SSOAR

Open Access Repository
www.ssoar.info

Does it get better with time? Web search consistency and relevance in the visual representation of the Holocaust

Makhortykh, Mykola; Urmann, Aleksandra; Ulloa, Roberto; Sydorova, Marya; Kulshrestha, Juhi

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

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

Empfohlene Zitierung / Suggested Citation:
Makhortykh, M., Urmann, A., Ulloa, R., Sydorova, M., & Kulshrestha, J. (2024). Does it get better with time? Web search consistency and relevance in the visual representation of the Holocaust. In E. Pfanzelter, D. Rupnow, É. Kovács, & M. Windsperger (Eds.), *Connected Histories: Memories and Narratives of the Holocaust in Digital Space* (pp. 13-32). Berlin: De Gruyter Oldenbourg. https://doi.org/10.1515/9783111329154-002

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

Terms of use:
This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more Information see:
https://creativecommons.org/licenses/by-nc/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-98241-1

Mykola Makhortykh, Aleksandra Urman, Roberto Ulloa,
Marya Sydorova, Juhi Kulshrestha

Does it get better with time? Web search consistency and relevance in the visual representation of the Holocaust

1 Introduction

Today's digital memory ecosystem is characterised by the abundance of memory-related content. Often referred to as a definitive feature of the "post-scarcity"[1] memory culture, such abundance is particularly pronounced in the case of the recent past, where the volume of content dealing with individual and collective recollections has increased dramatically due to the rise of portable digital devices (e.g. smartphones) generating this content and online storage spaces preserving it.[2] However, even in the case of historical events which happened decades or centuries ago, there is a growth in available content attributed both to the digitisation of institutional collections and the production of digital-born content.[3] In some cases, digital-born content is generated by the memory institutions (e.g. in the case of online materials countering the distortion of historical facts[4]), whereas in other cases such content originates from the users interested in commemorat-

1 Andrew Hoskins, "7/7 and connective memory: Interactional trajectories of remembering in post-scarcity culture," *Memory Studies* 4(3) (2011): 269–280.

2 Andrew Hoskins and Huw Halstead, "The new grey of memory: Andrew Hoskins in conversation with Huw Halstead," *Memory Studies* 14(3) (2021): 675–685.

3 See, for example, Matilde Eiroa San Francisco, "Primary sources for a digital-born history: The Hispanic blogosphere on the Spanish Civil War and Franco's regime," *Culture & History Digital Journal* 7(2) (2018): 1–10; Stefania Manca, Marcello Passarelli and Martin Rehm, "Exploring tensions in Holocaust museums' modes of commemoration and interaction on social media," *Technology in Society* 68 (2022): 1–13.

4 Manca, Passarelli and Rehm, "Exploring tensions in Holocaust museums' modes of commemoration and interaction on social media," 5–6.

Acknowledgments: This work was supported by the Alfred Landecker Foundation as part of the project titled "Algorithmic turn in Holocaust memory transmission: Challenges, opportunities, threats". The funder had no role in study design, data collection and analysis, the decision to publish, or manuscript preparation.

Open Access. © 2024 the author(s), published by De Gruyter. BY-NC This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
https://doi.org/10.1515/9783111329154-002

ing a specific personality or an episode of the past[5] or confronting hegemonic memory practices.[6]

The abundance of memory-related content enables new possibilities and challenges for individual and collective remembering. Because of the extensive volume of information about the past, it becomes difficult for individuals to navigate it without relying on digital mnemotechnologies. Defined by Stiegler[7] as a form of technology used for systematically ordering memories, mnemotechnologies come in different formats: from external storage devices (e.g. compact disks[8]) to the algorithm-driven platform affordances (e.g. the *On This Day* function of Facebook[9]). However, independently of the format, most digital mnemotechnologies can organise memory-related content and then retrieve it in response to explicit user requests (e.g. search queries) or implicit signals (e.g. internal system timers).

Among many digital mnemotechnologies, a special place is occupied by web search engines such as Google or Yandex. Search engines are algorithmic systems responsible for collecting Web data and retrieving and ranking these data in response to user queries.[10] The functionality of the search engines makes them one of the crucial types of information gatekeepers,[11] which determine what information sources and interpretations are prioritised to their users and how such prioritisation changes over time. Under these circumstances, search engines play a key role in shaping how the general public understands not only the present,[12] but also the past state of social reality.[13]

---

5 Yvonne Liebermann, "Born digital: The Black lives matter movement and memory after the digital turn," *Memory Studies* 14(4) (2021): 713–732.

6 Mykola Makhortykh and Maryna Sydorova, "Animating the subjugated past: digital greeting cards as a form of counter-memory," *Visual Communication* 21(1) (2022): 28–52.

7 Bernard Stiegler, *For a New Critique of Political Economy* (Cambridge: Polity Press, 2010), 67.

8 Bruno Lessard, "Archiving the gaze: Relation-images, adaptation, and digital mnemotechnologies," in *Save as … Digital memories*, ed. Joanne Garde-Hansen, Andrew Hoskins and Anna Reading (London: Palgrave Macmillan UK, 2009), 115–128.

9 Robert Prey and Rik Smit, "From personal to personalized memory: Social media as mnemotechnology," in *A networked self and birth, life, death*, ed. Zizi Papacharissi (London: Routledge, 2018), 209–223.

10 Dirk Lewandowski, *Understanding Search Engines* (Berlin: Springer Nature, 2023), 1.

11 Rasmus Kleis Nielsen, "News media, search engines and social networking sites as varieties of online gatekeepers," in *Rethinking journalism again. Societal role and public relevance in a digital age*, ed. Chris Peters and Marcel Broersma (London: Routledge, 2016), 93–108.

12 Juhi Kulshrestha, et al., "Search bias quantification: investigating political bias in social media and web search," *Information Retrieval Journal* 22 (2019): 188–227.

13 Andrei Zavadski and Florian Toepfl, "Querying the Internet as a mnemonic practice: how search engines mediate four types of past events in Russia," *Media, Culture & Society* 41(1) (2019): 21–37.

Despite the importance of search engines as a mnemotechnology amplified by their extensive use by internet users[14] and high trust towards them,[15] there are still many uncertainties about how search engines impact individual and collective engagement with the past. In addition to the concerns about search engines limiting the individual agency in the context of remembrance (e.g. by preventing individuals from controlling what information about their past is prioritised[16]), search engines are highly complex and highly non-transparent systems. The lack of transparency is particularly concerning when search engines are dealing with information about traumatic episodes of the past (e.g. mass atrocities) due to ethical and moral obligations to protect the memory of the victims and prevent it from being distorted and the frequent attempts to distort or appropriate memory about these episodes.

To better understand search engine performance in the context of mass atrocities, this chapter examines how three search engines – Bing, Google, and Yandex – prioritise visual information about the Holocaust. Our focus on the Holocaust is attributed to it being an iconic instance of mass atrocity (as well as one of the most studied and documented atrocities in the history of humankind) and the one which is often contested by denialist and antisemitic claims. By conducting a virtual agent-based algorithm audit, we investigate what search engines perceive as the most relevant content for the visual representation of the Holocaust and how such perception changes over time and address the following research questions: How thematically/historically relevant is visual content retrieved in relation to the Holocaust by search engines? What aspects of the Holocaust become more/less visible over time, and how does such visibility vary across individual search engines? What Holocaust sites become more/less visible over time, and what factors (e.g. historical significance) can influence their visibility?

14 Aleksandra Urman and Mykola Makhortykh, "You are how (and where) you search? Comparative analysis of web search behavior using web tracking data," *Journal of Computational Social Science* (2023): 1–16.

15 Sebastian Schultheiß and Dirk Lewandowski, "Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference," *Journal of Information Science* 49(3) (2023): 609–623.

16 Elena Esposito, "Algorithmic memory and the right to be forgotten on the web," *Big Data & Society* 4(1) (2017): 1–11.

2 Literature review

The research on digital mnemotechnologies is a quickly growing academic field located at the intersection of memory and media studies.[17] Originating from mnemotechnics used to preserve individual recollections and ranging from material (e.g. cave paintings[18]) to mental techniques (e.g. memory palaces[19]), mnemotechnologies were enabled by the industrialisation of memory-related processes via mass media that resulted in new large-scale formats of memory storage and transmission.[20] The immediate consequence of the rise of mnemotechnologies was the increase in the volume of preserved memory-related content that, in turn, created the pressing need to organise such content.

The advancement of digital technologies, particularly the establishment of Web 2.0, characterised by the ease of producing new content, enabled new possibilities and risks for organising memories via mnemotechnologies. Instead of focusing on storing information about the past (e.g. via external storage devices[21]), mnemotechnologies increasingly shifted towards new mechanisms for determining the relevance of stored information for individual users and delivering it in a new range of formats. In the case of heritage institutions, examples of such mnemotechnologies range from search systems integrated with institutions' collections of digitised and digital-born materials[22] to the use of conversational agents (e.g. in the form of holograms[23]) for delivering information about the past to the public.

However, the major change caused by Web 2.0 in the context of remembrance was related to the increasing presence of memory-related content outside of spaces curated by heritage institutions. Under these circumstances, mnemotechnologies have become adopted by commercial platforms such as Facebook/Meta

17 For examples, see Prey and Smit, "From personal to personalized memory"; Neal Thomas, "Social computing as a platform for memory," *Culture Machine* 14 (2013): 1–16; Jacek Smolicki, "You Press the Button, We Do the Rest. Personal Archiving in Capture Culture," *Towards a philosophy of digital media* (2018): 77–100.

18 Prey and Smit, "From personal to personalized memory," 221.

19 Frances A. Yates, *The art of memory* (London: Routledge and Kegan Paul, 1966), 46.

20 Stiegler, *For a New Critique of Political Economy*, 67.

21 Bruno Lessard, "Archiving the gaze."

22 Jefferson Heard, et al., "On mediated search of the United States holocaust memorial museum data," in *Next Generation Information Technologies and Systems. 7th International Conference, NGITS 2009 Haifa, Israel, June 16–18, 2009 Revised Selected Papers*, ed. Yishai A. Feldman, Donald Kraft and Tsvi Kuflik (Heidelberg: Springer, 2009), 38–46.

23 Victoria Grace Walden, "What is 'virtual Holocaust memory'?" *Memory Studies* 15(4) (2022): 621–633.

and Google[24] to deal with the evergrowing volume of content related to the individual and collective past. The constant growth in the volume of such content, the multiple formats in which it appeared and the multitude of interpretations it provided (including frequent attempts to distort the past) and diverse uses of memory-related content by individual users makes the use of mnemotechnologies "by commercial platforms" more challenging compared to the case of institutions. These complexities contributed to the commercial platforms adopting more advanced forms of mnemotechnologies capable of personalising user interactions with memory-related content, filtering out undesired content, and adapting to the appearence of more relevant sources and content items over time.

The variety of commercial mnemotechnologies has been attracting an increasing amount of scholarly attention in recent years, with studies looking at their use by social media platforms (e.g., Facebook[25] or Instagram[26]), video hosting services (e.g., TikTok[27]) or mobile apps (e.g., Apple Memories[28]). Web search engines are no exception in this case, with the volume of research on their role in today's digital memory ecosystem steadily growing. Pfanzelter[29] used aggregate data provided by Google Trends to look at the dynamics of the use of Google search in the context of the Holocaust and found that the use of Google for searching for Holocaust-related information in Germany was more consistent than internationally. Zavadski and Toepfl[30] conducted a large-scale study looking at how Google and Yandex text search algorithms represent information about different episodes of Russian history and found that both search engines tend to prioritise

24 See, for instance, Benjamin N. Jacobsen, "Sculpting digital voids: The politics of forgetting on Facebook," *Convergence* 27(2) (2021): 357–370; Rik Smit, Ansgard Heinrich and Marcel Broersma, "Witnessing in the new memory ecology: Memory construction of the Syrian conflict on YouTube," *New Media & Society* 19(2) (2017): 289–307.

25 Prey and Smit, "From personal to personalized memory."

26 Jennifer Krueckeberg, "Youth and Algorithmic Memory: Co-producing Personal Memory on Instagram," in *Culture and Computing. Design Thinking and Cultural Computing. 9th International Conference, C&C 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II*, ed. Matthias Rauterberg (Cham: Springer, 2021), 253–264.

27 Tom Divon and Tobias Ebbrecht-Hartmann, "#JewishTikTok: The JewToks' Fight against Antisemitism," in *TikTok cultures in the United States*, ed. Trevor Boffone (London: Routledge, 2022), 47–58.

28 Gabriel Pereira, "Apple Memories and Automated Memory-Making: The Networked Image Inside the iPhone Chip," *Digital Culture & Society* 7(2) (2021): 203–228.

29 Eva Pfanzelter, "At the crossroads with public history: Mediating the Holocaust on the Internet," *Holocaust Studies* 21(4) (2015): 253–254.

30 Zavadski and Toepfl "Querying the Internet as a mnemonic practice," 32–33.

pro-regime information sources. Finally, Makhortykh et al.[31] conducted algorithm audits of text search for a selection of search engines in relation to the Holocaust and the Holodomor and identified substantial differences in the types of information sources prioritised by different search engines regarding these two atrocities.

At the same time, despite the growing number of research inquiring into the implications of using web search engines as a form of mnemotechnology, there are still many gaps which are to be addressed, in particular considering the high complexity of algorithms underlying web search functionalities[32] and their tendency to change over time. So far, most studies focus on a single type of web search – i.e. text[33] – whereas the other forms of search (e.g. image or video search) remain less researched despite the importance of visual content for memory transmission.[34] Similarly, there is a tendency to focus on a single search engine – i.e. Google – despite the contrasting observations about how similar or different interpretations of the past prioritised by different search engines can be.[35] Furthermore, many studies focus on search data collected at a single point of time, despite a few comparative studies on representation of mass atrocities via web search indicating substantial fluctuation in

31 Mykola Makhortykh, Aleksandra Urman and Roberto Ulloa, "Memory, counter-memory and denialism: How search engines circulate information about the Holodomor-related memory wars," *Memory Studies* 15(6) (2022): 1330–1345; Mykola Makhortykh, et al., "Can an algorithm remember the Holocaust? Comparative algorithm audit of Holocaust-related information on search engines," in *Digital Memory. Neue Perspektiven für die Erinnerungsarbeit*, ed. Iris Groschek and Habbo Knoch (Göttingen: Wallstein Verlag, 2023), 79–93.

32 For the discussion of some factors influencing web search performance, in particular in the context of them deciding on the relevance of sources, see Lewandowski, "Understanding Search Engines"; Olof Sundin, Dirk Lewandowski and Jutta Haider, "Whose relevance? Web search engines as multisided relevance machines," *Journal of the Association for Information Science and Technology* 73(5) (2022): 637–642; Aniko Hannak, et al., "Measuring personalization of web search," in *Proceedings of the 22nd international conference on World Wide Web* (2013), 527–538.

33 For instance, Pfanzelter, "At the crossroads with public history"; Zavadski and Toepfl, "Querying the Internet as a mnemonic practice," Makhortykh, et al., "Memory, counter-memory and denialism."

34 For the importance of visual content for memory transmission, specifically in the context of the Holocaust, see Barbie Zelizer, *Remembering to forget: Holocaust memory through the camera's eye* (Chicago: University of Chicago Press, 1998); Sarah Farmer, "Going Visual: Holocaust Representation and Historical Method," *The American Historical Review* 115(1) (2010): 115–122. For a few studies looking at the role of image search on representation the historical and recent past, see Monica Lestari Paramita, et al., "Do you see what I see? Images of the COVID-19 pandemic through the lens of Google," *Information Processing & Management* 58(5) (2021): 1–16; Mykola Makhortykh, Aleksandra Urman and Roberto Ulloa, "Hey, Google, is it what the Holocaust looked like? Auditing algorithmic curation of visual historical content on Web search engines," *First Monday* 26(10) (2021): 1–24.

35 For instance, in Zavadski and Toepfl, "Querying the Internet as a mnemonic practice," researchers found little difference between memory-related content prioritised by Google and Yandex in relation to Russian history, whereas in Makhortykh, et al., "Memory, counter-memory and

how the past is represented by search engines over time.[36] Under these circumstances, it is important to investigate whether relevance of content prioritised by the search engines in relation to the visual representation of the past is also subject to fluctuations and how these fluctuations can influence the consistency of the historical representation.

3 Methodology

3.1 Data collection

To implement the study, we conducted a virtual agent-based algorithm audit of image search engines. Unlike other forms of algorithm audits which rely on crowdsourcing data from human users or the use of application programming interfaces (APIs),[37] virtual agent-based audits (also known as sock puppet audits) simulate human browsing behaviour (e.g. entering a search query[38]) via specialised software or programming scripts. The major advantage of virtual agent-based audits is that they allow controlling for personalisation[39] and randomisation[40] factors that can influence what content the search engine prioritises for specific queries. To control these factors, virtual agent-based audits allow the deployment of a large number of agents simultaneously (to account for some agents getting a more random selection of the outputs), using the IPs from the same location, and synchronising the activity of the agents to control for the effect of time.

To conduct the audits, we built a network of CentOS virtual machines (50 machines for the audit on 27 February 2020 and 30 machines for the audit on 18 March 2021) in the Frankfurt am Main region of the Amazon Elastic Compute Cloud. The deployment of machines in the Frankfurt region implied that each of

denialism," rather profound differences between Google and Yandex concerning whether the Holodomor was a genocide were observed.

36 For instance, Makhortykh, et al., "Memory, counter-memory and denialism"; Makhortykh, et al., "Can an algorithm remember the Holocaust?".

37 For a review, see Jack Bandy, "Problematic machine behavior: A systematic literature review of algorithm audits" *Proceedings of the ACM on human-computer interaction 5* (2021), 1–34.

38 Roberto Ulloa, Mykola Makhortykh and Aleksandra Urman, "Scaling up search engine audits: practical insights for algorithm auditing," *Journal of information science* (2022), https://doi.org/10.1177/01655515221093029.

39 Hannak, et al., "Measuring personalization of web search."

40 Mario Haim, Florian Arendt and Sebastian Scherr, "Abyss or shelter? On the relevance of web search engines' search results when people google for suicide," *Health Communication* 32(2) (2017): 253–258.

our simulated agents had an IP address coming from the same part of Germany to control for location-based personalisation. We deployed two virtual agents on each machine: one in the Chrome browser and one in Mozilla Firefox. Each agent consisted of two browser extensions: a tracker and a bot. The tracker collected the HTML of pages visited in the browser and sent them to a storage server. The bot emulated a sequence of actions, such as visiting an image search engine page, entering the "Holocaust" query, scrolling down the results page to load at least 50 images, and cleaning data accessible by the browser and the search engine's JavaScript to prevent earlier searches affecting the subsequent ones.

For the audits, we focused on three search engines – Google, Bing, and Yandex – between which we distributed a similar number of agents (i.e. 33 to 34 per engine for 2020 and 20 per engine for 2021). Our selection was attributed to these search engines being among the most often used worldwide: Google is currently the monopolist on the search market in most countries in the Global North, whereas Bing is usually the second most used search engine there. By contrast, Yandex, the largest Russian search engine, is prevalent in many post-Soviet countries. By comparing these search engines, we were interested to see what the implications of different sets of algorithmic principles are for the changing relevance in the visual representation of the Holocaust as well as whether such representation is influenced by the growing tendency of Yandex to reiterate pro-Kremlin narratives[41] (e.g. in the form of prioritising images showing perpetrators of Ukrainian origin or coming from the sites where such perpetrators were particularly active as part of an effort to stigmatise Ukrainians in the context of the Russian-Ukrainian war).

Our decision to focus on image search and not other forms of search (e.g. news or text search) is due to several reasons. First, visual content is known to be particularly effective in communicating meanings in relation to phenomena which can be difficult to describe verbally,[42] including mass atrocities such as the Holocaust.[43] Second, as we noted earlier, research on the role of web search engines as a form of mnemotechnology has focused on text search. In contrast, the role of image search in this context remained less studied despite the above-mentioned potential of visual content and its ability to stir strong affective reactions that make it additionally relevant in the context of remembrance.

41 Daria Krvets, et al., "Different platforms, different plots? The Kremlin-controlled search engine Yandex as a resource for Russia's informational influence in Belarus during the COVID-19 pandemic," *Journalism* (2023), https://doi.org/10.1177/14648849231157845.

42 Roland Bleiker, "Mapping visual global politics," in *Visual Global Politics*, ed. Roland Bleiker (New York: Routledge, 2018), 9.

43 Thomas Trezise, "Unspeakable," *The Yale Journal of Criticism* 14(1) (2001): 39–66.

3.2 Data analysis

For our analysis, we extracted URLs of image search results for each agent aggregated them across agents, and selected the 50 most frequent search results per search engine. These images were manually examined by one of the authors, who is a trained historian experienced in working with Holocaust-related archival and digital content. This examination aimed to detect the location and time at which the image was produced so it would be possible to identify whether it was related to the Holocaust (and not another historical event) and facilitate the identification of which aspect of the Holocaust the image showed.

The attribution process was facilitated with the use of authoritative information sources (e.g. the US Holocaust Memorial Museum collections) and existing knowledge of Holocaust materials by the author. This allowed to narrow the search process to specific locations/episodes and then verify the initial attribution. In these cases, when no direct match was found, the reverse search in Google and Yandex was used to locate the source of the image.

After the process of attribution was complete, we used descriptive statistics to determine how the proportion of content relevant to the Holocaust was changing over time; the same approach was used to evaluate the changes in the proportion of historical (i.e. content produced in the course of the Holocaust and immediately after it) to non-historical content (i.e. content produced in the post-war period). After it, we manually labelled each image to identify what aspect of the Holocaust it shows (e.g. deportations, life in camps, or post-war commemoration). Two coders conducted the labelling for 2020 data regarding the aspect of the Holocaust shown with the intercoder reliability tests showing high agreement (Krippendorff's alpha of 0.84); for 2021 data, the labelling was conducted by a single coder. The final step of the analysis again relied on descriptive statistics and examined the proportion of image outputs showing content from specific Holocaust extermination and concentration camps.

4 Findings

4.1 Relevance and historicity of web search outputs

We started our analysis by examining how relevant to the Holocaust image search outputs were in 2020 and 2021. Figure 1a shows that the relevance was high in 2020, with the lowest scores observed for Bing, where 91 percent of outputs were related to the Holocaust. A few examples of non-relevant outputs included content showing

images related to other mass atrocities (e.g. ethnic cleansing of Rohingya in Myanmar). In 2021, however, the number of relevant outputs increased even further, with 100 percent of outputs for Bing and Yandex being related to the Holocaust. Such a change can indicate an improvement in the performance of the algorithms powering web search engines, which prioritised more relevant content.

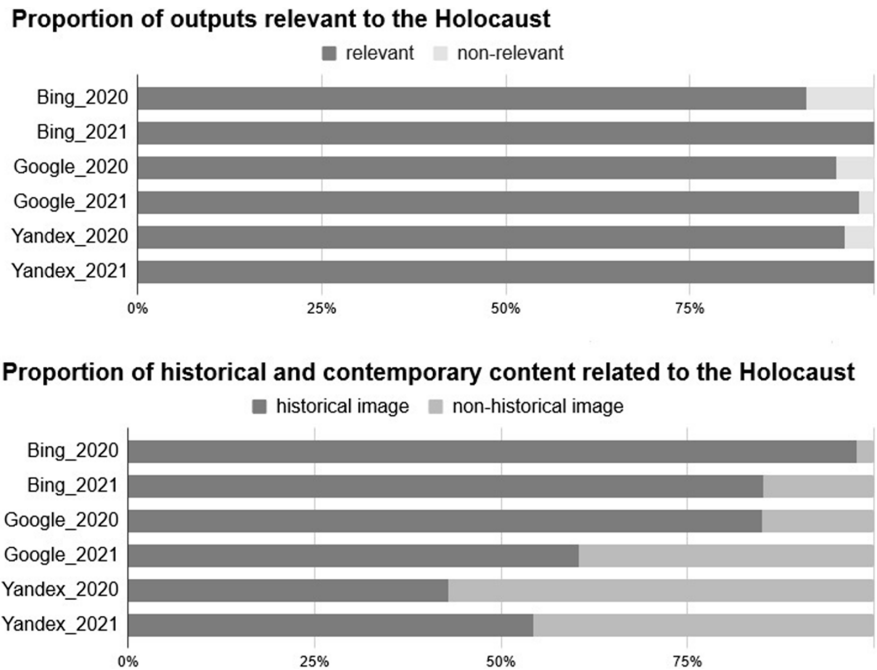

Figure 1: (a) Proportion of outputs relevant to the Holocaust. (b) Proportion of historical images among outputs relevant to the Holocaust.

Unlike relevance which increased over time, in the case of the historicity of outputs, we observed a reversed trend. Figure 1b demonstrates that for Bing and Google, in 2021, the number of outputs showing historical content (i.e. photos made in the course of the Holocaust or immediately after the liberation of the camps) has decreased from 98 percent to 85 percent and from 85 percent to 60 percent respectively. The Yandex is more of an exception in this context due to the increase in historical content in the search outputs; however, the proportion of such content remained lower than for Google and Bing.

The shift towards the present-day aspects of Holocaust remembrance can be attributed to several reasons. One of them is the tendency of search engines to prioritise content coming from journalistic outlets, both due to these sources

being viewed as more relevant by the search algorithms and journalistic outlets being more experienced in search engine optimisation. Under these circumstances, more present-day images are likely to appear due to journalistic outlets focusing on the latest developments, for instance, related to the commemoration of the Holocaust. At the same time, the higher visibility of present-day content can be related to the growing emphasis on using search engines for advertising and promoting Holocaust tourism.

4.2 Holocaust aspects prioritised by web search outputs

Following the analysis of the relevance and historicity of web search outputs, we shifted to examining what aspects of the Holocaust are more and less visible in the outputs. Figure 2a shows that all three engines put a major emphasis on content showing the liberation of the concentration camps in 1945. In addition to the liberation, the search engines (except Bing in 2020) prioritise content devoted to the post-war commemoration of the Holocaust, primarily by showing images of a few selected memorials (e.g. Auschwitz-Birkenau or the Memorial to the Murdered Jews of Europe). Between 2020 and 2021, the visibility of the memory aspect has increased substantially for Bing and Google while decreasing for Yandex (however, the memory aspect still remains the most visible there).

Besides the liberation and memory aspects common for all three search engines, we observed some distinctions in terms of the Holocaust aspects prioritised by the individual search engines. For instance, Bing was the only search engine that emphasised images showing Jews being arrested and deported by Nazis in 2020. In 2021, the visibility of these two aspects decreased, thus putting Bing more in line with Google and Yandex in terms of these issues, but instead, Bing started putting a large emphasis on images showing the consequences of mass murder. By contrast, Google was the only search engine which prioritised images of the pre-war life of Holocaust victims that reflects current representation practices in many Holocaust museums, which aim to go beyond the exclusive focus on the destruction of the Jewish population in Europe and contextualise the Holocaust for visitors.

Our analysis also demonstrates that a number of the aspects of the Holocaust consistently receive little visibility in search results. In addition to the limited number of images related to the life of victims before and after the Holocaust, there were only a few images showing ghettos and camps (except the liberation of the camps). Similarly, only a few search outputs on Google and Bing showed images of perpetrators, thus rendering the individuals responsible for the Holocaust somewhat invisible. At the same time, compared with 2020, we observed an

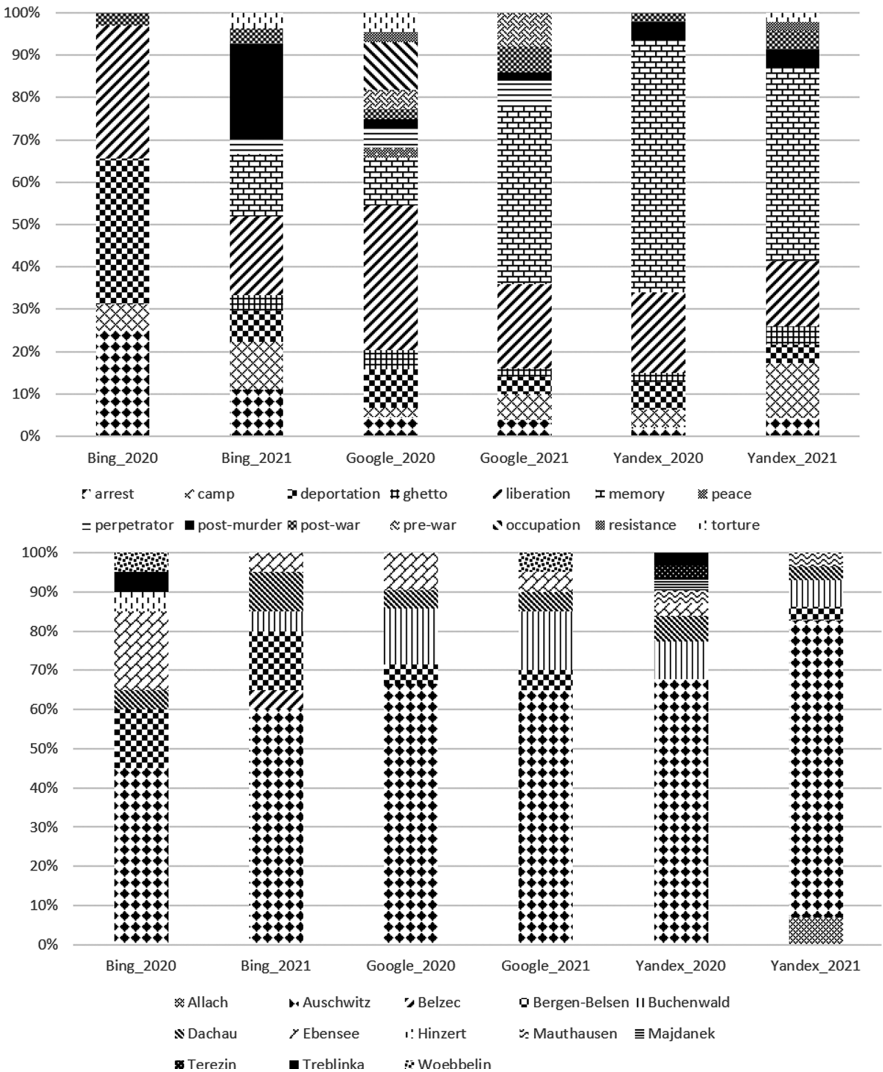

Figure 2: (a) Proportion of search outputs dealing with specific Holocaust aspects. (b) Proportion of search outputs showing content coming from specific Holocaust camps.

increase in the variation of the aspects, particularly for Bing and Yandex, which can indicate that in the long term the diversity of Holocaust representation by search engines can increase.

4.3 Holocaust sites prioritised by web search outputs

In the final part of our analysis, we looked at the distribution of search outputs from specific Holocaust camps. Figure 2b shows that most outputs come from a single site – i.e. Auschwitz-Birkenau. Already in 2020, this specific site was featured in at least 46 percent of outputs for Bing, with Google and Yandex showing an even higher proportion. The visibility of Auschwitz increased even further in 2021, where content from this specific site constituted 55 percent of outputs for Bing and 72 percent and 76 percent for Yandex and Google respectively. Besides Auschwitz, content from a few other camps which appeared more consistently in search outputs included Bergen-Belsen, Buchenwald, Dachau and Ebensee; however, for none of these camps featured more than in 20 percent of outputs.

The high visibility of Auschwitz that becomes even more pronounced over time can be attributed to multiple reasons. Not only is Auschwitz a site of particular historical significance due to the number of victims who died there, but it also occupies a central place in Holocaust-related popular culture.[44] This combination makes Auschwitz a particularly popular touristic destination, resulting in more user-generated content about the camp and a higher amount of commercial advertisement that can also contribute to the visibility of content coming from the site in web search outputs. Finally, there are also many mediatised visual representations of the Holocaust associated with Auschwitz, which are reiterated via journalistic media (the same reason can also explain the high visibility of Ebensee, which despite being a relatively small camp is highly visible in search outputs due to mediatised images of liberated inmates coming from there).

While there can be many reasons contributing to the consistent prevalence of Auschwitz in image search results, the implications of this 'rich get richer' principle in the context of visual representation of the Holocaust can be questioned from the normative point of view. One of the immediate consequences of the focus on the representation of the Holocaust via the prism of Auschwitz is that content from other camps remains less visible, thus resulting in the limited exposure of search users to other experiences of the Holocaust. What is particularly surprising is the extremely low visibility of other extermination camps, in particular the ones located in the eastern parts of Nazi-occupied Poland, despite these camps playing a key role in the Operation Reinhard and resulting in the massive number of deaths

44 Tim Cole, *Selling the Holocaust: From Auschwitz to Schindler; how history is bought, packaged and sold* (London: Routledge, 2017), 97–98.

(e.g. 925,000[45] in Treblinka 2 and 434,508 in Belzec which disappeared from search outputs after 2020 and at least 167,000 in Sobibor and Chelmno which were absent from the top 50 search results altogether).

5 Conclusions

In this chapter, we aimed to empirically investigate how one specific form of mnemotechnology – web search engines – influences the visual representation of historical mass atrocities. For this aim, we audited the performance of three search engines, Bing, Google and Yandex, in 2020 and 2021, in relation to one of the most well-known instances of mass atrocity, the Holocaust. We conducted audits to scrutinise what visual content these search engines prioritise in relation to the Holocaust and how the selection of such content varies over time.

Our analysis shows that the selection of content web search engines prioritised concerning the Holocaust in 2020 and 2021 was quite different. While independent of the search engine, most of such content is relevant to the Holocaust in the sense of showing historical evidence of the atrocity or forms of modern imaginary associated with its commemoration, other features of Holocaust representation are subject to change. Specifically, we observe that for Google and Bing, there is a shift towards prioritising more contemporary user-generated content (in contrast to less emphasis on historical evidence), particularly content from contemporary memory sites (with a particular focus on Auschwitz-Birkenau). Similarly, we found changes in the visibility of specific aspects of the Holocaust: for instance, if images showing the consequences of mass murders were absent on Bing in 2020, in 2021, content showing this specific aspect of the Holocaust constituted one of the largest categories of outputs for this search engine.

The major implication of these observations is that the way mnemotechnologies, such as search engines, organise information about the Holocaust and represent it is subjected to epistemic instability. While the degree of such instability varies between the search engines we examined, with Bing outputs being subject to more change than those of Google and Yandex, it stresses that mnemotechnologies make the representation of the past dynamic rather than static, despite following a set of logical principles which can be expected to result in the reiteration of

45 This and the following assessments of the victims' counts are based on the USHMM data. See USHMM, "Documenting numbers of victims of the Holocaust and Nazi persecution," https://encyclopedia.ushmm.org/content/en/article/documenting-numbers-of-victims-of-the-holocaust-and-nazi-persecution.

the same representation patterns over time. However, the increasing autonomy (as well as stochasticity) embedded in mnemotechnologies results in them keeping memories unstable by design and raises new possibilities and concerns for individual and collective remembrance, in particular in the case of mass atrocities.

In addition to the instability, our analysis shows that the way mnemotechnologies represent the Holocaust results in the unequal visibility of specific aspects of the Holocaust and Holocaust sites. Our findings demonstrate that the representation of the Holocaust is dominated by two or three key aspects largely represented by content from a single Holocaust site, Auschwitz. Under these conditions, many other aspects of the Holocaust, in particular the ones concerning Jewish life in the aftermath of the Holocaust and before it, as well as camp and ghetto experiences (besides the liberation of the camps), remain under-represented. The same under-representation concerns Holocaust sites, particularly some of the key extermination camps in which hundreds of thousands of victims perished (e.g. Sobibor, Belzec, and Treblinka).

The unequal and unstable representation of different aspects of the Holocaust raises the question of the baseline for representing mass atrocities by different forms of mnemotechnologies. Such a baseline is essential both for empirical evaluations of the performance of mnemotechnologies and for identifying the desired performance of these technologies in the context of atrocity remembrance as well as possibilities/risks posed by them in fulfilling ethical and legal obligations in relation to the traumatic past. For instance, is the prevalence of content related to the suffering in Auschwitz in response to general prompts about the Holocaust a form of desired or non-desired performance of a mnemotechnology?

Similar to the case of other forms of technology dealing with the organisation and retrieval of information,[46] the answer to this question depends on the normative expectations about the mnemotechnology's performance. If web search engines are expected to inform their users about the diverse nature of Holocaust remembrance, then the disproportionate focus on Auschwitz at the cost of other less known/visited memory sites, including some of the key extermination camps involved in Operation Reinhard, is likely to raise concerns. By contrast, if the intended aim of the search engine is to guide its users towards information about a few sites/episodes which the majority of search users are likely to associate with

46 For the discussion of the normative expectations concerning another form of information retrieval technology – i.e. news recommender systems – see Natali Helberger, "On the democratic role of news recommenders," in *Algorithms, Automation, and News. New Directions in the Study of Computation and Journalism*, ed. Neil Thurman, Seth C. Lewis and Jessica Kunert (London: Routledge, 2021), 14–33.

the Holocaust, then the performance we observed is to be expected. However, without identifying what performance of mnemotechnologies is desired and what its long-term implications for individual and collective remembrance will be, it is impossible to decide if the current performance is to be improved (and if yes, then how exactly). In the case of web search, the importance of such decisions is amplified by their performance having direct implications for the next generation of mnemotechnologies related to the generative AI[47] due to web search being increasingly integrated with generative AI and influencing what data generative AI has access to for generating outputs.

Finally, it is important to note several limitations of the conducted study. The first is relying on a single query in a Latin script (i.e. "Holocaust"). With the earlier research stressing substantial differences in web search engine performance in different languages,[48] it is important to evaluate how the prioritisation of Holocaust-related visual content can vary among search queries in different languages. Similarly, it is important to examine how different search queries influence the performance of image search in the context of the Holocaust; ideally, such an examination will be combined with soliciting queries from search users to minimise the potential researcher bias regarding the query selection. Another limitation concerns the focus on a single location – i.e. Frankfurt am Main region – that also may have implications for the selection of content the search engines prioritise. Future studies will benefit from comparing image search outputs from different locations to detect how generalisable or Germany-specific the patterns observed in the current study are.

Bibliography

Bandy, Jack. "Problematic machine behavior: A systematic literature review of algorithm audits." In *Proceedings of the ACM on human-computer interaction* 5: 1–34. 2021.

Bleiker, Roland. "Mapping visual global politics." In *Visual Global Politics*, edited by Roland Bleiker, 1–29. New York: Routledge, 2018.

Cole, Tim. *Selling the Holocaust: From Auschwitz to Schindler; how history is bought, packaged and sold*. London: Routledge, 2017.

47 For more information on those, see Wulf Kansteiner, "Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially Intelligent?," *History and Theory* 61(4) (2022): 119–133; Mykola Makhortykh, et al., "Shall androids dream of genocides? How generative AI can change the future of memorialization of mass atrocities," *Discover Artificial Intelligence* 3(1) (2023): 1–17.

48 Ov Cristian Norocel and Dirk Lewandowski, "Google, data voids, and the dynamics of the politics of exclusion," *Big Data & Society* 10(1) (2023): 1–14.

Divon, Tom, and Tobias Ebbrecht-Hartmann. "#JewishTikTok: The JewToks' Fight against Antisemitism." *TikTok cultures in the United States*, edited by Trevor Boffone, 47–58. London: Routledge, 2022.

Esposito, Elena. "Algorithmic memory and the right to be forgotten on the web." *Big Data & Society* 4(1) (2017): 1–11.

Haim, Mario, Florian Arendt and Sebastian Scherr. "Abyss or shelter? On the relevance of web search engines' search results when people google for suicide." *Health Communication* 32(2) (2017): 253–258.

Hannak, Aniko, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. "Measuring personalization of web search." In *Proceedings of the 22nd international conference on World Wide Web*, 527–538. 2013.

Heard, Jefferson, Jordan Wilberding, Gideon Frieder, Ophir Frieder, David Grossman and Larry Kane. "On mediated search of the United States holocaust memorial museum data." In *Next Generation Information Technologies and Systems. 7th International Conference, NGITS 2009 Haifa, Israel, June 16–18, 2009 Revised Selected Papers*, edited by Yishai A. Feldman, Donald Kraft and Tsvi Kuflik, 38–46. Heidelberg: Springer, 2009.

Helberger, Natali. "On the democratic role of news recommenders." In *Algorithms, Automation, and News. New Directions in the Study of Computation and Journalism*, edited by Neil Thurman, Seth C. Lewis and Jessica Kunert, 14–33. London: Routledge, 2021.

Hoskins, Andrew. "7/7 and connective memory: Interactional trajectories of remembering in post-scarcity culture." *Memory Studies* 4(3) (2011): 269–280.

Hoskins, Andrew, and Huw Halstead. "The new grey of memory: Andrew Hoskins in conversation with Huw Halstead." *Memory Studies* 14(3) (2021): 675–685.

Farmer, Sarah. "Going Visual: Holocaust Representation and Historical Method." *The American Historical Review* 115(1) (2010): 115–122.

Jacobsen, Benjamin N. "Sculpting digital voids: The politics of forgetting on Facebook." *Convergence* 27(2) (2021): 357–370.

Kansteiner, Wulf. "Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially Intelligent?." *History and Theory* 61(4) (2022): 119–133.

Kravets, Daria, Anna Ryzhova, Florian Toepfl and Arista Beseler. "Different platforms, different plots? The Kremlin-controlled search engine Yandex as a resource for Russia's informational influence in Belarus during the COVID-19 pandemic." *Journalism* (2023). https://doi.org/10.1177/14648849231157845.

Krueckeberg, Jennifer. "Youth and Algorithmic Memory: Co-producing Personal Memory on Instagram." In *Culture and Computing. Design Thinking and Cultural Computing. 9th International Conference, C&C 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II*, edited by Matthias Rauterberg, 253–264. Cham: Springer, 2021.

Kulshrestha, Juhi, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi and Karrie Karahalios. "Search bias quantification: investigating political bias in social media and web search." *Information Retrieval Journal* 22 (2019): 188–227.

Lessard, Bruno. "Archiving the gaze: Relation-images, adaptation, and digital mnemotechnologies." In *Save as . . . Digital memories*, edited by Joanne Garde-Hansen, Andrew Hoskins, and Anna Reading, 115–128. London: Palgrave Macmillan UK, 2009.

Lewandowski, Dirk. *Understanding Search Engines*. Berlin: Springer Nature, 2023.

Liebermann, Yvonne. "Born digital: The Black lives matter movement and memory after the digital turn." *Memory studies* 14(4) (2021): 713–732.

Makhortykh, Mykola, and Maryna Sydorova. “Animating the subjugated past: digital greeting cards as a form of counter-memory.” *Visual Communication* 21(1) (2022): 28–52.

Makhortykh, Mykola, Aleksandra Urman, and Roberto Ulloa. “Hey, Google, is it what the Holocaust looked like? Auditing algorithmic curation of visual historical content on Web search engines.” *First Monday* 26(10) (2021): 1–24.

Makhortykh, Mykola, Aleksandra Urman, and Roberto Ulloa. “Memory, counter-memory and denialism: How search engines circulate information about the Holodomor-related memory wars.” *Memory Studies* 15(6) (2022): 1330–1345.

Makhortykh, Mykola, Eve M. Zucker, David J. Simon, Daniel Bultmann and Roberto Ulloa. “Shall androids dream of genocides? How generative AI can change the future of memorialization of mass atrocities.” *Discover Artificial Intelligence* 3(1) (2023): 1–17.

Makhortykh, Mykola, Aleksandra Urman, Roberto Ulloa and Juhi Kulshrestha. “Can an algorithm remember the Holocaust? Comparative algorithm audit of Holocaust-related information on search engines.” In *Digital Memory. Neue Perspektiven für die Erinnerungsarbeit*, edited by Iris Groschek and Habbo Knoch, 79–93. Göttingen: Wallstein Verlag, 2023.

Manca, Stefania, Marcello Passarelli, and Martin Rehm. “Exploring tensions in Holocaust museums’ modes of commemoration and interaction on social media.” *Technology in Society* 68 (2022): 1–13.

Nielsen, Rasmus Kleis. “News media, search engines and social networking sites as varieties of online gatekeepers.” In *Rethinking journalism again. Societal role and public relevance in a digital age*, edited by Chris Peters and Marcel Broersma, 93–108. London: Routledge, 2016.

Norocel, Ov Cristian, and Dirk Lewandowski. “Google, data voids, and the dynamics of the politics of exclusion.” *Big Data & Society* 10(1) (2023): 1–14.

Paramita, Monica Lestari, Kalia Orphanou, Evgenia Christoforou, Jahna Otterbacher, and Frank Hopfgartner. “Do you see what I see? Images of the COVID-19 pandemic through the lens of Google.” *Information Processing & Management* 58(5) (2021): 1–16.

Pfanzelter, Eva. “At the crossroads with public history: Mediating the Holocaust on the Internet.” *Holocaust Studies* 21(4) (2015): 250–271.

Pereira, Gabriel. “Apple Memories and Automated Memory-Making: The Networked Image Inside the iPhone Chip.” *Digital Culture & Society* 7(2) (2021): 203–228.

Prey, Robert, and Rik Smit. “From personal to personalized memory: Social media as mnemotechnology.” In *A networked self and birth, life, death*, edited by Zizi Papacharissi, 209–223. London: Routledge, 2018.

San Francisco, Matilde Eiroa. “Primary sources for a digital-born history: the Hispanic blogosphere on the Spanish Civil War and Franco’s regime.” *Culture & History Digital Journal* 7(2) (2018): 1–10.

Schultheiß, Sebastian, and Dirk Lewandowski. “Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference.” *Journal of Information Science* 49(3) (2023): 609–623.

Smit, Rik, Ansgard Heinrich, and Marcel Broersma. “Witnessing in the new memory ecology: Memory construction of the Syrian conflict on YouTube.” *New Media & Society* 19(2) (2017): 289–307.

Smolicki, Jacek. “You Press the Button, We Do the Rest. Personal Archiving in Capture Culture.” *Towards a philosophy of digital media*, 77–100. 2018.

Sundin, Olof, Dirk Lewandowski, and Jutta Haider. “Whose relevance? Web search engines as multisided relevance machines.” *Journal of the Association for Information Science and Technology* 73(5) (2022): 637–642.

Stiegler, Bernard. *For a New Critique of Political Economy*. Cambridge: Polity Press, 2010.

Thomas, Neal. “Social computing as a platform for memory.” *Culture Machine* 14 (2013): 1–16.

Trezise, Thomas. “Unspeakable.” *The Yale Journal of Criticism* 14(1) (2001): 39–66.
Ulloa, Roberto, Mykola Makhortykh, and Aleksandra Urman. “Scaling up search engine audits: practical insights for algorithm auditing.” *Journal of information science* (2022). https://doi.org/10.1177/01655515221093029.
Urman, Aleksandra, and Mykola Makhortykh. “You are how (and where) you search? Comparative analysis of web search behavior using web tracking data.” *Journal of Computational Social Science* (2023): 1–16.
Yates, Frances A. *The art of memory*. London: Routledge and Kegan Paul, 1966.
Walden, Victoria Grace. “What is ‘virtual Holocaust memory?.” *Memory Studies* 15(4) (2022): 621–633.
Zavadski, Andrei and Florian Toepfl. “Querying the Internet as a mnemonic practice: how search engines mediate four types of past events in Russia.” *Media, Culture & Society* 41(1) (2019): 21–37.
Zelizer, Barbie. *Remembering to forget: Holocaust memory through the camera’s eye*. Chicago: University of Chicago Press, 1998.