SSOAR

Open Access Repository
www.ssoar.info

Shall androids dream of genocides? How generative AI can change the future of memorialization of mass atrocities

Makhortykh, Mykola; Zucker, Eve; Simon, David; Bultmann, Daniel; Ulloa, Roberto

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Makhortykh, M., Zucker, E., Simon, D., Bultmann, D., & Ulloa, R. (2023). Shall androids dream of genocides? How generative AI can change the future of memorialization of mass atrocities. *Discover Artificial Intelligence*, *3*, 1-17. https://doi.org/10.1007/s44163-023-00072-6

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by/4.0/deed.de

Terms of use:
This document is made available under a CC BY Licence (Attribution). For more Information see:
https://creativecommons.org/licenses/by/4.0

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-98480-6

Discover Artificial Intelligence

Discover

Perspective

Shall androids dream of genocides? How generative AI can change the future of memorialization of mass atrocities

Mykola Makhortykh[1] · Eve M. Zucker[2] · David J. Simon[3] · Daniel Bultmann[4] · Roberto Ulloa[5,6]

Received: 8 May 2023 / Accepted: 27 June 2023
Published online: 18 July 2023
© The Author(s) 2023 OPEN

Abstract
The memorialization of mass atrocities such as war crimes and genocides facilitates the remembrance of past suffering, honors those who resisted the perpetrators, and helps prevent the distortion of historical facts. Digital technologies have transformed memorialization practices by enabling less top-down and more creative approaches to remember mass atrocities. At the same time, they may also facilitate the spread of denialism and distortion, attempt to justify past crimes and attack the dignity of victims. The emergence of generative forms of artificial intelligence (AI), which produce textual and visual content, has the potential to revolutionize the field of memorialization even further. AI can identify patterns in training data to create new narratives for representing and interpreting mass atrocities—and do so in a fraction of the time it takes for humans. The use of generative AI in this context raises numerous questions: For example, can the paucity of training data on mass atrocities distort how AI interprets some atrocity-related inquiries? How important is the ability to differentiate between human- and AI-made content concerning mass atrocities? Can AI-made content be used to promote false information concerning atrocities? This article addresses these and other questions by examining the opportunities and risks associated with using generative AIs for memorializing mass atrocities. It also discusses recommendations for AIs integration in memorialization practices to steer the use of these technologies toward a more ethical and sustainable direction.

1 Introduction

Memorialization of mass atrocities, such as war crimes and genocides, has a special meaning to humankind. Memorialization involves the practice of preserving the memory of those who suffered, the experiences of what they endured, and the mass atrocity event itself. As such, the memorialization of collective suffering addresses many ethical and moral obligations that may be specific to a particular culture and society and reflect ideas drawn from transnational human rights and transitional justice. These obligations range from respecting and tending to the dead [66, 68] and honoring those

✉ Mykola Makhortykh, mykola.makhortykh@unibe.ch | [1]Institute of Communication and Media Studies, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland. [2]Department of Anthropology, Yale University/Weatherhead Institute of East Asian Studies, Columbia University, New York, USA. [3]Jackson School of Global Affairs, Yale University, New Haven, USA. [4]Humboldt-Universität zu Berlin, Berlin, Germany. [5]University of Konstanz, Konstanz, Germany. [6]GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany.

who fought against or resisted the perpetrators [31, 42][1] to countering the distortion of historical facts and preventing the repetition of atrocities [17]. The construction of social memory of atrocities via memorialization is thus integral to the moral reconstitution and social repair of societies in the aftermath of atrocities [18, 99]. However, these obligations are not easy to fulfill. Challenges include contestation over individual and collective memories with the imperative of accounting for various perspectives of the past, but without legitimizing patently false ones [68]; balancing individual narratives versus collective narratives and local versus national and international ones [44]; temporal, political, and other situational factors that may influence different narrative constructions among individuals or groups [80, 100]; limited access to human and financial resources required for commemoration [68]; and the need to resist efforts to instrumentalize memorialization efforts, which subordinate social needs to short-term political gains [47].

As with other societal practices, digital technologies have transformed the memorialization of mass atrocities [4, 20, 62, 63, 89, 91, 101]. While mass media had a major influence on memorialization already in the analog era [90], the rise of digital platforms caused a profound disruption in memory practices. The unprecedented amount of digitized and digital-born content, the increased connectivity between consumers and producers of memories [35], and the fluidity between various digital formats, including images, text, and video [39] have created space for less top-down memory practices and greater autonomy and creativity in memorial works. These practices offer the prospect of giving voice to under-represented communities and enable possibilities for engaging with the difficult pasts that might not otherwise have been accessible via heritage institutions and archives [77]. At the same time, digital technologies create challenges for memorialization, potentially facilitating denialism, and distortion [32, 55], while contributing to the trivialization of past suffering through the amplification of cynical attitudes toward it [36]. This latter phenomenon emerges from the potent combination of the unprecedented availability of atrocities-related content [10] and the ease with which such evidence of violence can be fabricated. Furthermore, the plethora of online memorial activities may obscure some memorial efforts in the noise of others or spark the banality that accompanies surface-level repetition in a manner reminiscent of Walter Benjamin's thoughts on the risks of mechanical reproduction [3].

The extensive volume of atrocities-related digital content prompts the need to help individuals navigate it. Since the 2000s, this task has been increasingly delegated to artificial intelligence (AI)-driven systems, such as search engines and recommendation systems, which filter content items in response to user input (e.g., search queries) and retrieve items that are viewed by the system as the most relevant for the users [48]. Despite the increasing use of such systems by the heritage institutions (e.g., USHMM digital collections' search; https://collections.ushmm.org/search/) and commercial platforms (e.g., Google search), the discussion of the opportunities and risks that AI systems create for the memorialization of mass atrocities remains limited (for some exceptions, see [6, 43, 56, 88]). For example, recent research has exposed the role of different search engines in memorialization by analyzing the search results in relation to the Holocaust [59] and the Ukrainian Holodomor[2] [60] and potential distortions in the representation of mass atrocities which these systems may cause.

The situation becomes more complex with the emergence of generative AIs, which represent a profound transformation in the context of the memorialization of mass atrocities. Instead of retrieving information about atrocities and prioritizing certain content sources, generative AIs produce the textual/visual content themselves [12], as illustrated in Fig. 1. The implications of this change for the memorialization of mass atrocities are substantial: we can expect not only a rapid increase of content engaging with memory about past suffering but also significant challenges in distinguishing between human- and AI-made content. The shift to generative AI also implies major changes in how information about mass atrocities is retrieved and how humans interact with such content in various contexts (e.g., education, community NGOs, international development and humanitarian entities, and government).

The efficiency of generative AIs in producing text and illustrations can facilitate the processes of compiling and analyzing memorialization data and make efforts to memorialize the past more accessible for individuals worldwide. Yet the long-term consequences of adopting generative AIs in the context of memorializing both historical (e.g., the Holocaust)

[1] Khan's work focuses on rescuers in genocide and other conflict situations with the aim of highlighting their moral choices as a model for younger generations. Similarly, institutions such as Yad Vashem recognize the noble acts of those who took great risks to help and save Jewish people during the Holocaust [92]. Gruner's work, on the other hand, details the resistance by Jewish people to Nazi policies and persecution. All three cases, in different ways, honor those who resisted and worked to save those in dire situations.

[2] Holodomor is a human-made famine that took place in the 1930s and took a particularly heavy toll on the Ukrainian population of the Soviet Union. A growing number of countries around the world recognize the Holodomor as a genocide perpetrated against the Ukrainian people.

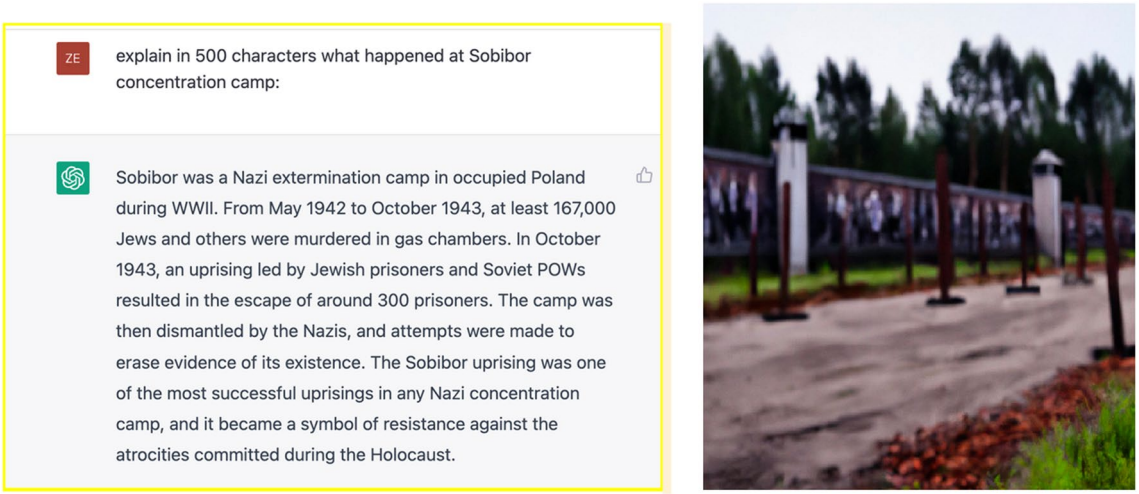

Fig. 1 **a** ChatGPT (https://chat.openai.com/) produced in a few seconds the piece of text responding to the prompt inquiring to explain in 500 characters "what happened at Sobibor concentration camp". **b** The AI platform Rytr (https://rytr.me/), also in seconds, produced a digital rendition of an oil painting using the prompt "digital memorial of Sobibor concentration camp with representation of victims in oil paint"

and recent atrocities (e.g., Russian war crimes in Ukraine) are uncertain. Some of the related concerns include the non-transparency of sources used to automatically generate narratives about the atrocities, the possibilities for misrepresentations of historical facts in rendered content, and the ethical problems arising from the generation of fake evidence or the manipulation of real one in a manner that changes the meaning of personal testimonies or artifacts. Additionally, it is unclear how the generated content might influence beliefs and attitudes toward a given atrocity, or how a growing distrust of some forms of digital media[3] might interact with an increasing awareness of AI capabilities and the pervasiveness of AI-rendered content in highly sensitive contexts such as the ones dealing with the memorialization of suffering.

Through our interdisciplinary perspective and experience of working in the field of memorialization, we aim to address some of these concerns and questions by reviewing the possibilities and challenges generative AI brings to the online memorialization of mass atrocities. As AI-related technologies continue to approach and exceed human capabilities in pattern recognition and processing large volumes of data in an increasing number of areas[4], it is crucial to evaluate their potential impact. Such evaluations are integral for deciding whether (and how) to incorporate generative AI in an area where technologies have to be adopted particularly carefully due to the ethical implications or whether recommending against AI use in this field is more beneficial due to its inadequacy to account for multiple ethical considerations guiding the process of atrocities' memorialization.

2 Multiple facets of AI: from information retrieval to information generation

The use of AI in the context of mass atrocity memorialization started long before the rise of generative AI. Defined as the ability of human-made artifacts to engage in intellectual behavior [102], AI has been applied to a broad range of memorialization-related tasks both by heritage institutions and commercial platforms. Most of the tasks[5] involving these non-generative forms of AI concerned information retrieval, a prominent field of computer science focused on information identification and retrieval for user information needs in response to user input [105]. Initially inheriting

[3] On this phenomenon, see [70, 71]. For instance, Newman and Fletcher [70] used survey data from nine OECD countries and found that amid growing distrust in all media, social media is the least trusted news medium.

[4] Notably, AI development seems well ahead of the pace of AI capability development predicted by [29, p. 729], which included "translating languages (by 2024) [and] writing high-school essays (by 2026)". See, for example, [53] on language translation and [33] on high school essay writing.

[5] Some other applications of non-generative AIs in the context of memorialization included, for instance, pattern recognition in data related to the historical event itself or its commemoration. For some examples, see [6, 103, 104].

many features of the information organization systems used by human practitioners to curate collections in the analogue archives and museums, non-generative AI systems, such as recommender systems and search engines, brought the ability to scale the retrieval tasks and automatically update databases with information about new sources and items (e.g., as in the case of web search engines constantly re-indexing the Internet with the help of crawlers [106]). Over time, non-generative AI systems started including additional mechanisms aiming to customize information delivery for individual users via personalization mechanisms (e.g., to select more relevant sources depending on the language of the query [107]) or introducing new formats for information retrieval (e.g., conversational exchanges as in the case of the Holocaust holograms [85]).

As a portal to information relevant for memorialization for the general public, non-generative AI systems have become the gatekeepers of information [108]. However, the expanding role of these systems in information retrieval and knowledge production highlights a number of shortcomings associated with their performance, in particular in relation to ontologically contested subjects, that is, subjects for which there are multiple viewpoints and a lack of general consensus. Existing research shows that the exact ways in which non-generative AI can malperform vary from non-systematic errors (e.g., search engines retrieving irrelevant or factually incorrect information [59]) to systematic bias resulting in a skewed representation of social reality [109, 110]. Researchers have also demonstrated the potential of non-generative AI systems to reiterate and amplify stereotypes, in particular in relation to discriminated gender and racial groups [110–114].

The evidence of limitations of non-generative AI systems together with their fragmented regulation [115, 116] has triggered a response from the scholarly community and also steps from the AI companies themselves (e.g., staged rollouts of products and calls for state regulation). The growing interest in non-generative AI (mal)performance resulted in the establishment of a new set of research techniques used to systematically audit the performance of non-generative AI systems [117, 118]. The exact implementations of these auditing approaches (for a comprehensive review, see [119] vary from sock puppet approaches using software simulating human behavior to generate inputs for the AI system [120], to techniques collecting system outputs from crowdworkers [121] to system code audits [122]. While the research on AI bias in the context of memorialization-related issues so far remains limited, a few studies [56, 60, 123] highlighted the tendency of non-generative AIs for focusing on a limited set of memorialization-related sources (often, the ones reiterating predominantly state narratives [123]), or prioritizing only a few selected aspects of the atrocity (e.g., prioritizing information drawn from a single atrocity site and downplaying the significance of other sites [59]).

Despite the lack of transparency of non-generative AIs, these systems would provide sources and excerpts that, although potentially subject to decontextualization, could still be traced back to their origins. This has changed with the use of generative AI which produce new content items, instead of finding and retrieving already existing ones[6]. For example, to generate text, language models (LMs), which power generative AI-systems (e.g., chatgGPT), are trained to predict the likelihood of the next token (e.g., a word) given information about the preceding or surrounding text [124]. In principle, the mathematical procedures allowing generative AI to make content do not allow for the emergence of understanding of the meaning behind such content [2]. Instead, generative AI can achieve only a shallow definition of meaning, known as distributional semantics—i.e., the description of the contexts in which a specific token appears [125, 126]. However, some evidence suggests that AI is capable of generating content of substantial internal complexity that goes beyond simple statistical modeling, although it may be just the result of parroting the gargantuan amount of data used for the training [127]. In either case, the ability of generative AI to imitate human capacities for generating content and its implications for different sectors has sparked heated academic and societal debates (e.g., [128]).

3 Possibilities of generative AI for digital memorialization of mass atrocities

Generative AI enables new possibilities for creating content that can be used for memorialization purposes. Such content can be either visual (motion or still) or textual and can reproduce multiple formats which are currently used in the context of digital grassroots memorialization of mass atrocities; examples of these formats range from video tributes commemorating specific instances of mass violence [55], to the Internet memes reinforcing [28] or challenging established genocide narratives [7], to the images imitating photos made at the sites of violence [52], to the drawings aiming to attract the attention of the general public to the past and present suffering [81], to the online encyclopedia entries

[6] Additionally, the recent attempts to integrate generative and non-generative AI systems (e.g., in the case of Bing search) highlight possibilities for combining content generation with information retrieval (e.g., to highlight information sources used to produce a text)

aiming to find consensus between the different viewpoints on the past [54]. As a result, AI might induce a variety of positive developments, including capacities for creation and participation in the context of memorialization, expanding public access to knowledge about the past, helping researchers learn more about mass atrocities and the memorialization thereof, and serving as a tool to help detect and counter false narratives.

- *AI can expand the capacity to memorialize*

  Generative AIs can easily produce representations of mass atrocities. The ease of access and use of these platforms enables new possibilities for expressing mourning, producing and analyzing testimony, and communicating loss—and for a fraction of the cost it would take to hire a human to create the textual and image content to fulfill these purposes. The outputs of AI models can then be employed on their own or combined with other digital- or analog-born materials to memorialize past and present atrocities (e.g., by populating digital memorials with AI-generated content where such content is lacking, or in cases where an AI-derived image or voice might serve to protect the privacy or safety of the victims or their relatives). This is particularly true in relation to survivors engaging with an AI platform to tell their stories or in cases where AI is employed as a therapist assistant [21].

  Similarly, the textual descriptions of atrocities or commemorative statements generated by text-based AIs may augment international participation and engagement by enabling individuals to participate in online memorial practices in different languages without the need for translators, copyeditors, or other production assistants. Under these circumstances, image- and text-focused AIs can empower individuals who want to create, contribute to, or expand existing grassroots memorialization practices with almost limitless creativity providing their projects remain within the accepted use requirements of the platforms (e.g., Midjourney's use guidelines aiming to prevent the creation of graphic or potentially harmful images through an explicit list of banned prompts [30, 65]).
- *AI can enable new ways for the general public to learn about mass atrocities*

  Engaging with generative AIs to learn about mass atrocities can be a form of memorialization practice itself. Earlier, non-generative versions of AI-driven systems focused on information retrieval tasks [1]. For example, the *Let Them Speak*[7] platform [133] incorporated databases of Holocaust textual testimonies that could be queried for specific words and used to create lists of testimonial fragments [69]. The Shoah Foundation-produced holograms of Holocaust survivors [85, 88] is a technologically similar form of AI that uses a finite database of questions with pre-recorded answers from Holocaust survivors and is employed to enable an interactive experience for individuals and groups engaging with the holograms.

  A key distinction of generative AI is the ability to produce *new* narratives, store these narratives, and use them as additional data for iterative training. It allows generative AIs not only to retrieve content in response to an (often limited) set of user inputs, but to generate content and sustain a broader range of interactions with individuals interested in exploring the past. Thus a user may have a vested interest in continuing the dialogue with the generative AI chatbot on a given topic rather than having to start at the beginning each time as in the case of more traditional conversational agents. These interactions, which occur on one-on-one user and machine bases over private accounts, can be understood as a form of memorialization in itself: for instance, when individuals engage with AIs to generate a text- or image-based narrative about a mass atrocity.
- *AI can be a means for researchers to collect/analyze data*

  The advancements in the field of AI also transform the processes for collecting and analyzing data concerning both historical and present mass atrocities. Instead of simply retrieving and ranking information sources relevant to the user queries (as non-generative AIs used to do), generative AI platforms can directly respond to user requests while also providing suggestions of information sources for further exploration (e.g., as in the case of generative AI interface for Bing search). Such functionality can potentially accelerate the process of analyzing data on mass atrocities and lead to profound transformations in how institutions provide access to this data. For example, instead of using a conventional archive search, a generative AI can be integrated into the interface to provide recommendations to the researcher in a conversational format; similarly, generative AI can be used to generate visualizations based on text descriptions to offer additional perspectives for the researcher. In addition to retrieving information in a more digestible format, generative AIs can also be used for automated content labeling that can help identify content related to

[7] For a more detailed discussion of the platform and its relationship to digital memorialization see [69]. The platform itself can be found here: https://lts.fortunoff.library.yale.edu/.

mass atrocities, given AIs' capacity for content identification can be comparable or superior to that of humans [26, 86].

There are also other capabilities of generative AI which can facilitate the analysis of data related to mass atrocities. Similar to earlier AI-driven information retrieval systems, generative AI can personalize outputs (e.g., textual content generation prompted by individual user needs and interests). For example, AI can summarize data and present several summaries emphasizing various elements to let the researcher think through the data from various perspectives. Furthermore, both text- and image-focused generative AIs can be used for creating synthetic data (see, for instance [75]), which can then be employed for training computational approaches for automated content analysis. For example, the Iraq war crimes investigation initiative UNITAD and the Germany-based non-governmental organization Mnemonic have reportedly started using AI to scan hundreds of thousands of hours of video to identify evidence of war crimes in the context of recent armed conflicts in Iraq and Syria [84].

- *AI can be a tool for detecting distortion and denialism*

  In addition to the above-mentioned possibilities, the ability of generative AIs to interprete and engage with user input makes them an effective means of content classification. Content classification involves the assignment of specific labels to content items to detect specific attributes of such content (e.g., whether it is related to a specific subject, such as politics, or expresses a specific sentiment, for instance, positive or negative one). In the case of content related to mass atrocities, such classification can take multiple forms, including whether the statement that AI is asked to evaluate comes from a certain source or whether it might contain denialist claims. While the degree to which generative AI can perform these functions, specifically in the case of content dealing with mass atrocities, is yet to be studied, generative AI platforms, such as chatGPT, have demonstrated the ability to identify whether claims entered by the users are false in relation to diverse subjects [34] and health-related matters in particular [41], and to evaluate the degree of news outlets' credibility [94]. Under these circumstances, generative AI can potentially be used to filter out content promoting distortion or denialism of mass atrocities, with the possibility that individuals will be more open to engaging with views they find contrasting to their own if these views are expressed via a human-like conversational interface (for examples related to conventional chatbots, see [95]).

4 Threats of generative AI for digital memorialization of mass atrocities

Despite the many advantages generative AI presents for the memorialization of mass atrocities, there are also risks. An amoral machine entity, AI does not assign specific meaning to the data it processes. Rather, AI is more akin to Bender and Kohler's "octopus test" [2], where a clever octopus fools people into believing it is a sentient human by communicating in the English language through statistical patterns until it fails to comprehend the implications of an unfamiliar context, or Searle's "Chinese Room" proposition [76] in which he posits he could create coherent text in Chinese by learning to manipulate symbols without actually understanding the language itself. Both exercises demonstrate how the capacity to impart information should not be conflated with the ability to understand the meaning—nor should it, beyond that point, be confused with the capability to make context-driven moral judgments. Yet memorialization is a realm of human thought and activity for which the abilities to make fine distinctions in meaning and to exercise moral judgment are essential.

- *AI can serve as a means of enforcing hegemonic narratives and practices*

  Generative AIs are trained on specific sets of data. Unless their training process is diversified, AIs can take prevalent patterns in data for granted and then reiterate them in content which AIs generate. In the case of atrocities-related content, it may result in the enforcement of hegemonic narratives and representation practices, for instance, by prioritizing Western-centric views on how mass atrocities shall be remembered or interpreted and uncritically translating these views to other contexts. Such reinforcement of memory hegemonies might result in the silencing or erasure of experiences of minority groups (including the ones which were disproportionately affected by the mass atrocities) and suppression of alternative practices of memorialization [19]. It is additionally concerning in the case of authoritarian states, where hegemonic memory practices often serve as an integral means of state ideology and propaganda and where generative AIs can become an effective tool for consolidating the national memory regimes, such as in the case of recent efforts by China [16].

The capabilities of generative AIs to enforce hegemonic narratives are further amplified by the risks of keeping users in information bubbles, which was a common concern about the non-generative AI-driven systems dealing with information retrieval tasks (e.g., [9, 71]). While the existence of AI-amplified information bubbles has so far found little empirical support [9] and, in some cases, such bubbles can be beneficial for society by nurturing independent thinking (e.g., [58]), the possibility of generative AIs leading their users into the "rabbit holes"[8] [8] of memory hegemonies can not be currently excluded. Further support to such concerns is provided by the observations indicating that some generative AI models can remember the previous interactions with a particular user [96], thus contouring their answers to what the system perceives as the desired type of response that the specific user is seeking.

Self-reinforcement or feedback loops wherein model outputs become increasingly less diverse over time [82] are another aspect of generative AIs which can potentially serve to amplify hegemonic narratives of atrocities. While the possibility of collecting user input (e.g., positive/negative evaluations of outputs of chatGPT) is important for improving the system functionality, in the case of memorialization, it might also result in prioritization of small sets of possible outputs (e.g., standardized textual or visual narratives of atrocities). Given the highly political battles over narrative creation that are often embedded in memorialization efforts (see, for instance, the case of the Holodomor [45, 72] or the genocide in Rwanda [40]), AI platforms—assiduously apolitical by design—might nevertheless have profound political and social consequences.

- *AI can introduce bias in analysis and make some research practices more obscure/obsolete*

  The popularity of AI can entice (or even force) researchers towards relying on AI-driven tools instead of traditional qualitative practices of studying the memorialization of mass atrocities. While it can enable new possibilities for research in the field of memorialization, it might diminish the role of the human in the loop by making obsolete interactions and forms of knowledge that require human embodiment and, by doing so, allow for a wider range of communication beyond what the digital systems can produce or relay. Such a shift can be concerning for a multitude of reasons, including the possibility of generative AIs being subjected to certain forms of bias that, in the case of earlier AI systems, varied from the retrieval of factually incorrect information to systematic skewness of outputs (e.g., in terms of visibility of specific memorial sites and practices [59]).
- *AI can be used to facilitate and enhance censorship*

  The ability to program what AI can (and cannot) generate creates not only possibilities but also risks for memorialization. It might improve efforts to detect and combat denial and distortion, but it could just as easily inhibit individuals or groups in vulnerable situations, such as in authoritarian countries, from participating in memorial activities that counter the state's narrative. For instance, AIs can be prevented from generating information about particular atrocities, thus limiting memorial efforts of communities who wish to remember these events publicly. Authoritarian-leaning states seeking to censure memorialization activities may use AI to reinforce a preferred narrative of the past while stifling expressions that stray from that norm.[9]

  Alternatively, AI can potentially produce watered-down or even irrelevant generalities about a given atrocity at the expense of more nuanced complexities that make memorialization meaningful, thus enabling a different form of algorithm-driven "masked censorship" [57, p. 38]. In some cases, the motivation for censorship can be benign—i.e., in the case of commercial companies behind AIs forbidding the generation of atrocities-related content to avoid potential misuse of AIs for trivialization and denialism—but it nevertheless can interfere with the use of technology for memorialization.
- *AI can be used to generate narratives that support distortion and denialism*

  Similar to the possibilities for amplifying genuine memory practices, generative AIs can be abused to amplify distortion and denial in the context of mass atrocities. Such abuses can be intentional and unintentional. In the former case, AIs can be used to produce a large volume of diverse content promoting claims that deny that specific mass atrocities happened (e.g., via text-focused AIs) or promoting a distorted representation of the atrocities (e.g., via image-focused AIs generating sexualized images of Holocaust victims, such as Anne Frank, or images glorifying the Holocaust perpetrators). Distorted representations could even be quite nuanced, for example, by producing generative images of atrocities that look similar to the authentic ones but lack important details or by replacing the faces of perpetrators to confuse the public.

[8] The rabbit hole denotes the process of moving from the diverse information environment to an ideologically extreme echo chamber.

[9] Some examples include several countries in Southeast Asia [46, 73, 78], China [15, 16], Russia [50], Turkey [24].

Aside from the risks of intentional technology abuse, generative AIs are prone to hallucinations [5], in which the systems fill up information voids with generated content that is not factually supported. An example of such hallucinations in the case of mass atrocities can be the invention of fictional details about a particular perpetrator or a victim to respond to the user prompt in response to which the AI has no factual information. In this case, the risk is intrinsic to the technology itself (i.e., the ability of AI to generate content): without establishing a system of controls (e.g., informing the user that there is no information available about their inquiry), the use of AI can lead to the distortion of historical facts.

- *AI can undermine trust in the use of technology for memorializing mass atrocities*

  The humanlike responses of AI chat systems to users' questions make these systems inherently more trustworthy to the humans who engage with them [27]. Science fiction films like Spike Jonze's *Her* [132] and Alex Garland's *Ex Machina* [131] have contributed to the popular perception of AI as compelling—but also potentially dangerous and untrustworthy. The inherent uncertainty that accompanies trust [67] can be integral to engaging in new experiences and, in some cases, goes so far as to be a 'leap of faith'[10] [98]—for example, *in extremis* in Garland's film or potentially for some of today's generative AI users. With respect to the transmission of memory about genocides and other mass atrocity events, however, the impacts of a potential 'leap of faith' by users engaging with generative AIs(e.g., chat systems such as chatGPT) could have negative results. For example, users may fail to check questionable claims generated by AIs and potentially repeat them.

  The unsuccessful leaps of faith in relation to the use of generative AI can interfere with the successful integration of these technologies in memorialization practices due to such integration being predicated (as it is for other tasks [22]) on a degree of trust in the technology. There needs to be some level of certainty that the AI system is conforming to expectations both in its use and its output. When applied to subjects as delicate as mass atrocity memorialization, accuracy is often viewed as a key imperative. The guiding transnational belief is that the atrocities should be remembered, and the suffering of victims should be honored and not forgotten [19], and for this purpose to be achieved, the accuracy of representation of atrocities and suffering must be trusted. With this in mind, if a specific form of generative AI was to acquire a reputation for inventing its facts or promoting skewed interpretations of mass atrocities, it would likely lead to an erosion of that underlying trust.

  Given that generative AI's output is (at least partially) determined by the training data selected by its creators, critics of the technology sector might seek to disparage AI based on the trope that "big tech" routinely censors knowledge production in the interests of a hidden agenda—which, per the logic of conspiracy theory, need not even be specified to be believed. Moreover, irrespective of the systems themselves, individuals may still distrust the AI-produced text and narratives due to unverified claims made about the content's veracity. Under these circumstances, the malperformance of certain forms of generative AIs could create a sort of moral panic, causing severe distrust towards new forms of innovation similar to what had happened when earlier non-generative AI systems spawned fears of filter bubbles [9], and led to distrust towards their use in societally relevant sectors, such as journalism.

- *AI can disclose sensitive information about individuals*

  Generative AI models are trained on vast amounts of training data collected from the Internet, and such data may include personal information. Since 2021, research has demonstrated that it is possible to extract such information from earlier AI models [13], and ChatGPT is not exempt from this issue [49]. Similarly, it is also possible to retrieve training images (or sections of them) from models that generate visual content [14]. The sudden launch of these models to the market has left governments scrambling to take action against generative AI providers, for instance, in the case of Italian data protection regulators demanding OpenAI stop using training data of Italian citizens as it is not compliant with the General Data Protection Regulation (GDPR). In response, OpenAI has stopped providing access to chatGPT in Italy [11].

  While the disclosure of sensitive information per se is problematic, in the case of the memorialization of mass atrocities, it raises additional concerns due to the care and sensitivity integral to respecting and honoring victims. Sensitive details about the lives of atrocity victims, including, for instance, images of them being murdered or tortured, prisoner intake photographs, or other visual representations (including images taken before the atrocities, such as victims' childhood photos) may be problematic in the context of the production of new types of content for public (or even

[10] Anthony Giddens [25] notes that trust involves risk, and therefore a 'leap of faith'.

commercial[11]) purposes. However, the exclusion of this material is challenging given the volume of sensitive content available for AI model training as well as the difficulty of drawing boundaries regarding exclusion of such content.

- *AI can create non-authentic content in relation to mass atrocities*

  The ability of AI to generate content that can be hardly differentiated from the content produced by a human (e.g., [83]) raises concerns that generative AIs could produce inauthentic content about a mass atrocity which might then be presented and treated as authentic. Such content can vary from fabricated historical documents assigning the blame to specific actors or whitewashing the actual perpetrators to fake visual evidence offering a distorted representation of the atrocities (e.g., to scapegoat a specific group of victims or bystanders). While companies behind AIs will likely try to limit such uses, the susceptibility of models to blindly follow user instructions can result in users circumventing these restrictions in a process known as "jailbreaking" as demonstrated in studies examining the potential of generative AIs to produce extremist content (e.g., [64]). If people believe potentially false statements made by an AI, then they can fail to check the results against other sources and it can lead to perpetuating false claims and exacerbate the problems of denial and distortion already present. Even if users of generative AIs remain capable of detecting and rejecting misleading or extremist claims, there could still be a cumulative effect of undermining the trust in historical sources through the generation of more subtle forms of non-authentic content that steers the narrative toward interests beyond the genuine purpose of atrocity memorialization itself.[12]

5 Other considerations of using generative AI for digital memorialization of mass atrocities

Beyond the possibilities and potential risks identified in the previous sections, there are additional considerations and dimensions related to applying generative AI technologies to mass atrocity memorialization practices. Some of these considerations emanate from challenges that may arise in research efforts to understand the impacts of AI on memorialization, whereas others relate to the long-term implications of integrating generative AIs in the broad field of digital memorialization.

- *Access to data about mass atrocities*. The widely used AI models usually rely on information available on the Internet, but most of the content related to mass atrocities is stored in traditional archives, which are often only partially digitized. Therefore, the knowledge that AIs possess about atrocities is limited to already digitized or digital-born content (e.g., PDFs, images, and other content items, including ones coming from non-institutional sites such as Pinterest), which the model can access. Another aspect of data access concerns the tendency of some generative AIs (e.g., chatGPT-3) to rely on a snapshot of data produced at some time point (e.g., only up to 2021). Time-range boundaries can have implications for content generation in relation to mass atrocities, especially concerning the atrocities which happened after the model underlying the AI has been trained and deployed. At the same time, many AI models integrate possibilities for retrieving data from the Web to overcome these time limits; however, doing so also increases the possibility of using unverified or false data.
- *Matters of user privacy*. Currently, there is limited understanding of how generative AI platforms use the data provided by their users (e.g., whether they track information requests generated by individual users) and what applications of these data they might consider. In the case of data about mass atrocities, certain requests might be rather sensitive (for instance, inquiries about the war in Ukraine contradicting the official state narrative in Russia for Russian users, in particular when using Russia-based AI platforms).
- *Matters of representation*. With respect to the *Let Them Speak* project, Naron and Toth [69] assert that the aim of the project is to make the fragmented collection of Holocaust testimonies representative of those who perished and, therefore, could not speak. Applying generative AI could take such projects a step further by using separate testimonies to give AI the power to speak and tell coherent stories about victims' experiences. On the positive side, such a system arguably permits a representation of those who have no voices anymore. However, this move begins to approach the uncanny valley where the chatbot speaker could become a victim of a sort itself or, interpreted through

[11] See [97] for the discussion of how photographs of prisoners at Tuol Sleng were digitally altered for commercial wall art and sold on major shopping sites.

[12] This idea stems from the observation by [87] concerning that exposure to AI-generated deepfakes can erode trust in social media in general.

a supernatural lens, as a medium for the ghosts and spirits of the deceased. A chatbot could conceivably become seen as an entity unto itself, distinct from the inputs used to construct its persona (and memories, of a sort). Representing those who have perished in a mass atrocity through a technology that uses existing testimonies to fabricate new ones may also raise ethical concerns about atrocity remembrance along with epistemological ones about how we understand and represent the past.

- *Replacement of the human working force involved in memorialization*. The ability of generative AI to produce a large volume of textual/visual content related to mass atrocities enables new possibilities for decreasing the costs of digital memorialization activities. However, these possibilities also create risks for humans working in memorialization-related fields. For example, visual artists, copywriters, data analysts, and certain types of archivists could potentially be reduced or replaced by generative AI technologies. Additionally, generative AIs can perform some functions associated with the roles of curators working in heritage institutions connected with mass atrocities. Not only shall we consider the economic implications of job loss or displacement, but the fact that the involved individuals, directly or indirectly, increase awareness of mass atrocities and propagate the sensitivities of the past suffering to groups that would otherwise not engage with it. Under these circumstances, it is imperative to consider integrating generative AIs into memorialization practices while prioritizing the input and expertise professionals in these sectors have and the importance of social interaction in informing individuals in other societal sectors.

6 An agenda for future research on the role of generative AI in the context of atrocity memorialization

In light of the trends, potentials, and uncertainties in AI-generated atrocity memorialization that we have identified, we now turn to the question of what we, as a scholarly community, can do to realize the possibilities and counter the risks of using generative AI for digital memorialization. Below, we share four recommendations that, in our view, are integral for steering the use of these technologies toward a more ethical and sustainable direction.

Experimenting with AI in the context of mass atrocities. We recognize threats related to genocide distortion/denialism, which can be amplified by AI, but we should not ignore the possibilities provided for the memorialization of mass atrocities. To examine these possibilities, we need to keep experimenting with generative AI and also keep options for experimentation open for other stakeholders, including ordinary citizens. The ability to experiment with AI is essential for understanding the full range of its possible uses for memorialization. While banning the atrocities-related prompts (as done by certain AIs, such as Bing Image Creator) can be a safe option, it could also, problematically, contribute to erasing memories about past suffering by undermining memorialization (and prevention and rescue) efforts. For example, human rights groups seeking to collect and display evidence of war crimes and genocide may have the content of their projects censored out from AI training data (and, consequently, AI system outputs), thus making it difficult for them to show the crimes being committed with the hope of some action [37, p. 10]. Several studies [23, 37, 38] have observed a similar phenomenon concerning human and algorithmic moderation systems used by social media platforms to block potentially offensive content (including evidence of mass atrocities) and then delete such content before perpetrators can be held accountable. Finally, even if major AI companies ban queries dealing with mass atrocities, it might be hard to do it comprehensively (e.g., so users would not find a way to circumvent the limitations), whereas such bans can push individuals towards using generative AI models deployed in fringe environments like the dark web.

Studying how generative AI is used for preserving and distorting memories of mass atrocities in and outside academia. There are indications (e.g., [79]) that generative AIs could become another source of moral panic, similar to earlier forms of AI-driven technology (e.g., recommender systems and the associated filter bubble fears [9]). To counter it, it is important to understand how generative AIs are used in the memorialization of mass atrocities and whether the concerns about the risks of their adoption in the field of memorialization are justified. The complexity of memorialization, which includes a broad range of individual and group practices focused both on remembering past suffering and on researching such practices, adds to the challenging nature of this task.

How to achieve such an understanding is an open question. Unlike content released via websites, blog posts, or social media, memorial activity relying on AI platforms is often confined to private accounts. The inputs and outputs of AI systems are hidden from public view unless or until the user shares them in a public (presumably digital) forum. While there are exceptions (e.g., the rendering of images on Discord servers using models like Midjourney), even in these cases, the degree of communal engagement with generative AI appears to be more limited than with more conventional digital memorialization practices such as commentaries for statements of memorial institutions published on social media.

Under these circumstances, a first step can be to examine cases where individuals or groups integrate AI-rendered text, image, video, or voice content into established memorialization practices (e.g., by adding them to memorial websites devoted to a mass atrocity) and, where possible, conduct interviews with practitioners involved in such integration. Though few currently, such cases are likely to proliferate as individuals and groups seek the capabilities that AI can provide to enhance the message they want to convey about past suffering. Another alternative can be to use experimental research designs to examine how individuals use AIs to generate memorialization-related content or engage with such content in a controlled environment. Similarly, one might track the extent to which heritage institutions and human rights organizations adopt generative AIs, along with their respective motivations and expectations.

Arguing for the importance of integrating transparency principles in the generative AI design and functionality. Generative AI design must follow the principles of transparency both in data acquisition for training the underlying models (e.g., to ensure that data was acquired without breaching intellectual rights) and in providing possibilities for analyzing the performance of generative AIs and their potential bias. Similarly, greater transparency is needed in the functionality of generative AIs: for instance, by providing sources for data that are used to generate outputs (when possible) so potential privacy or copyright breaches would be easy to identify. Transparency is especially important in the context of using generative AIs for producing content related to mass atrocities due to associated ethical and moral considerations (e.g., the importance of treating the memory of victims respectfully) as well as the need for public cooperation in the effort to identify potential risks of AI misuse. Under these circumstances, transparency of AI design and functionality becomes a main prerequisite for AIs becoming reliably truthful and establishing a reputation as such.

Watch AI developments beyond the West and Silicon Valley. The development and application of generative AIs to mass atrocity memorial practices are not confined to the West and Silicon Valley. Generative AIs are employed, refined, and redesigned to meet the needs and tastes of various cultures and nationalities around the globe. For instance, the Cambodian government is currently creating its chatGPT system for the public [51], and another version was created in Saudi Arabia (miniGPT-4 [61]). While it is not clear yet when or how these platforms will be used for memorial purposes, there is little question they will eventually be adopted to memorialize past and present suffering. In China, for example, the government has sought to place guardrails on the range of sources a chatbot might use to discuss particular events, such as the Tiananmen Square massacres [16], and the Cambodian government is evaluating the risks of the public using the Khmer chatGPT as a primary source of information [93]. Moreover, the speed at which these technologies are being developed worldwide is astounding. Cambodia, which is considered a developing nation and has been struggling for decades to recover from the devastating Khmer Rouge era, has already produced a robot that speaks using the model underlying the ChatGPT platform [74]. What do these developments suggest for memorialization? How will these varying forms of ChatGPT impact the memorial narratives produced? We can expect the unexpected: developments that Silicon Valley could not have anticipated as the AI models are increasingly trained implicitly and explicitly based on different core ideologies, cultures, and political agendas.

7 Discussion

Generative AIs defy conceptual premises regarding the relationship between digital technologies and memorialization by surpassing the information retrieval functionalities which are commonly associated with the use of AI for digital memorialization. Instead of ranking existing memorialization-related content in response to human queries, generative AIs utilize complex models to produce responses that to a certain degree emulate the human ability to interpret and comprehend information. While the emulation capabilities of generative AI may fall short at times, they can expand the analytical capacities of researchers studying memorialization and amplify possibilities for the general public to learn about past suffering. However, the possibilities enabled by generative AI also come with a price (as demonstrated by the discussion of shortcomings of its use in the context of memorialization above). Therefore there is an urgent need for careful critical assessment of appropriateness of using generative AI and its capacity in this sensitive context. Adding to considerations of its application to mass atrocity memorialization is the importance of extensive contextualization of each unique instance of atrocity and the loss of life and suffering that these events entail.

The major challenge of enabling such contextualization relates to the difference between human and AI capacities of understanding semantics of articulations and representations of specific instances of mass atrocities and their memorialization. For generative AI to be meaningful for memorialization, it needs to achieve a certain understanding of the meaning of texts, images, and utterances involved in the production of stories about suffering as well as the ethical and moral dimensions of these stories. The ability to understand these complexities is also crucial for how AI-generated

content about mass atrocities will be received by the human audience and what meanings embedded in memorialization practices will be rendered or lost. In this context, Walter Benjamin's [3] observations regarding the aura of an artwork that is grounded in its originality and immediate connection to its human maker are particularly relevant. Benjamin argues that the mechanical reproduction of a work of art—e.g., by AI models which reuse elements of existing texts or images to create new ones—severs the art from its human maker, effectively removing the work of art's aura.

Applying this to generative AI memorialization of mass atrocities a similar and more poignant conclusion can be drawn when the "work of art" is witness testimonies or other representations of collective trauma and suffering. Even if a generative AI can produce a relatively convincing version of a textual or visual narrative of a mass atrocity, it begs the question (beyond the obvious ethical issues) whether the human element was lost along the way. Turning to the human audience, there is then the question of perception. Beyond perceiving the AI-generated story (as film, image, or narration) as passable by objective standards, is there still something that is lost in its reading? Roland Barthes [129, p. 27] identifies what he calls the punctum that photographic images can evoke: "that accident which pricks" that are the details in a photograph that elicit human meaning and connection—a missing button on a jacket, a particular look of the subject, a wilted flower. These are the bits that connect the viewer to the subject by being subjective and born from human experience as in the case of. the small familiarities or notable absences. Bender's Octopus [2] lacks the experience and imagination to make these connections and inferences. Arguably, neither can generative AI, and even if it could, what meaning or purpose would that serve?

These considerations are integral for critically assessing the implications of using generative AI in the context of memorialization in different areas, including public policy and international relations. In the context of public policy, generative AI can enable new modes of accessing information about mass atrocity (e.g., as part of school education and life-long learning) that can increase societal awareness about the mass atrocities which is essential for their prevention. Similarly, generative AI can shape new standards of using atrocity-related content for memorialization purposes (e.g., by determining what can and can not be generated). This can lead to harmonizing knowledge about the past and sustaining the production of content that can be used for grassroots memory practices while countering the spread of distortion. However, without careful understanding of the context in which the atrocity occurred and the semantic complexities of accounting for the diversity of individual experiences, such knowledge harmonization can result in the simplified representation of the atrocity that can undermine human-led memorialization efforts or even silence victim voices (e.g., in the case they belong to the groups which are marginalized in the present).

The implications of relying on generative AI for memorialization in the context of international politics can also be rather ambiguous. The ongoing rollout of generative AI systems in different parts of the world suggests that the national agendas regarding the use of AI for mass atrocity memorialization will not necessarily be driven by international consensus. Such a consensus can be even more difficult to achieve considering that different countries may rely on their own generative AI platforms together with country-specific datasets used for training models powering the respective platforms and distinct sets of AI guardrails. Under these circumstances, the possibility of generative AIs being used to fuel cross-state online memory wars—i.e., the discursive battles dealing with contrasting interpretations of the past [130]—becomes more likely with AI serving as a source of content promoting specific narratives and subjugating the alternative interpretations.

The discussion of the implications of the use of generative AI for mass atrocity memorializations in the context of public policy and international relations guides us to the crucial question about how such use will affect the role of human practitioners in the processes of atrocity memorialization and the ways to keep the human in the loop. While it is too early to know how exactly generative AI will add value to human memorialization practices, our work suggests that it can enable a more multidimensional and comprehensive understanding of the past, for instance, by summarizing insights coming from the large volume of testimonies, legal documents, and other historical materials relevant to memorialization. It can also provide representation of personal experiences that are difficult to convey and where in the interest of protecting the identities of victims is essential.

At the same time, any benefits of generative AI for memorialization must be considered with the potential harms it may cause and this is why it is imperative that humans guide, analyze, and review any memorialization-related process incorporating AI. Without such supervision, generative AI can easily create confusion and distrust by producing false representations of past suffering that can make individuals doubt the accuracy of related atrocity narratives. Such uncertainty (or false certainty) can exacerbate concerns about the rise of denialism and distortion in relation to mass atrocity which are associated with social media and other digital developments. To address this problem, it is essential to implement mechanisms for human supervision of AI with a particular emphasis on transparency of using AI-driven platforms as well as data and procedures used for training AI models. Such transparency mechanisms shall be supplemented by capacities for monitoring the performance

of generative AI and nurturing digital literacies for different demographic groups (and not only teenagers/young adults) to prevent the potential erosion of trust towards AI-enhanced memorialization practices.

The implementation of the above-mentioned mechanisms is essential for integrating generative AI with human memorialization practices, in particular considering the limited capacities of AI to comprehend the full scope of human semantics in the context of memorializing mass atrocities as it was discussed earlier. Any memorialization process needs to be human-driven rather than machine-driven for it to be a true representation of the human experiences it is intended to convey and to respect the dignity of those who perished. Generative AIs can augment human understanding provided it does not directly shape that understanding (e.g., by representing only particular subsets of information or by highlighting or repressing certain experiences). The practical procedures that might be employed by humans are difficult to predict in their entirety, however, the ability of human practitioners to stay informed about how generative AI works in the context of memorialization and also be able to guide its design and usage may be considered as a central requirement to the process. For this, establishing national and international guidance and policies for its use, especially pertaining to mass atrocity cases is fundamental.

8 Conclusion

The foregoing insights about the impact of generative AI on mass atrocity memorialization must come, as is the case for any future-looking discussion on AI, with the caveat that there are more unknowns than there are knowns. As the developments in the field of AI accelerate, academic research and societal discussions must remain responsive to the ongoing developments. This article offers some guidelines in terms of what scholars, as well as practitioners in both the AI and memorialization fields, may need to consider at this moment. As is the case with other rapidly evolving subjects, the discussion of possibilities and risks of relying on generative AI in the context of memorialization, as well as considerations associated with such reliance, can help to establish a framework of reference for anticipating practical and conceptual issues and guiding future debates on the topic. Transparent and collaborative efforts drawing from various academic disciplines can help to address the challenges presented by this emergent technology to support its ethical integration with existing memorialization practices.

Author contributions Conceptualization: EM.Z, MM, RU. Writing—original draft preparation: MM, EM.Z, RU. Writing—review and editing: EM.Z, MM, DJ.S, RU, DB. Literature search: MM, EM.Z, RU, DJ.S. Funding acquisition: MM. Supervision: MM, EM.Z.

Funding The research has been supported by the Alfred Landecker Foundation, which provided financial support for the research time of Mykola Makhortykh, who contributed to the article as part of his project titled "Algorithmic turn in Holocaust memory transmission".

Data availability Not applicable.

Code availability Not applicable.

Declarations

Competing interests Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Agrawal A, Gans J, Goldfarb A. Prediction machines: the simple economics of artificial intelligence. Brighton: Harvard Business Press; 2018.
2. Bender EM, Koller A. Climbing towards NLU: On meaning, form, and understanding in the age of data. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, ACL; 2020. pp. 5185–5198; https://doi.org/10.18653/v1/2020.acl-main.463.
3. Benjamin W. The work of art in the age of mechanical reproduction. Harlow: Penguin Books; 2008.
4. Benzaquen S. Looking at the Tuol Sleng Museum of genocidal crimes, Cambodia, on Flickr and youtube. Media Cult Soc. 2014;36:790–809. https://doi.org/10.1177/0163443714532983.

5. Beutel G, Geerits E, Kielstein JT. Artificial hallucination: GPT on LSD? Crit Care. 2023;27:1–3. https://doi.org/10.1186/s13054-023-04425-6.
6. Blanke T, Bryant M, Hedges M. Understanding memories of the holocaust—a new approach to neural networks in the digital humanities. Digit Scholarsh Humanit. 2020;35:17–33. https://doi.org/10.1093/llc/fqy082.
7. Bright J. History under attack: holocaust denial and distortion on social media. UNESCO. 2022. https://www.unesco.org/en/articles/survivor-and-her-family-stand-strong-against-holocaust-denial-and-distortion. Accessed 1 May 2023.
8. Brown MA, Bisbee J, Lai A, Bonneau R, Nagler J, Tucker JA. Echo chambers, rabbit holes, and algorithmic bias: how YouTube recommends content to real users. SSRN. 2022. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4114905. Accessed 28 Apr 2023. https://doi.org/10.2139/ssrn.4114905.
9. Bruns A. Are filter bubbles real? Hoboken: John Wiley & Sons; 2019.
10. Bultmann D, Makhortykh M, Simon D, Ulloa R, Zucker EM. Digital archive of memorialization of mass atrocities (DAMMA) workshop whitepaper. Yale University Genocide Studies Program. 2022. https://gsp.yale.edu/digital-archive-memorialization-mass-atrocities-damma-workshop-whitepaper. Accessed 1 May 2023.
11. Burgess M. ChatGPT has a big privacy problem. Wired. 2023. https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/. Accessed 1 May 2023.
12. Cao Y, Li S, Liu Y, Yan Z, Dai Y, Yu PS, Sun L. A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. arXiv. 2023. https://arxiv.org/abs/2303.04226. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2303.04226.
13. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss, A, Lee K, Roberts A, Brown T. B., Song D., Erlingsson U., Oprea A, Colin Raffel. (2021). Extracting training data from large language models. USENIX Security Symposium. 2021. https://www.usenix.org/system/files/sec21-carlini-extracting.pdf. Accessed 1 May 2023; https://doi.org/10.48550/arXiv.2012.07805.
14. Carlini N, Hayes J, Nasr M, Jagielski M, Sehwag V, Tramèr F, Balle B, Ippolito D, Wallace E. Extracting training data from diffusion models. arXiv. 2023. https://arxiv.org/abs/2301.13188. Accessed 1 May 2023; https://doi.org/10.48550/arXiv.2301.13188.
15. Chen T. Made in censorship. New York: Columbia University Press; 2022.
16. Che C. China says chatbots must toe the party line. New York Times. 2023. https://www.nytimes.com/2023/04/24/world/asia/china-chatbots-ai.html. Accessed 28 Apr 2023.
17. Cox J, Khoury A, Minslow S. Denial: the final stage of genocide? 1st ed. Abingdon: Routledge; 2021.
18. Das V, Kleinman A. Introduction. In: Das V, Kleinman A, Lock MM, Ramphele M, Reynolds P, editors. Remaking a world: violence, social suffering, and recovery. Berkeley: UC Press; 2001.
19. David L. Against standardization of memory. Hum Rights Q. 2017;39:296–319. https://doi.org/10.1353/hrq.2017.0019.
20. Ebbrecht-Hartmann T, Stiassny N, Henig L. Digital visual history: historiographic curation using digital technologies. Rethink Hist. 2023;27:1–28. https://doi.org/10.1080/13642529.2023.2181534.
21. Eshghie M, Eshghie M. ChatGPT as a therapist assistant: a suitability study. arXiv. 2023. https://arxiv.org/abs/2304.09873. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2304.09873.
22. Feldman RC. Technology law: Artificial intelligence: trust and distrust. Judges Book. 2019;3:115–22.
23. Freeman L. Digitally disappeared: the struggle to preserve social media evidence of mass atrocities. Georget J Int Affairs. 2022;23(1):105–13. https://doi.org/10.1353/gia.2022.0017.
24. Geybullayeva A. In Turkey, the state resorts to censorship majeure. Global voices. 2023. https://globalvoices.org/2023/02/13/in-turkey-the-state-resorts-to-censorship-majeure. Accessed 5 May 2023.
25. Giddens A. The consequences of modernity. Cambridge: Polity; 1990.
26. Gilardi F, Alizadeh M, Kubli M. ChatGPT outperforms crowd-workers for text-annotation tasks. arXiv. 2023. https://arxiv.org/abs/2303.15056. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2303.15056.
27. Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. Acad Manag Ann. 2020;14(2):627–60. https://doi.org/10.5465/annals.2018.0057.
28. González-Aguilar JM, Makhortykh M. Laughing to forget or to remember? Anne Frank memes and mediatization of Holocaust memory. Media Cult Soc. 2022;44:1307–29. https://doi.org/10.1177/01634437221088951.
29. Grace K, Salvatier J, Dafoe A, Zhang B, Evans O. When will AI exceed human performance? Evidence from AI experts. J Artif Intell Res. 2018;62:729–54. https://doi.org/10.1613/jair.1.11222.
30. Great AI Prompts. The complete list of banned words In Midjourney you need to know. Great AI Prompts. n.d. https://www.greataiprompts.com/imageprompt/list-of-banned-words-in-midjourney/. Accessed 4 May 2020.
31. Gruner W. Resisters: how ordinary Jews fought persecution in Hitler's Germany. New Haven: Yale University Press; 2023.
32. Guhl J, Davey J. Hosting the 'holohoax': a snapshot of holocaust denial across social media. London: The Institute for Strategic Dialogue; 2020.
33. Hern A. AI bot ChatGPT stuns academics with essay-writing skills and usability. The guardian. 2022. https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability. Accessed 5 May 2023.
34. Hoes E, Altay S, Bermeo J. Using ChatGPT to fight misinformation: ChatGPT nails 72% of 12,000 verified claims. PsyArXiv. 2023. https://psyarxiv.com/qnjkf. Accessed 28 Apr 2023; https://doi.org/10.31234/osf.io/qnjkf
35. Hoskins A. Media, memory, metaphor: remembering and the connective turn. Parallax. 2011;17:19–31. https://doi.org/10.1080/13534645.2011.605573.
36. Hoskins A. Media and compassion after digital war: why digital media haven't transformed responses to human suffering in contemporary conflict. Int Rev Red Cross. 2020;102:117–43. https://doi.org/10.1017/S1816383121000102.
37. Koenig A. et al. Digital lockers: archiving social media evidence of atrocity crimes. Berkeley: Human Rights Center at the University of California. 2021. https://humanrights.berkeley.edu/sites/default/files/digital_lockers_report5.pdf. Accessed 1 May 2023.
38. Human Rights Watch. "Video unavailable": social media platforms remove evidence of war crimes. 2020. https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes. Accessed 1 May 2023.
39. Ibrahim Y. Transacting memory in the digital age: modernity, fluidity and immateriality. Fudan J Humanit Soc Sci. 2018;11:453–64. https://doi.org/10.1007/s40647-018-0222-2.

40. Jessee E. Negotiating genocide in Rwanda: the politics of history. Cham: Palgrave Macmillan; 2017.
41. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. JNCI Cancer Spectr. 2023;7:1–9. https://doi.org/10.1093/jncics/pkad015.
42. Kahn L. The rescuers: the role of testimony as a peacebuilding tool to create empathy. In: Zucker EM, McGrew L, editors. Coexistence in the aftermath of mass violence: imagination, empathy, and resilience. Ann Arbor: University of Michigan Press; 2020. p. 149–69.
43. Kansteiner W. Digital doping for historians: can history, memory, and historical theory be rendered artificially intelligent? Hist Theory. 2022;61:119–213. https://doi.org/10.1111/hith.12282.
44. Kent L. Local memory practices in East Timor: disrupting transitional justice narratives. Int J Transit Justice. 2011;5:434–55. https://doi.org/10.1093/ijtj/ijr016.
45. Kovaliv C. Memory wars. Prandium: The Journal of Historical Studies at U of T Mississauga. 2022;11:1–10.
46. Laungaramsri P. Mass surveillance and the militarization of cyberspace in post-coup Thailand. ASEAS. 2016;9:195–213; https://doi.org/10.14764/10.ASEAS-2016.2-2.
47. Letschert R, Van Boven T. Providing reparation in situations of mass victimization. Key challenges involved. In: Letschert RN, Haveman R, de Brouwer AM, Pemberton A, editors. Victimological approaches to international crimes. Cambridge: Larcier Intersentia; 2011. p. 153–83.
48. Lewandowski D. Ranking search results. In: Lewandowski D, editor. Understanding search engines. Berlin: Springer International Publishing; 2023. p. 83–118. https://doi.org/10.1007/978-3-031-22789-9_5.
49. Li H, Guo D, Fan W, Xu M, Song Y. Multi-step jailbreaking privacy attacks on ChatGPT. arXiv. 2023. https://arxiv.org/abs/2304.05197. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2304.05197.
50. Litavrin, M. & Frenkel, D. (2023). Inside Russia's internet monitoring. How the censorship agency tracks online activity with the help of tech companies. Mediazona. https://en.zona.media/article/2023/03/08/neurorkn.
51. Long K. Ministry using ChatGPT AI to 'ease workload'; Khmer version planned. Phnom Penh Post. 2023. https://www.phnompenhpost.com/national/ministry-using-chatgpt-ai-ease-workload-khmer-version-planned. Accessed 28 Apr 2023.
52. Lundrigan M. #Holocaust #Auschwitz: performing holocaust memory on social media. In: Earl H, Gigliotti S, editors. A companion to the Holocaust. Chichester: Wiley; 2020. p. 639–55.
53. Macaulay T. Exclusive: Swiss startup unveils 'world-first' AI translation service. TWN. 2023. https://thenextweb.com/news/interprefy-startup-unveils-worlds-first-advanced-automated-speech-translation-service-for-online-and-live-eventsinterprefy, Accessed 5 May 2023.
54. Makhortykh M. Framing the Holocaust online: memory of the Babi Yar massacres on wikipedia. Stud Russ Eurasian Cent Eur New Media. 2017;18:67–94.
55. Makhortykh M. Nurturing the pain: audiovisual tributes to the Holocaust on YouTube. Holocaust Stud. 2019;25:441–66. https://doi.org/10.1080/17504902.2018.1468667.
56. Makhortykh M. Memoriae ex machina: how algorithms make us remember and forget. Georget J Int Affairs. 2021;22:180–5.
57. Makhortykh M, Bastian M. Personalizing the war: perspectives for the adoption of news recommendation algorithms in the media coverage of the conflict in Eastern Ukraine. Media War Confl. 2022;15:25–45.
58. Makhortykh M, Wijermars M. Can filter bubbles protect information freedom? Discussions of algorithmic news recommenders in Eastern Europe. Digit J. 2021. https://doi.org/10.1080/21670811.2021.1970601.
59. Makhortykh M, Urman A, Ulloa R. Hey, Google, is it what the Holocaust looked like? Auditing algorithmic curation of visual historical content on web search engines. First Monday. 2021. https://doi.org/10.5210/fm.v26i10.11562.
60. Makhortykh M, Urman A, Ulloa R. Memory, counter-memory and denialism: how search engines circulate information about the holodomor-related memory wars. Mem Stud. 2022;15:1330–45. https://doi.org/10.1177/17506980221133732.
61. Malhotra T. Meet MiniGPT-4: An open-source AI model that performs complex vision-language tasks like GPT-4. Marktechpost. 2023. https://www.marktechpost.com/2023/04/18/meet-minigpt-4-an-open-source-ai-model-that-performs-complex-vision-language-tasks-like-gpt-4/. Accessed 24 Apr 2023.
62. Manca S. Bridging cultural studies and learning science: an investigation of social media use for Holocaust memory and education in the digital age. Rev Educ Pedag Cult Stud. 2021;43:226–53. https://doi.org/10.1080/10714413.2020.1862582.
63. Manca S. Digital Holocaust memory on social media: how Italian Holocaust museums and memorials use digital ecosystems for educational and remembrance practice. Int J Herit Stud. 2022;28:1152–79. https://doi.org/10.1080/13527258.2022.2131879.
64. McGuffie K, Newhouse A. The radicalization risks of GPT-3 and advanced neural language models. arXiv. 2020. https://arxiv.org/abs/2009.06807. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2009.06807
65. Midjourney. Community guidelines. Midjourney. n.d. https://docs.midjourney.com/docs/community-guidelines. Accessed 28 Apr 2023.
66. Moon C. What remains? Human rights after death. In: Squires K, Errickson D, Márquez-Grant N, editors. Ethical approaches to human remains: a global challenge in bioarchaeology and forensic anthropology. Cham: Springer International Publishing; 2020. p. 39–58.
67. Misztal BA. Trust in modern societies: the search for the bases of social order. Cambridge: Polity; 1996.
68. Naidu E. From memory to action: a toolkit for memorialization in post conflict societies. The International Coalition of Sites of Conscience. 2016. https://www.sitesofconscience.org/wp-content/uploads/2016/08/From-Memory-to-Action-A-Post-Conflict-Memorialization-Toolkit.pdf. Accessed 28 Apr 2023.
69. Naron S, Toth G. Let them speak: an effort to reconnect communities of survivors in a digital archive. In: Zucker EM, Simon DJ, editors. Mass violence and memory in the digital age: memorialization unmoored. Cham: Palgrave Macmillan; 2020. p. 71–94.
70. Newman N, Fletcher R. Bias, bullshit and lies: Audience perspectives on low trust in the media. SSRN. 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3173579. Accessed 28 Apr 2023.
71. Pariser E. The filter bubble: how the new personalized web is changing what we read and how we think. London: Penguin; 2011.
72. Portnov A. Memory wars in post-Soviet Ukraine (1991–2010). In: Blacker U, Etkind A, Fedor J, editors. Memory and theory in Eastern Europe. London: Palgrave Macmillan; 2013. p. 233–54.
73. Quitzon J, Ear S. Hun Sen's fight to control the Cambodian infosphere. East Asia Forum. 2 May 2023. https://www.eastasiaforum.org/2023/05/02/hun-sens-fight-to-control-the-cambodian-infosphere. Accessed 4 May 2023.

74. Sarom K. ChatGPT links to robot greeter. Phnom Penh Post. 2023. https://www.phnompenhpost.com/national/chatgpt-links-robot-greeter. Accessed 24 Apr 2023.
75. Savage N. Synthetic data could be better than real data. Nature. 2023. https://www.nature.com/articles/d41586-023-01445-8?utm_campaign=Data_Elixir&utm_source=Data_Elixir_434. Accessed 5 May 2023.
76. Searle J. Minds, brains and programs. Behav Brain Sci. 1980;3:417–57.
77. Simon DJ, Zucker EM. Introduction: mass violence and memory in the digital age—memorialization unmoored. In: Zucker EM, Simon DJ, editors. Mass violence and memory in the digital age: memorialization unmoored. Cham: Palgrave Macmillan; 2020. p. 1–8.
78. Sinpeng A. Southeast Asian cyberspace: politics, censorship, polarisation. New Mandala. 2017. https://www.newmandala.org/southeast-asian-cyberspace-politics-censorship-polarisation. Accessed 28 Apr 2023.
79. Sohail SS, Madsen DØ, Himeur Y, Ashraf M. Using ChatGPT to navigate ambivalent and contradictory research findings on artificial intelligence. SSRN. 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4413913. Accessed 28 Apr 2023.
80. Sorabji C. Managing memories in post-war Sarajevo: individuals, bad memories, and new wars. J Roy Anthropol Inst. 2006;12:1–18.
81. Stargardt N. Drawing the Holocaust in 1945. Holocaust Stud. 2005;11:25–37.
82. Stray J. The AI learns to lie to please you: preventing biased feedback loops in machine-assisted intelligence analysis. Analytics. 2023;2:350–8. https://doi.org/10.3390/analytics2020020.
83. Susnjak T. ChatGPT: The end of online exam integrity? arXiv. 2022. https://arxiv.org/abs/2212.09292. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2212.09292.
84. The Economist. AI helps scour video archives for evidence of human-rights abuses. The Economist. 2021. https://www.economist.com/international/2021/06/05/ai-helps-scour-video-archives-for-evidence-of-human-rights-abuses. Accessed 28 Apr 2023.
85. Traum D, Jones A, Hays K, Maio H, Alexander O, Artstein R, Debevec P, Gainer A, Georgila K, Haase K, Jungblut K. New dimensions in testimony: digitally preserving a Holocaust survivor's interactive storytelling. In: Schoenau-Fog H, Bruni LE, Louchart S, Baceviciute S, editors. Interactive storytelling: 8th international conference on interactive digital storytelling. Switzerland: Springer International Publishing; 2015. p. 269–81.
86. Törnberg P. ChatGPT-4 outperforms experts and crowd workers in annotating political Twitter messages with zero-shot Learning. arXiv. 2023. https://arxiv.org/abs/2304.06588. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2304.06588.
87. Vaccari C, Chadwick A. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media+Society. 2020;6:1–13; https://doi.org/10.1177/2056305120903408.
88. Walden VG. What is 'virtual Holocaust memory'? Mem Stud. 2022;15:621–33. https://doi.org/10.1177/1750698019888712.
89. Walden VG (2022b). The Memorial Museum in the Digital Age. Sussex: REFRAME. https://reframe.sussex.ac.uk/wp-content/uploads/2022/12/Memorial_Museum_in_Digital_Age_Hi_res.pdf. Accessed 28 Apr 2023.
90. Winter J. The generation of memory: reflections on the "memory boom" in contemporary historical studies. Canadian Military History. 2001;10:57–66.
91. Woolford A, Muller A, Sinclair S. Risky times and spaces: settler colonialism and multiplying genocide prevention through a virtual Indian residential school. Genocide Stud Prevent Int J. 2019;13:79–96.
92. Yad Vashem. The righteous among the nations. Yad Vashem. n.d. https://www.yadvashem.org/righteous.html. Accessed 28 Apr 2023.
93. Yalirozy T. Experts warn of danger of broad use of artificial intelligence in government or education. Cambodianess. 2023. https://cambodianess.com/article/experts-warn-of-danger-of-broad-use-of-artificial-intelligence-in-government-or-education. Accessed 28 Apr 2023.
94. Yang KC, Menczer F. Large language models can rate news outlet credibility. arXiv. 2023. https://arxiv.org/abs/2304.00228. Accessed 28 Apr 2023.
95. Zarouali B, Makhortykh M, Bastian M, Araujo T. Overcoming polarization with chatbot news? Investigating the impact of news content containing opposing views on agreement and credibility. Eur J Commun. 2021;36:53–68. https://doi.org/10.1177/0267323120940908.
96. Zuccon G, Koopman B. Dr ChatGPT, tell me what I want to hear: how prompt knowledge impacts health answer correctness. arXiv. 2023. https://arxiv.org/abs/2302.13793. Accessed 28 Apr 2023; https://doi.org/10.48550/arXiv.2302.13793.
97. Zucker EM. Time, medium, memory: Remembering the Cambodian genocide. In Clavé E, Tran QT, editors. Handbook on memory studies: Southeast Asian contexts. Leiden: Brill, forthcoming.
98. Zucker EM. Trust and distrust in a highland Khmer community after thirty years of war. In: Marston JA, editor. Anthropology and community in Cambodia: reflections on the work of May Ebihara. Melbourne: Monash University Press; 2011. p. 79–103.
99. Zucker EM. Forest of struggle: memorialization of remembrance in upland Cambodia. Honolulu: University of Hawai'i Press; 2013.
100. Zucker EM. Contested narratives of victimhood: the tales of two Former Khmer Rouge soldiers. In: Klein D, editor. Societies emerging from conflict: the aftermath of atrocity. New York: Cambridge Scholars Press; 2017. p. 34–52.
101. Zucker EM, Simon DJ. Mass violence and memory in the digital age: memorialization unmoored. Cham: Palgrave Macmillan; 2020.
102. Nilsson NJ. Artificial intelligence: a new synthesis. Burlington: Morgan Kaufmann; 1998.
103. Wevers M, Smits T. The visual digital turn: using neural networks to study historical images. Digit Scholarsh Humanit. 2020;35:194–207.
104. Makhortykh M, Lyebyedyev Y, Kravtsov D. Past Is another resource: remembering the 70th anniversary of the victory day on livejournal. Natly Papers. 2021;49:375–88.
105. Salton G, McGill MJ. Introduction to modern information retrieval. New York: McGraw-Hill Book; 1983.
106. Lewandowski D. Web searching, search engines and information retrieval. Inf Serv Use. 2005;25:137–47. https://doi.org/10.3233/ISU-2005-253-402.
107. Zeynali Tazehkandi M, Nowkarizi M. Evaluating the effectiveness of Google, Parsijoo, Rismoon, and Yooz to retrieve Persian documents. Library Hi Tech. 2021;39:166–89. https://doi.org/10.1108/LHT-11-2019-0229.
108. Wallace J. Modelling contemporary gatekeeping. Digit J. 2018;6:274–93. https://doi.org/10.1080/21670811.2017.1343648.
109. Just N, Latzer M. Governance by algorithms: reality construction by algorithmic selection on the Internet. Media Cult Soc. 2017;39:238–58. https://doi.org/10.1177/0163443716643157.
110. Noble SU. Algorithms of oppression. New York: New York University Press; 2018.

111. Otterbacher J, Bates J, Clough P. Competent men and warm women: gender stereotypes and backlash in image search results. Proceedings of the 2017 conference on human factors in computing systems. New York: ACM Press; 2017. p. 6620–31.
112. Urman A, Makhortykh M. "Foreign beauties want to meet you": the sexualization of women in google's organic and sponsored text search results. New Media Soc. 2022. https://doi.org/10.1177/1461444822109953.
113. Makhortykh M, Urman A, Ulloa R. Detecting race and gender bias in visual representation of AI on web search engines. Advances in bias and fairness in information retrieval: second international workshop on algorithmic bias in search and recommendation. Cham: Springer; 2021. p. 36–50.
114. Kay M, Matuszek C, Munson SA. Unequal representation and gender stereotypes in image search results for occupations. Proceedings of the 2015 conference on human factors in computing systems. New York: ACM Press; 2015. p. 3819–28.
115. Leerssen P. The soap box as a black box: regulating transparency in social media recommender systems. Eur J Law Technol. 2020;11:1–50.
116. Gbolahan F. Towards Regulating Gender and Racial Discrimination on Google Search Engine. SSRN. 2021. Accessed 24 Jun 2023; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4046123.
117. Mittelstadt B. Automation, algorithms, and politics| auditing for transparency in content personalization systems. Int J Commun. 2016;10:4991–5002.
118. Brown S, Davidovic J, Hasan A. The algorithm audit: scoring the algorithms that score us. Big Data Soc. 2021;8:1–8. https://doi.org/10.1177/2053951720983865.
119. Bandy J. Problematic machine behavior: a systematic literature review of algorithm audits. Proc ACM Conf Hum Comput Interact. 2021;5:1–34.
120. Ulloa R, Makhortykh M, Urman A. Scaling up search engine audits: practical insights for algorithm auditing. J Inf Sci. 2022. https://doi.org/10.1177/01655515221093029.
121. Hannak A, Sapiezynski P, Molavi KA, Krishnamurthy B, Lazer D, Mislove A, Wilson C. Measuring personalization of web search. Proceedings of the 22nd international conference on World Wide Web. New York: ACM Press; 2013. p. 527–38.
122. Weber MS, Kosterich A. Coding the news: the role of computer code in filtering and distributing news. Digit Journal. 2018;6:310–29. https://doi.org/10.1080/21670811.2017.1366865.
123. Zavadski A, Toepfl F. Querying the Internet as a mnemonic practice: how search engines mediate four types of past events in Russia. Media Cult Soc. 2019;41:21–37. https://doi.org/10.1177/016344371876456.
124. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. New York: ACM Press; 2021. p. 610–23.
125. Boleda G, Herbelot A. Formal distributional semantics: introduction to the special issue. Comput Linguist. 2016;42:619–35. https://doi.org/10.1162/COLI_a_00261.
126. Manning CD. Human language understanding and reasoning. Daedalus. 2022;151:127–38. https://doi.org/10.1162/daed_a_01905.
127. Musser G. How AI Knows Things no one told it. Scientific American. Scientific American. 2023. https://www.scientificamerican.com/article/how-ai-knows-things-no-one-told-it/. Accessed 24 Jun 2023.
128. Weil E. You Are Not a Parrot. Intelligencer. New York magazine. 2023. https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html. Accessed 24 Jun 2023.
129. Barthes R. Camera Lucida, reflections on photography. New York: Hill and Wang; 1980.
130. Rutten E, Fedor J, Zvereva V. Memory, Conflict and New Media. Web Wars in Post-Socialist States. London: Routledge; 2013.

Films

131. Ex Machina. Directed by Garland E. United Kingdom: Film4 and DNA Films. 2014.
132. Her. Directed by Jonze S. Los Angeles: Annapurna Pictures. 2013.

Websites

133. "Let Them Speak" https://lts.fortunoff.library.yale.edu/ Created by Gabor Toth in collaboration with Yale Fortunoff Archive, USC Shoah Foundation, United States Holocaust Memorial Museum. Accessed April 12, 2023.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.