Auditing YouTube algorithms in relation to Holocaust and COVID misinformation

Keywords: algorithm audit, YouTube, Holocaust, COVID, misinformation

Extended Abstract

With more than 2.5 billion users (Datareportal, 2023), YouTube is one of the most widely used video hosting platforms worldwide. In addition to its extensive user base, YouTube is also distinguished by its intense use of algorithms for curating and promoting content hosted by the platform. These algorithms power a cross-platform search which is used to retrieve videos in response to user queries as well as a content recommendation system which suggests videos based on the user interactions with the platform.

Despite the importance of algorithms for YouTube's functionality, the platform's algorithmic affordances also came under intense criticism, in particular concerning YouTube recommendation algorithms. A number of studies (e.g. Roth et al., 2020; Kirdemir et al., 2021) suggest that YouTube recommendations tend to prioritize a small subset of videos sharing similar characteristics (e.g. viewing time). In some cases, such skewness results in users being nudged towards conspiratorial (Faddoul et al., 2020) or extremist content (Ribeiro et al., 2020), in particular when the user has been watching conspiratorial videos. However, to what degree these observations apply to YouTube search and how the performance of YouTube recommendation and search algorithms in relation to different types of misinformation overtime currently remains unclear. To address this uncertainty, we investigate the following research questions in this work-in-progress submission:

- RQ1: How do YouTube search algorithms deal with the requests for content related to the COVID and Holocaust misinformation?
- RQ2: How engagement with conspiratorial videos influences YouTube recommendation algorithms in the context of COVID and the Holocaust?
- RQ3: How does the performance of YouTube search and recommendation algorithms in relation to COVID- and Holocaust-related misinformation change over time?

To answer these questions, we conduct a virtual agent-based algorithm audit of YouTube search and recommendation algorithms in relation to COVID- and Holocaust-related misinformation in Switzerland. Our selection of these two types of misinformation was attributed to our interest in the capabilities of YouTube algorithms to deal with more established (i.e. the ones related to the Holocaust) and more recent false narratives (i.e. the ones related to COVID). Unlike other audit approaches (see, for the review, Bandy 2021), virtual agent-based audits enable more control over the effects of algorithmic personalization and randomization by simulating human browsing activity in a controlled environment to automatically generate user inputs and record system outputs. To implement the audit, we built a cloud-based network of Linux machines made from scratch and deployed via Google Compute Engine using Zurich area IPs. On each machine, two virtual agents were deployed (one in the Chrome browser and one in the Firefox browser); overall, we deployed 48 agents.

To simulate agent activity, we used Selenium. For search audits, the agents were programmed to open a YouTube search page and enter 28 Germanoptone search queries one by one. Out of

these queries, 14 queries expressed interest in specific misinformation narratives related to COVID and the Holocaust (e.g. whether gas chambers are fake or COVID is harmless) and 14 queries inquired about similar types of information, but without an explicit interest in their conspiratorial aspects (e.g. information about Holocaust concentration camps or health impacts of COVID). In the case of recommendation audits, agents were assigned to four seed videos. Two videos were dealing with misinformation (video 1) and factual information about the Holocaust (video 2) and two videos followed the same principle, but on the topic of COVID. After watching the seed video for 45 seconds, agents shifted to the top recommended video and then repeated the procedure for this video; the process of shifting to the next video continued until agents engaged with 50 recommendations.

To examine how the performance of YouTube algorithms changes over time, we conducted three rounds of audits in May and June 2022. Additionally, we used the concept of search personas - i.e. sequences of browsing actions (Haim et al., 2018) - to investigate how earlier history of visits to websites with different political leanings (e.g. websites of right- and left-leaning Swiss political parties) affects the visibility of misinformation in YouTube search and recommendations. To analyse data, we extracted the top 10 search outputs for each query and the first 20 top recommended videos per agent and coded them to identify the type of YouTube channel from which the result is coming, its stance on disinformation (e.g. debunking or promoting), and the exact type of disinformation which the result is related to.

Currently, we completed coding data and are in the process of analysing the results. Our preliminary findings show little impact of search personas for YouTube search; in the case of recommendation algorithms, however, there is higher variation in outputs which can be possibly attributed to personalisation. Based on the analysis of one round of data collection (Fig. 1), we observe a higher presence of content promoting misinformation for the COVID queries as well as substantial variation in the presence of debunking content per query.

References

Bandy, J. (2021). Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the ACM on human-computer interaction*, *5*(CSCW1), 1-34.

Datareportal. (2023). YouTube Statistics and Trends. https://datareportal.com/essential-youtube-stats

Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube's promotion of conspiracy videos. *arXiv preprint arXiv:2003.03318*.

Haim, M., Graefe, A., & Brosius, H. B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital journalism*, *6*(3), 330-343.

Kirdemir, B., Kready, J., Mead, E., Hussain, M. N., & Agarwal, N. (2021, June). Examining video recommendation bias on YouTube. In *Advances in Bias and Fairness in Information Retrieval: Second International Workshop on Algorithmic Bias in Search and Recommendation, BIAS 2021, Lucca, Italy, April 1, 2021, Proceedings* (pp. 106-116). Cham: Springer International Publishing.

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020, January). Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 131-141).

Roth, C., Mazières, A., & Menezes, T. (2020). Tubes and bubbles topological confinement of YouTube recommendations. *PloS one*, *15*(4), e0231703.

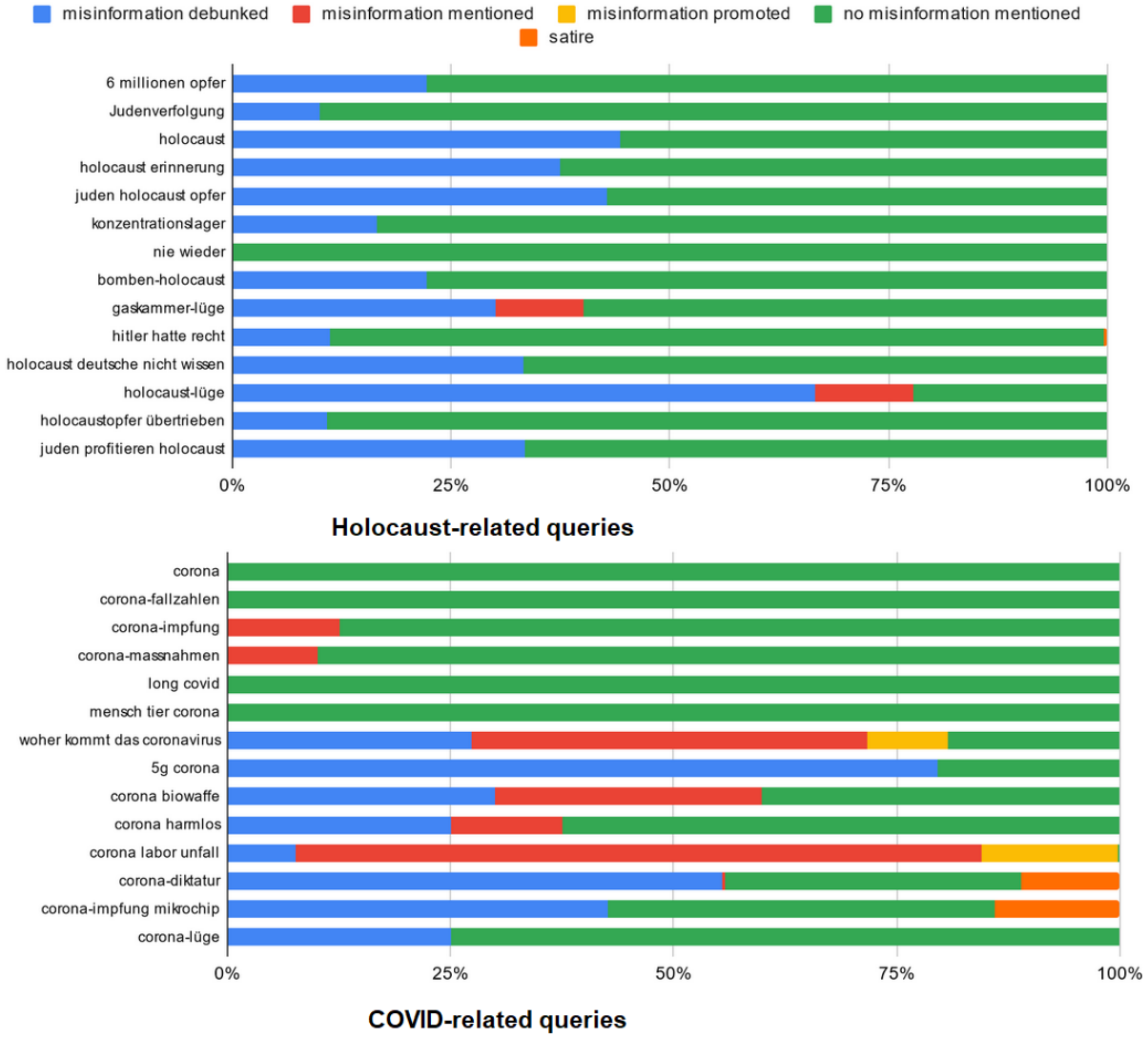

Figure 1. The distribution of YouTube search results for the Holocaust- and COVID-related queries for the data collection on June 25 2022 (by the stance on misinformation)