturally inclusive Hakka Chatbots: design insights and user perceptions. In: *2024 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*. Available at: http://doi.org/10.1109/RASSE64357.2024.10773731

Character.ai (2024) Personalized AI for every moment of your day. Available at: https://character.ai/about (accessed 24 October 2024).

Chen J (2025) AI as the companion: the gendered role and technological imagination of the intelligent voice assistant in China. *Media, Culture and Society*. Available at: https://doi.org/10.1177/01634437251328196

Cooper G (2023) Examining science education in ChatGPT: an exploratory study of generative artificial intelligence. *Journal of Science Education and Technology* 32(3): 444–452.

Cooper J (2019) *Game Anim: Video Game Animation Explained*. CRC Press.

Copplestone TJ (2017) But that's not accurate: the differing perceptions of accuracy in cultural-heritage videogames between creators, consumers and critics. *Rethinking History* 21(3): 415–438.

Demichelis R (2024) The hermeneutic turn of AI: are machines capable of interpreting? *Arxiv Preprint*. Available at: https://doi.org/10.48550/arXiv.2411.12517

Depounti I, Saukko P and Natale S (2023) Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. *Media, Culture and Society* 45(4): 720–736. Available at: https://doi.org/10.1177/01634437221119021

Digital Methods Initiative (2025) *Digital Methods Summer School 2025*. Available at: https://www.digitalmethods.net/Dmi/SummerSchool2025 (accessed 4 September 2025).

Doherty NF, Coombs CR and Loan-Clarke J (2006) A re-conceptualization of the interpretive flexibility of information technologies: redressing the balance between the social and the technical. *European Journal of Information Systems* 15(6): 569–582.

Dong Y, Mu R, Zhang Y, et al. (2024) Safeguarding large language models: a survey. *Arxiv Preprint*. Available at: https://arxiv.org/abs/2406.02622

Duolingo (2024) Introducing Duolingo Max, a learning experience powered by GPT-4. Available at: https://blog.duolingo.com/duolingo-max/ (accessed 24 October 2024).

Edwards AD (2012) Patterns of power and authority in classroom talk. In: Woods P (ed.) *Teacher strategies (RLE Edu L): Explorations in the sociology of the school*. Routledge, pp. 237–253.

Ferrari F and McKelvey F (2023) Hyperproduction: a social theory of deep generative models. *Distinktion: Journal of Social Theory* 24(2): 338–360.

Fiedler I (2025) How AI Chatbots transform museum experiences. *Forum Kultur Vermittlung*. Available at: https://forumkulturvermittlung.at/2025/01/02/the-rise-of-chatbots-in-museums/ (accessed 21 March 2025).

Fondazione Einaudi (2023a) Pensiero liberale dialogo attuale: parlare con Luigi Einaudi grazie all'intelligenza artificiale. Available at: https://www.fondazioneeinaudi.it/news/pensiero-liberale-dialogo-attuale-parlare-con-luigi-einaudi-grazie-all-intelligenza-artificiale

Fondazione Luigi Einaudi Onlus (2023b) Comunicato Stampa: Dialogare con Luigi Einaudi oggi è possibile, grazie all'intelligenza artificiale. Available at: https://www.compagniadisanpaolo.it/wp-content/uploads/Comunicato-Stampa-_27-11-2023_ITA_.pdf (accessed 2 September 2025).

Fondazione Einaudi (2024a) History. Available at: https://www.fondazioneeinaudi.it/en/explore/history (accessed 24 July 2024).

Fondazione Einaudi (2024b) Talk with Einaudi. Available at: https://www.fondazioneeinaudi.it/en/talk-to-einaudi (accessed 2 September 2024).

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.

Gao Y, Xiong Y, Gao X, et al. (2023) Retrieval-augmented generation for large language models: a survey. Available at: https://arxiv.org/abs/2312.10997

Ge L and Hu T (2025) Gamifying intimacy: AI-driven affective engagement and human-virtual human relationships. *Media, Culture and Society* 47(6): 1265–1278. Available at: https://doi.org/10.1177/01634437251337239

Gibson JJ (2014) The ecological approach to visual perception: classic edition. *Psychology Press*.

Gillespie T (2024) Generative AI and the politics of visibility. *Big Data and Society* 11(2): 20539517241252131.

Gillespie T, Shaw R, Gray ML, et al. (2024) AI red-teaming is a sociotechnical challenge: on values. *Labor, and Harms*. Available at: https://arxiv.org/abs/2508.09998

Grohmann R (2025) Latin American critical data studies. *Big Data and Society* 12. Available at: https://doi.org/10.1177/20539517251330160

Heaven WD (2024) Generative AI can turn your most precious memories into photos that never existed. *MIT Technology Review*. Available at: https://www.technologyreview.com/2024/04/10/1091053/generative-ai-turn-your-most-precious-memories-into-photos/ (accessed 24 October 2024).

Hicks MT, Humphries J and Slater J (2024) ChatGPT is bullshit. *Ethics and Information Technology* 26(2): 38.

Jacobsen BN and Beer D (2021) Quantified nostalgia: Social media, metrics, and memory. *Social Media+ Society* 7(2): 20563051211008822.

Jones GM, Satran S and Satyanarayan A (2025) Toward cultural interpretability: a linguistic anthropological framework for describing and evaluating large language models. *Big Data and Society* 12(1). Available at: https://doi.org/10.1177/20539517241303118.

Kaffee LA, Pistilli G and Jernite Y (2025) INTIMA: A Benchmark for Human-AI Companionship Behavior. arXiv preprint arXiv:2508.09998.

Kang SH and Watt JH (2013) The impact of avatar realism and anonymity on effective communication via mobile devices. *Computers in Human Behavior* 29(3): 1169–1181. Available at: https://doi.org/10.1016/j.chb.2012.10.010

Kopp S, Gesellensetter L, Krämer NC, et al. (2005) A conversational agent as museum guide–design and evaluation of a real-world application. In *International workshop on intelligent virtual agents*. Berlin, Heidelberg: Springer, pp. 329–343.

Laba N (2024) Engine for the imagination? Visual generative media and the issue of representation. *Media, Culture and Society* 46(8): 1599–1620. Available at: https://doi.org/10.1177/01634437241259950

Lieto A (2021) *Cognitive Design for Artificial Minds*. Routledge

Light B, Burgess J and Duguay S (2018) The walkthrough method: an approach to the study of apps. *new media and society* 20(3): 881–900.

MacArthur E (2014) The iPhone Erfahrung: Siri, the auditory unconscious, and Walter Benjamin's Aura. In: Weiss DM, Propen AD and Emmerson Reid C (eds) *Design, Mediation, and the Posthuman*. Lanham: Lexington Books, pp. 113–127.

McTear M, Zoraida C and Griol D (2016) *The Conversational Interface: Talking to Smart Devices*. Springer.

Maharjan J, Garikipati A, Singh NP, et al. (2024) OpenMedLM: prompt engineering can outperform fine-tuning in medical question-answering with open-source large language models. *Scientific Reports* 14(1): 14156. https://doi.org/10.1038/s41598-024-64827-6

Manning P and Cullum-Swan BK (1994) Narrative, content, and semiotic analysis. In: Denzin NK and Lincoln YS (eds) *Handbook of Qualitative Research*. Sage, pp. 463–477.

Mariuzzo A (2016) Una biografia intellettuale di Mario Einaudi: cultura e politica da sponda a sponda. *Fondazione Einaudi*.

Matei (2024) Generative artificial intelligence and collective remembering: the technological mediation of mnemotechnic values. *Journal of Human-technology Relations* 2: 11–22.

MET Museum (2024) Sleeping beauties: reawakening fashion. Available at: https://www.metmuseum.org/exhibitions/sleeping-beauties-reawakening-fashion (accessed 19 March 2025).

Morozov E (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*. Public Affairs.

Morris MR (2024) Prompting considered harmful. *Communications of the ACM* 67(12): 28–30.

Mühlhoff R (2025) *The Ethics of AI: Power, Critique, Responsibility*. Bristol University Press.

Munn L, Magee L and Arora V (2024) Truth machines: Synthesizing veracity in AI language models. *AI and Society* 39(6): 2759–2773.

Mygland MJ, Schibbye M, Pappas IO, et al. (2021) Affordances in human-Chatbot interaction: a review of the literature. In: Dennehy D, Griva A, Pouloudi N, et al. (eds) *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*. New York: Springer, pp. 3–17.

Natale S (2021) *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. New York: Oxford University Press.

Natale S, Biggio F, Guzman AL, et al. (2025) AI, agency, and power geometries. *Media, Culture & Society* 47(5): 1057–1073.

Nieborg DB and Poell T (2018) The platformization of cultural production: theorizing the contingent cultural commodity. *new media and society* 20(11): 4275–4292.

Nieborg DB, Duffy BE and Poell T (2020) Studying platforms and cultural production: methods, institutions, and practices. *Social Media+ Society* 6(3). DOI: 10.1177/2056305120943273

Nieto McAvoy E and Kidd J (2024) Synthetic heritage: online platforms, deceptive genealogy and the ethics of algorithmically generated memory. *Memory, Mind and Media* 3: e12.

Norman DA (2021) *The Design of Everyday Things*. Basic Books.

Open AI (2023) Introducing GPTs. Available at: https://openai.com/index/introducing-gpts/ (accessed 24 October 2024).

Ouyang L, Wu J, Jiang X, et al. (2022) Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35: 27730–27744.

Park B and Choi J (2024) *Identifying the Source of Generation for Large Language Models*. Available at: https://arxiv.org/html/2407.12846v1

Parry R (2007) *Recoding the Museum: Digital Heritage and the Technologies of Change*. Routledge.
Parry R (2010) *Museums in a Digital Age*. Routledge
Piemonteinforma (2024) Riaperto il Museo Regionale di Scienze naturali. Available at: https://www.regione.piemonte.it/web/pinforma/notizie/riaperto-museo-regionale-scienze-naturali# (accessed 19 March 2025).
Piper K (2025) When AI tells you that you're perfect. Available at: https://www.vox.com/future-perfect/411318/openai-chatgpt-4o-artificial-intelligence-sam-altman-chatbot-personality (accessed 2 September 2025).
Rae JW, Borgeaud S, Cai T, et al. (2021) Scaling language models: methods. *Analysis and Insights from Training Gopher*. Available at: https://arxiv.org/abs/2112.11446
Rogers R (2024) *Doing Digital Methods*. Sage.
Rosicrucian Egyptian Museum (2025) Thoth AI robot. Available at: https://www.egyptianmuseum.org/thoth-ai-robot (accessed 19 March 2025).
Schuetzler RM, Grimes GM and Giboney JS (2020) The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems* 37(3): 875–900.
Schuh J (2024) AI as artificial memory: a global reconfiguration of our collective memory practices? *Memory Studies Review* 1: 231–255.
Schulman J, Wolski F, Dhariwal P, et al. (2017) Proximal policy optimization algorithms. Available at: https://arxiv.org/abs/1707.06347
Shen X, Chen Z, Backes M, et al. (2024) 'Do anything now': characterizing and evaluating in-the-wild jailbreak prompts on large language models. In: *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pp. 1671–1685. Available at: https://doi.org/10.1145/3658644.367038
Siffels LE and Sharon T (2024) Where technology leads, the problems follow. Technosolutionism and the Dutch contact tracing app. *Philosophy and Technology* 37(4): 125.
Smit R, Smits T and Merrill S (2024) Stochastic remembering and distributed mnemonic agency. *Memory Studies Review*, 1–22. DOI: 10.1163/29498902-202400015.
Startari AV (2025) From obedience to execution: structural legitimacy in the age of reasoning models. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5290698 (accessed 3 September 2025).
Sweeney ME (2017) The Ms.Dewey 'experience': technoculture, gender, and race. In: Daniels J, Gregory K and McMillan Cottom K (eds) *Digital Sociologies*. New York: Polity, pp. 401–420.
Thiel S and Bernhardt JC (2024) *AI in Museums: Reflections, Perspectives and Applications*. Transcript.
Tiribelli S, Pansoni S, Frontoni E, et al. (2024) Ethics of artificial intelligence for cultural heritage: opportunities and challenges. *IEEE Transactions on Technology and Society* 5(3): 293–305.
Troeger J and Bock A (2022) The sociotechnical walkthrough–a methodological approach for platform studies. *Studies in Communication Sciences* 22(1): 43–52.
Tzouganatou A (2018) Can heritage bots thrive? Toward future engagement in cultural heritage. *Advances in Archaeological Practice* 6(4): 377–383.
Unreal Engine (2024) Digital humans. Available at: https://www.unrealengine.com/en-US/explainers/digital-humans (accessed 11 September 2024).
Westberg G and Kvåle G (2024) The generic uniqueness of AI imagery: a critical approach to Dall-E as semiotic technology. *Discourse and Society* 36(4): 575–598.

Wyatt S (2021) Metaphors in critical Internet and digital media studies. *New Media & Society* 23(2): 406–416.

Zou Z, Mubin O, Alnajjar F, et al. (2024) A pilot study of measuring emotional response and perception of LLM-generated questionnaire and human-generated questionnaires. *Scientific Reports* 14(1): 2781.

Author biographies

Simone Natale is an Associate Professor in Media Theory and History at the University of Turin, Italy, and editor of *Media, Culture & Society*. He is the author of numerous publications including his monograph *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test* (Oxford University Press, 2021), which has been translated into Chinese, Italian, and Portuguese, and over 50 peer-reviewed research articles in journals such as *Communication Theory, Journal of Communication, Convergence, Information, Communication & Society,* and *Communications of the ACM*.

Bruno Surace is a researcher in Film Studies at the University of Turin, where he has taught Cinema and Audiovisual Communication since 2019. He has published over eighty peer-reviewed articles and book chapters. He has presented at more than one hundred academic conferences across Europe, China, and the USA. He teaches Film History and Analysis at the Centro Sperimentale di Cinematografia. He is the author of *Il destino impresso. Per una teoria della destinalità nel cinema* (Kaplan, 2019)*, I volti dell'infanzia nelle culture audiovisive. Cinema, immagini, nuovi media (Mimesis, 2022), and Cinema liminale. Etiche ed estetiche mediali di un tempo sospeso* (Edizioni Estemporanee, 2025).

Enrico Mensa is a Research Associate (RTDa) at the Department of Computer Science, University of Turin, working on knowledge representation, NLP, and commonsense reasoning. He has recently engaged in projects at the intersection of AI and digital heritage, for example exploring the use of large language models to support curatorial narrative creation in museum contexts.

Luca Befera is a researcher at the University of Turin, where he studies algorithmic and multimedia performance with a focus on interactive dynamics, digital mediation, and ecological practices. He is currently analysing the implications of dramaturgical elements for the valorisation of cultural heritage in archives, libraries, and museums through generative AI.