DIGITAL MEMORY DIALOGUES

Dialogue II

Provocation: In Search of the North Star for AI and Holocaust Memory

By Mykola Makhortykh and Maryna Sydorova

The growing experimentation with new forms of artificial intelligence technologies across various sectors presents both new possibilities and risks for Holocaust education and memory. In this provocation, we discuss whether the North Star -i.e., a fundamental vision for orienting research efforts - regarding the use of AI in the context of Holocaust remembrance can help address these risks and what challenges have to be tackled for formulating such a vision.

Artificial intelligence (AI) is often defined as a technology that enables data and information processing in a way that resembles intelligent behaviour, including aspects of reasoning, learning, planning or control. The degree to which AI should be attributed the resemblance of intelligent (human) behaviour remains debated. Recent studies highlight both the fundamental differences between human and AI reasoning, demonstrating the limited capacities of (generative) AI for understanding complex semantics, and the potential of anthropomorphising language for misinterpreting both the past and future of AI technologies, particularly in the context of human remembrance. However, the advantage of the above definition is that it captures, even if clumsily, the growing agency of AI in the context of processes traditionally associated with human intelligence, including memory (for more information, see works by Smit and colleagues (2024), Richardson-Walden and Marrison (2024), and Makhortykh (2024)).

Much of this emerging agency is attributed to the ongoing push towards adopting AI across different sectors and the integration of AI-assisted information processing and, increasingly, information generation in digital infrastructures, both in the private and public sectors. With many of these infrastructures dealing with historical information (in a broad sense), the adoption of AI inevitably affects our individual and collective memory. By retrieving, organising, and, increasingly, generating information about the past, AI models and applications - the latter ranging from chatbots to search engines to personalised news feeds - change how we remember and learn about the past. These changes concern both cognitive processes associated with individual remembering, for instance, the ability to encode and, importantly, retain information, and societal

practices of preserving and activating the past that increasingly rely on digital technologies.

Holocaust memory is not exempt from these changes. While the Holocaust heritage and education sectors are often cautious in terms of adopting new technologies – as discussed in the previous *Dialogue* and several earlier studies (for instance, by Wulf Kansteiner) – many institutions and projects have been experimenting with AI and its subfields, such as machine learning. Machine learning focuses on training models to perform tasks without explicitly programming the implementation of each task, but letting a model learn from data. Examples of such experiments include well-known cases such as *Dimensions of Testimony* or *Survivor Stories*, which use machine learning as just one component of complex infrastructures facilitating the preservation of Holocaust testimonies and keeping them engaging for future generations. In these cases, the use of AI usually remains relatively limited: for instance, it can be applied to develop a machine learning-based classifier to match the video recording (e.g. a survivor's response to a question) with a user input in the form of text or speech.

However, in addition to these cases, there is a growing number of more unconventional applications, such as *Young Again/Never Again*, *Let Them Speak* or *Auschwitz Laundry* that transform existing testimonies and memorabilia to produce new narratives or visualisations of the past. These uses of AI are characterised by the large diversity of purposes as well as technologies applied: for instance, *Young Again/Never Again* applied a selection of AI techniques, including voice cloning and facial mapping, to digitally de-age Holocaust survivors and produce video recordings with the aim of making Holocaust education more relatable for younger generations. By contrast, *Let Them Speak* applied machine learning, in particular a Latent Dirichlet Allocation technique for topic modelling, to process testimonies of Holocaust survivors in order to reconstruct the experiences of the Voiceless, namely those victims of the Holocaust who perished and did not have a chance to share their stories. *Auschwitz Laundry,* a work-in-progress at the Film Universität Babelsberg Konrad Wolf, uses (as far as we know at present) text-to-image translation to create visual representations of scenarios mentioned in testimonies for exhibition purposes.

Simultaneously, AI has been actively adopted outside heritage and education institutions to deal with Holocaust-related information. Such adoption has involved a diverse range of actors, from big tech companies such as Google or Microsoft to activist groups and collectives to individual web users. While the motivations of these actors to use AI in this particular context vary substantially, two key rationales can be inferred. Firstly, AI has become integral for organising the growing amount of digitised and digital-born content dealing with the Holocaust and available across commercial platforms,

such as YouTube, Instagram, TikTok or Google search. Similar to the institutional uses of AI, which we briefly discussed above, the specific applications of AI vary substantially across specific platforms. For instance, search engines apply AI models and frameworks, such as Multitask Unified Models or Bidirectional Encoder Representations from Transformers, to understand the meaning of human-made search queries and predict intent behind them in order to select the most relevant search results, including on topics related to the Holocaust. In the case of other platforms, like TikTok, computer vision models are combined with recommender systems to learn from the user interactions with the platform and provide them with a personalised selection of content.

Secondly, AI models have enabled the creation of new content, enhancing and reflecting upon existing Holocaust-related content, but also replacing and sometimes distorting it for purposes ranging from enriching or creatively challenging existing practices of Holocaust remembrance to propagating antisemitism and amplifying hate speech. Like before, the selection of AI models and applications which can be applied for these aims varies vastly across platforms, from the AI Alive feature on TikTok, used to transform static photos into videos, to large language models (LLMs) powering chatbots such as Gemini, Grok, or ChatGPT. In the case of Holocaust denial and distortion, the concerns are particularly pronounced regarding the potential of generative AI tools to propagate antisemitic hate speech, as in the case of the recent debacle with Grok referring to itself as MechaHitler and spreading hate speech, or Google's Bard inventing fake eyewitnesses of the Holocaust and their testimonies.

Despite the growing number of uses of AI in Holocaust memory both within and outside heritage institutions, much about them remains uncertain. Many emerging discussions remain conceptual or are driven by anecdotal evidence and examples, as well as extrapolations from other contexts. The recommendations emerging from the recent series of participatory workshops organised by the Digital Holocaust Memory Project, including the workshop on the use of AI and machine learning for Holocaust memory and education, highlight some of these uncertainties. For instance, while tensions between computational and human logic in the context of Holocaust memory are likely to emerge, the exact scope of these tensions and their implications remain unclear. To test these assumptions, more empirical (and critical) assessments are required, but so far there are only a few studies providing them (for instance, the work of Presner and colleagues (2024) and that of Makhortykh, Urman and Ulloa (2021) and, mostly, in relation to a few particular types of AI applications which are easier to assess (e.g. in terms of use or data availability).

The discussions (and concerns) about the implications of AI logic for Holocaust memory are closely connected to several other debates, for instance, regarding "Holocaust-AI

literacies". The concept acknowledges the contextual nature of digital media and AI literacies, but exactly what it means in the context of Holocaust education is still to be decided. Part of the uncertainty in this case is directly related to the limited understanding of the empirical validity of particular AI-related risks: should, for instance, such literacies emphasise the ability to cope with emotionally triggering (historical) content or proactively protect historical facts in online environments? Another discussion regards the ethical principles that should be accounted for when developing AI systems dealing with Holocaust-related information: while the importance of these principles for informing AI design is obvious, their exact selection and conceptualisation is a less trivial task. While there are a few studies discussing the potential for integrating specific principles and values in the development of AI systems used in the context of genocide-related information (e.g. Makhortykh (2023) and Zucker et al. (2024)), this problem is yet to be solved.

Consequently, while it is obvious that AI is likely to profoundly transform the field of Holocaust remembrance and education, the concrete risks and benefits of such transformation are yet to be empirically examined. For instance:

- What are the exact cognitive effects of different AI applications on individuals' ability to learn about the Holocaust, and what exactly does learning mean in this case – the ability to recall factual information? The increase in empathy?
- How do we measure these cognitive effects in the context of Holocaust memory and education, and to what degree may these effects vary according to the individual characteristics of learners or AI applications used?
- How widespread is the distribution of AI-generated historical materials regarding the Holocaust, and are there reliable means of differentiating between artificial and authentic materials, especially in the long term?
- What exactly is meant by historical authenticity? Are digitised copies of historical documents authentic, and if yes, then what about AI-facilitated translations into other languages or media, for instance, speech?
- Is authenticity always preferred to artificiality, especially considering that much of the historical materials regarding the Holocaust were made by perpetrators?
- What is the likelihood for individuals to be incidentally exposed to content denying or distorting the Holocaust in the increasingly AI-curated online environments?
- And, crucially, what is the ultimate vision - or the North Star - of using AI in Holocaust memory?

The last point is what we would like to focus on in this provocation. We suggest that much of the engagement between AI and Holocaust memory has so far been reactive, with the

latter adopting (and often adapting to) the advancements in technology. Due to many constraints, in particular financial and expertise-related, institutions dealing with Holocaust memory and education have limited capacities for pushing forward technological innovation in the field of AI. As a result, to fulfil their mission, these institutions often outsource the implementation of AI applications to third parties for internal systems and adapt their educational and commemorative strategies to account for the AI-transformed media landscape. However, recent developments, in particular regarding the increased accessibility of generative AI and its meteoric adoption by major online platforms, have resulted in a situation where earlier reactive approaches may not work anymore. Paraphrasing the famous quote from Alan Wake, what we see now is not a lake, but an ocean of possibilities (but also threats) created by AI for preserving and interacting with the past. The new memory ecosystem, which is emerging in front of our eyes, is similar to an ocean not only in its vastness but also in the complex set of factors – similar to winds and currents – which are to determine the success or failure of not only an individual memory project journey but also of the larger connective and transnational efforts countering low levels of Holocaust knowledge and increasing awareness about the relevance of Holocaust memory for today. Under these circumstances, it is crucial to identify a reference point which can be followed to navigate to the end destination and, potentially, help us consider how to nudge the development of technology in order to get there.

We use the concept of the North Star to emphasise the need to identify such a reference point and also to follow it. Historically, Polaris provided a crucial waypoint for mariners who relied on it to navigate their "odysseys of discovery" by giving them a fixed point against which their course could be calculated and corrected. In the scientific context, the North Star is a fundamental principle or a vision that orients research efforts, especially in times of disruption and uncertainty. It often has idealistic (or even utopian) underpinnings, being associated with guidance towards the public good. To a certain degree, "Never again" can be viewed as a form of North Star for Holocaust memory, but its applicability for the specific task of using AI for remembering this past can be questioned, especially at the current point in time. Such questioning is due both to the rather general focus of the "Never again" argument and the existing criticism, for instance, regarding the potential ambiguity of the vision (e.g. how it shall be achieved and how broadly/narrowly it shall be applied) together with the mounting evidence of the practical impossibility of fulfilling the promise (as exemplified in the argument that in reality we face the "time and again" situation). Considering the limited capacities of current forms of AI to understand complex ethical issues and act on them accordingly, it is rather dubious to expect it to be able to comprehend why "Never again" is important

and also follow this principle, considering that its human creators consistently fail to do it themselves.

Under these circumstances, we suggest that we need to find the North Star for AI in Holocaust memory and education that can serve as a beacon, providing the long-term direction for the adoption of technology and highlighting the overarching goals of its use. We also acknowledge that it is not an easy or immediate process: unlike Polaris, which is usually relatively easy to locate in the night sky (at least in the Northern Hemisphere), the process of identifying the North Star for Holocaust remembrance faces multiple challenges that have to be addressed. In this provocation, we do not aim to provide an exhaustive list of these challenges, but instead, focus on three challenges that we think are of particular relevance for the current moment and which we would like to explore in more detail in this Dialogue.

The first challenge regards the diversity of AI applications in the context of Holocaust remembrance. Current AI models are yet to reach the potential associated with artificial general (and not narrow) intelligence – i.e. AI capable of matching or even exceeding human capacities for any particular task – and there is debate about whether it will be possible and what technical advancements will be required for this. However, despite these limitations, present forms of AI already enable an extremely broad set of possible use cases. Drawing from the existing function-based typologies of AI applications for cultural heritage (for example, see Weicong et al. (2025) and Gîrbacia (2024)), we can group these use cases into the following broad categories:

- *Organisation and retrieval of Holocaust-related information*. For these use cases, AI is applied to help individuals and institutions navigate the large volumes of information about the Holocaust. In the case of heritage institutions, the most well-known is the Arolsen Archives applying AI to index documents and extract structured information from them. These use cases are also particularly prominent outside of the heritage sector, with many platforms, for instance, search engines, applying AI to retrieve information regarding the Holocaust in response to user search queries, or recommending Holocaust-related content in response to user interest.

- *Storytelling about the Holocaust*. This category includes a broad range of non-generative and generative AI applications which facilitate the transmission of Holocaust memory in different immersive formats. One of the most recognisable formats of such transmission is the creation of interactive biographies of Holocaust survivors (e.g. as in the case of *Dimensions for Testimony*). However, there are also other cases: for instance, the use of generative AI applications, such as Midjourney, to help Holocaust survivors reconstruct and visualise their

memories, as in the case of the project run by Chasdei Naomi in Israel. The accessibility of AI applications also makes it easier for non-institutional actors to engage in Holocaust storytelling, for instance, by producing static images and videos using commercial AI applications and then disseminating them online. Some recent examples include AI-generated visuals for political messages on Instagram commemorating the liberation of the Holocaust camps and fake images of Holocaust victims (e.g. a prisoner playing a violin in Auschwitz) shared on X and Facebook.

- *Analysis of Holocaust-related historical materials and their presentation*. These use cases apply AI capacities for information processing and pattern recognition to produce insights about specific aspects of Holocaust-related materials. Examples range from Yad Vashem applying AI-facilitated named-entity recognition to process large volumes of historical materials to identify victims' names and other historical details to the *From Numbers to Names* project using face recognition to identify victims on historical photos. Other projects like *Decoding Antisemitism* apply AI for identifying different forms of antisemitic hate speech, including Holocaust denial, in online platform data. The latter uses are also common outside heritage institutions, with online platforms increasingly adopting AI for automated content moderation that includes analysing content that is moderated, even while the effectiveness of such moderation remains debated (e.g. see this study for Facebook).

An immediate consequence of such diversity is the difficulty of establishing a common vision of how these diverse applications of AI are expected to contribute to sustaining memory about the Holocaust and to drawing lessons from it, and what criteria can be used to assess whether applications meet these aims. Perhaps, we should start from the opposite position: is it possible to agree on what these applications should not do and search for the North Star by defining what it certainly is not?

The second challenge is closely related to the first one. Not only is the use of AI in the context of Holocaust memory characterised by an impressive diversity of applications, but also, increasingly, a diversity of actors using them. While much debate on AI until now focuses on its use within heritage institutions (as reflected in many typologies tracing such uses), the applications of AI outside institutions remain somewhat neglected, despite their number and thus significance for Holocaust remembrance quickly growing. As companies like Google and OpenAI become universal information gatekeepers, they also (even if unwillingly) become gatekeepers of information about the Holocaust. Furthermore, these companies increasingly provide possibilities to use AI to generate new content, including (if not safeguarded against) content about the

Holocaust, in text, image, or video format. To a certain degree, it makes these companies Holocaust-related content providers, together with the individual and collective actors using these companies' affordances to creatively engage with the past.

Such an expansion of actors has very immediate implications for the sector and the potential vision guiding its future. Probably, the most obvious implication is the massive disruption of Holocaust memory, which we have already started to observe. In addition to hallucinations and factually incorrect statements produced by commercial AI applications in relation to the Holocaust, there is also a growing amount of AI-made content representing and interpreting the Holocaust in different ways that may soon flood digital environments. Examples of such content range from TikTok videos showcasing, with the help of AI, what an inmate's life in Auschwitz would look like to fake historical materials, for instance, photos of apparent Holocaust victims, to AI-generated images propagating antisemitic messages and promoting Nazi leaders. The implications of this are yet unclear: will it invigorate the interest of the general public to further explore history and historical mass atrocities? Or will it undermine the trust of the public towards historical evidence, including evidence of the Holocaust, or amplify already increasing levels of cynicism towards present and historical suffering? And how can we account for these risks when looking for the North Star of AI?

However, these immediate risks for historical authenticity and its perception by the public are not the only implications of the growing role of tech companies in the context of Holocaust-related information. The current situation highlights that, unlike heritage institutions, commercial platforms usually do not perceive the application of technology towards history in general and the Holocaust in particular as taboo. While some companies try to implement safeguards to address the problem of their AI applications propagating distortion and denial, these efforts often lack consistency that in some cases results in impeding legitimate forms of engagement with Holocaust memory. While we can debate the reasons for it at the current point in time - lack of expertise and understanding of the importance of the past? Lack of commercial incentive? The reality is that the impact of non-conventional actors on Holocaust education and memory will keep growing. Consequently, any vision of the future in this domain will need to account for it. For instance, how to preserve the impact of digital museums and archives if platforms such as Google will apply AI to shift away from their traditional role as information gatekeepers and gateways to the larger universe of content to become "answer machines"? Currently, this process is already happening as users can find answers to their questions (potentially about the Holocaust) directly via chatbots embedded in search engines. This phenomenon leads to the cannibalisation of external website visits.

The third challenge regards the lack of clarity about how (in particular, non-expert) users interact with AI systems to acquire information about the Holocaust and what effects this may have for these users' engagement with Holocaust memory, especially outside Holocaust institutions. Many of these interactions occur on a one-to-one basis, for instance, when individuals use search engines or chatbots to seek information about a specific aspect of the Holocaust or are incidentally exposed to it by different types of (commercial) recommender systems (e.g. YouTube's algorithms). As a result, it is difficult to assess how common such interactions are (e.g., how many individuals actually search for information about the Holocaust or, alternatively, Holocaust distortion) and what implications this could have for Holocaust knowledge and attitudes towards Holocaust remembrance. While there was a growing number of AI audits looking at what outputs AI models and applications can produce in response to Holocaust-related user queries (see, for examples, our studies on search engines and AI chatbots), these approaches have many limitations due to the difficulties of modelling authentic user behaviour and simulating multiple signals which can affect AI outputs.

This lack of transparency regarding user interactions with the Holocaust-related content is particularly pronounced in the case of commercial platforms using AI information curation. It is attributed both to the extreme diversity of Holocaust-related content with which individuals can potentially interact and the increasingly hyper-personalised nature of such interactions due to AI adapting content selection (and, increasingly, format) to individual user preferences. However, similarly obscure are some uses of AI applications in the case of Holocaust institutions: for instance, how many individuals ask questions to digital copies of survivors via web interfaces and also what questions do they ask?

Acquiring such information is crucial for refining the North Star of AI in the context of Holocaust memory. Without understanding how exactly AI is used and misused, it is hardly possible to identify what the current problems are and how they can be solved (or worsened) in the future. Such understanding shall also involve scrutinising motivations for the use of AI – both within heritage and education institutions and outside them. What exactly do people making AI-generated videos of the typical day of an Auschwitz prisoner want to achieve: do they want to gain more likes? Do they want to attract attention to Holocaust memory? Do they want to educate people in a way they see as an effective means under the conditions of today's (online) attention economy? Similarly important is understanding the actual effects of engagement with AI: can it increase empathy? Undermine trust towards history and facts? How much might these effects vary across particular groups of users? Trying to follow the North Star without addressing these questions will be similar to following Polaris without accounting for the ocean currents.

None of these challenges is trivial to deal with, but it does not mean that the search for the North Star for the use of AI in Holocaust memory and education is an impossible mission. While at the current moment AI poses many risks, from amplifying denialism to undermining the credibility of historical evidence due to the emerging spread of artificial Holocaust content, it also offers a unique opportunity to consider what the future of Holocaust memory could be in the AI and possibly post-AI age. It is hard to disagree with the premise that advancements in AI may have a disruptive impact on Holocaust education and memory, but the outcome of such a disruption should encourage a critical reconsideration of what these sectors have managed to achieve in recent decades, and what they did not achieve. Which in turn, should make us reflect on whether the use of AI could improve the current state of affairs or not. In turn, such reconsideration can create more possibilities for nudging the development of disruptive AI technologies towards more societally desirable forms and impacts, especially as these technologies are still at relatively early stages of development and adoption.

We would like to end this provocation by formulating several questions to the respondents, which, in our view, are particularly relevant for assessing the relationship between AI and Holocaust memory and finding the North Star for it:

- Should the use of AI in the context of Holocaust memory and education be regulated (or at least self-regulated)? If yes, then what principles should be used and who should (or can) enforce them?

- Can we learn how AI is used for learning about the Holocaust and remembering it, especially outside heritage institutions? What are the technical and ethical challenges of acquiring such knowledge, and how can these challenges be countered?

- Is there a way to protect the authenticity of Holocaust evidence and preserve (or restore) public trust in historical facts when facing a wave of AI-generated histories and memories? What do we mean by authenticity, and how do we measure it?

- How can we imagine AI affecting the work of heritage and education institutions in the near future? Will archives and museums limit access to their collections to prevent data cannibalisation for AI training or adopt new functions (e.g. of historical fact-checkers)?

- What could motivate the general public to rely on Holocaust-specific – or even Holocaust-sensitive – AI systems and models more than (likely more accessible) general use applications like ChatGPT or Gemini?

- How general or specific should our vision of the (digital) future of Holocaust memory and education be, and how can it be translated into a transnational and inter-organisational action?