

This is a post-print version of the blogpost submitted to Digital Holocaust Memory project and accessible via the following link:

<https://reframe.sussex.ac.uk/digitalholocaustmemory/2021/02/04/algorithmic-auditing-the-holocaust-and-search-engine-bias/>

To cite this piece:

Makhortykh, M. (2021, February 4). Algorithmic Auditing, the Holocaust, and Search Engine Bias. *Digital Holocaust Memory*.

<https://reframe.sussex.ac.uk/digitalholocaustmemory/2021/02/04/algorithmic-auditing-the-holocaust-and-search-engine-bias/>

Algorithmic Auditing, the Holocaust, and Search Engine Bias

Mykola Makhortykh

The ongoing digitisation of individual and collective remembrance results in the growing role of algorithmic curation of information about the past by web search engines. This process, however, raises multiple concerns, in particular considering that performance of complex algorithmic systems, including the ones dealing with web search, is often subjected to bias. In this guest post, Mykola Makhortykh shares his experience of studying biases in visual representation of the Holocaust across six major search engines, and discusses its implications for Holocaust remembrance.

Searching for “the Holocaust” Online

Visual images are integral components of individual and collective remembrance. In the case of the Holocaust, photo and video records of atrocities not only serve as evidence of Nazi crimes, but also communicate the traumatic experiences of the past to future generations ([Bathrick, Prager, and Richardson 2008](#)). Unsurprisingly, images of camps, survivors, and perpetrators are common means through which human curators narrate the story of the Holocaust in museums and education centres. However, as interactions with Holocaust memory become digital, the selection, filtering, and ordering of Holocaust-related content is increasingly delegated from humans to algorithms.

Search engines, such as Google, Yahoo, or Baidu, are at the forefront of algorithmic curation of visual information. Trusted by the users to be impartial, search engines increasingly become the preferred source of information both about the present and the past. However, there is a growing recognition that outputs of search engine algorithms can also promote distorted representation of social reality, for instance, by [retrieving only male images for “CEO” query](#) or [prioritizing pornography-related sources for queries related to black women](#). Considering multiple instances of such biases identified in relation to matters of gender and race on search engines raises the question of the implications of algorithmic information curation and its potential biases for representation of past atrocities, such as the Holocaust.

One of the major difficulties of studying bias in algorithmic curation of historical information is its operationalisation. In the field of information retrieval, bias is usually treated as a systematic skewness of outputs towards specific individuals or groups ([Friedman and Nissenbaum 1996](#)). While it is clear how such skewness can be quantified in the case of measurable phenomena, [such as gender distribution for specific occupations and its representation in image search](#), it is less obvious how it can be applied to something that is often subjective, such as remembrance of the past.

While difficult, applying the concept of bias to Holocaust-related content is not impossible. Similar to other historical events, the Holocaust is a composite phenomenon constituted by multiple aspects, such as the pre-war life of European Jews, different forms of Nazi persecution, liberation of the camps, and post-war commemoration. Similar to human curators, who have concentrated on specific themes for representing the Holocaust at certain points in time (e.g. [photos showing liberation of camps in the post-war years or images portraying pre-war life of Jewish communities nowadays](#)), algorithms prioritize images showing certain aspects of the Holocaust. This results in a higher

retrievability of visual content associated with these aspects and, potentially, can diminish public awareness of other aspects which are not prioritized by the algorithms.

Besides retrievability of specific aspects of the Holocaust, the skewness of which can be measured by comparing their appearance in search outputs, the presence (or absence) of specific Holocaust sites can be another form of bias. It is known that the Holocaust took place at multiple places, [including among others more than 44,000 incarceration sites](#). Because of their sheer number, it is barely possible to see visual evidence associated with all of the Holocaust sites presented in search outputs, but it raises the question how algorithms choose which sites to highlight and which to play down. Considering that Holocaust sites also serve as physical embodiments of memory about the victims, it means that search engines are to decide whose suffering the public will be more aware about (e.g. shall Auschwitz victims be highlighted more than those who died in Majdanek?).

To understand whether above-mentioned concerns are applicable to visual representation of the Holocaust by search engines, members of our team (me, [Aleksandra Urman](#) from University of Bern/University of Zurich, and [Roberto Ulloa](#) from GESIS) employ algorithmic auditing, that is a set of techniques used to investigate “the functionality and impact of decision-making algorithms” ([Mittelstadt 2016](#)). Because of the lack of access to search engines’ source code and databases, researchers have to rely on different techniques (e.g. [API querying](#) or [crowd-sourcing](#)) for collecting data about search outputs and detecting whether they are biased. It is also important to account for factors that affect data collection, in particular personalisation, that is the provision of individualized results (e.g. based on search history or location), or randomisation, namely the engines’ tendency to randomly reorder search outputs for maximizing user engagement.

To address these challenges, we designed a cloud-based infrastructure composed of multiple virtual machines on which we deployed 200 virtual agents (also known as robots; for more information on method see [Makhortykh, Urman, and Ulloa 2020](#)). We programmed each robot so they can simulate simple browsing activity (e.g. entering search queries or scrolling the page). Then, on the 26th February, 2020, we asked the robots to simultaneously search for the term “Holocaust” in the six world’s largest search engines: Baidu, Bing, DuckDuckGo, Google, Yandex, and Yahoo, and send us the results. Because of the large number of robots, we were able to counter effects of search randomisation. Their deployment in a controlled environment within the same range of IPs and clean browsers allowed us to minimize the possible effects of search personalisation.

For the analysis, we used 50 images per each search engine which occurred in the top 50 results most frequently. Each image was matched with the online collections of [United States Holocaust Memorial Museum](#). In a few cases, when no direct match was found (or the image was obviously not related to the Holocaust as in the case of exploitation movie posters), we used reverse search in Google and Yandex to locate the source of the image. Specifically, we were interested in three aspects:

- a) whether the image search output is actually related to the Holocaust;
- b) what aspect of the Holocaust it portrays (e.g. deportation, murder, or commemoration);
and
- c) what Holocaust site it portrays.

Our Observations

#1: Not everything that is retrieved for the Holocaust query is actually about the Holocaust. Figure 1 shows that while the majority of search outputs are related to the Holocaust, between 6% to 35% of outputs deal with other subjects. The proportion of non-relevant content is particularly high for Baidu, a Chinese search engine, where the large number of search results deal with [death metal](#) and [exploitation movies](#) (here and below linked materials are actual search results retrieved by our robots). For other search engines, non-relevant content is constituted mostly by historical images related to other atrocities (e.g. [photos showing suppression of Mau Mau uprising in 1950s](#)). One particularly tricky case is [the image of children behind the barbed wire](#) that was prioritized by several engines, but did not actually relate to the Holocaust (instead, it shows Russian children in a Finnish-run transition camp in Petrozavodsk).

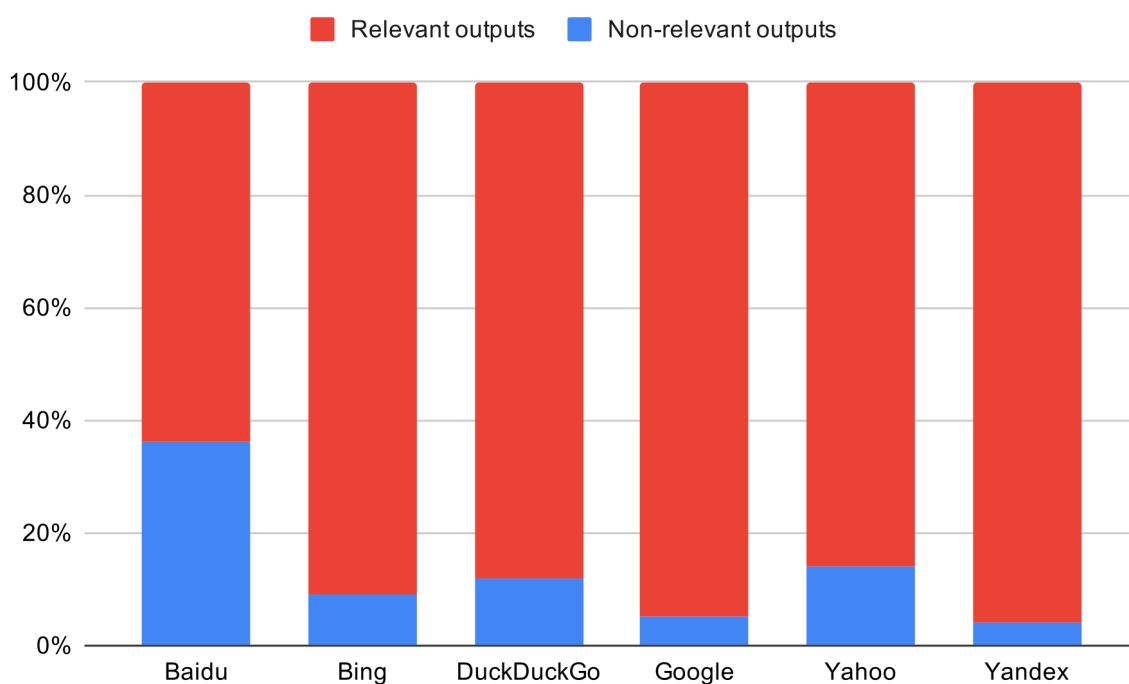


Figure 1. Proportion of Holocaust-relevant and Holocaust-irrelevant outputs per engine

#2: The retrievability of content associated with specific aspects of the Holocaust varies substantially between the search engines. Figure 2 highlights a large gap between Western search engines, which focus on different historical aspects of the Holocaust (e.g. [arrests](#), [deportations](#), and [liberation of the camps](#)) and non-Western search engines, which put major emphasis on commemoration of the Holocaust by focusing on images of monuments and museums, in particular [the Memorial to the Murdered Jews of Europe](#) and [Auschwitz-Birkenau museum](#). There is also substantial variation among Western search engines with some of them putting larger emphasis on images showing arrests and deportations, whereas others highlighting life under occupation. Interestingly, only Google and Baidu returned a few images associated with [pre-war life of Jewish population](#), whereas other engines omitted this specific aspect of the Holocaust.

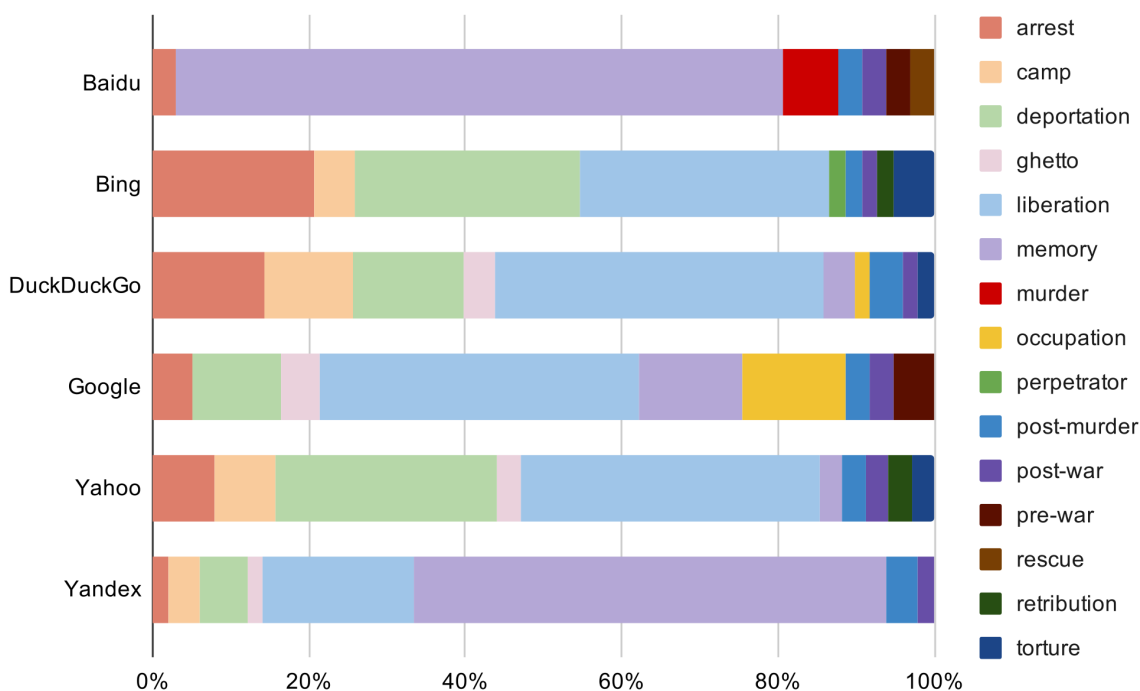


Figure 2. Proportion of Holocaust aspects in search engine outputs

#3: The retrievability of content associated with specific Holocaust sites is skewed towards a few places. Figure 3 shows that for all search engines except Bing more than half of search outputs related to Nazi camps is associated with a single Holocaust site: **Auschwitz**. There is some cross-engine variation for the rest of the outputs with some engines prioritizing images coming from [Bergen-Belsen](#) and [Ebensee](#) (Bing and DuckDuckGo) or [Dachau](#) and [Buchenwald](#) (Yandex and Baidu). However, with the exception of Auschwitz, which is also one of the most attractive destinations for “**dark tourism**”, there are extremely few images associated with major camps located in Eastern Europe. Only half of the engines included a few images from Treblinka, whereas Majdanek appeared only in Yandex outputs. There were no images related to **Sobibor, Chelmno, or Belzec**, despite these three camps being major killing centres, where hundreds of thousands of victims perished. Similarly, only 1-2 outputs across all search engines showed sites related to the so-called “**Holocaust by bullets**” ([Desbois 2008](#)), which was carried out at Eastern European sites not specifically designed for mass murder.

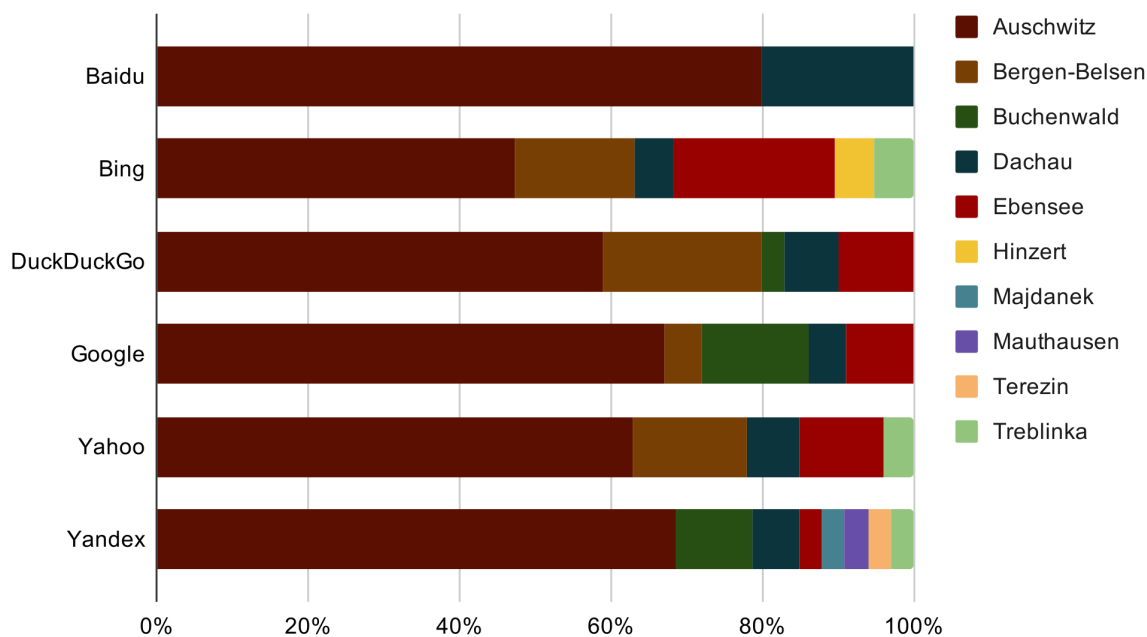


Figure 3. Proportion of specific sites in search engine outputs related to Holocaust camps

Implications of our Findings

First, whilst the proportion of Holocaust-irrelevant outputs for image search is quite low (except for Baidu), their presence is still concerning. Considering that some irrelevant outputs are hard to distinguish from the ones genuinely related to the Holocaust, it raises a possibility of search engines **misinforming their users**. Such possibility is particularly worrisome as image search is often used by individuals who do not necessarily have in-depth knowledge of Holocaust history (e.g. media practitioners), but rely on search outputs to inform the public about specific phenomena.

Second, the **unequal retrievability** of Holocaust-related content between the search engines results in different presentation (and, potentially, different perceptions) of the Holocaust depending on the engine which is used. While cross-engine differences are not unexpected considering the different algorithms and databases used to train them, their presence can amplify **fragmentation of Holocaust remembrance** by reinforcing distinct perspectives on its history and meaning. Particularly in the case of Western and non-Western engines, there is a profound gap between the formers' focus on Holocaust history and the latter's focus on Holocaust remembrance that can lead to different perceptions of the event.

Third, the skewness of outputs towards specific Holocaust themes or sites can be viewed as a form of bias that leads to **unequal** and, possibly, **unfair representation of its aspects**. It is hard to assess whether this bias is attributed to search algorithms or the data they are trained on, but the situation where more than half of outputs is dominated by a single aspect of the event or a single site among 44,000 is worrisome. The possibility of these inequalities being driven by **commercial interests** (e.g. the promotion of the most successful tourist destinations) raises concerns about the logic behind the algorithmic curation of historical information, in particular its ethical dimension. These concerns prompt the need for the more active involvement of memory scholars and curators in the ongoing

debate about **algorithmic fairness and diversity** and raise the question whether the growing interest towards value-driven system design can also accommodate the possibility of **memory-sensitive design** of information retrieval systems dealing with the troubled past.