Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Sister Conferences Best Papers Track

Black-box Audit of YouTube's Video Recommendation: Investigation of Misinformation Filter Bubble Dynamics (Extended Abstract)*

Matus Tomlein[1], **Branislav Pecher**[1,2], **Jakub Simko**[1], **Ivan Srba**[1], **Robert Moro**[1], **Elena Stefancova**[1], **Michal Kompan**[1,3], **Andrea Hrckova**[1], **Juraj Podrouzek**[1] and **Maria Bielikova**[1,3]
[1]Kempelen Institute of Intelligent Technologies
[2]Faculty of Information Technology, Brno University of Technology
[3]Slovak Centre for Research of Artificial Intelligence — slovak.AI
matus.tomlein@kinit.sk, branislav.pecher@kinit.sk, jakub.simko@kinit.sk, ivan.srba@kinit.sk, robert.moro@kinit.sk, elena.stefancova@kinit.sk, michal.kompan@kinit.sk, andrea.hrckova@kinit.sk, juraj.podrouzek@kinit.sk, maria.bielikova@kinit.sk

Abstract

In this paper, we describe a black-box sockpuppeting audit which we carried out to investigate the creation and bursting dynamics of misinformation filter bubbles on YouTube. Pre-programmed agents acting as YouTube users stimulated YouTube's recommender systems: they first watched a series of misinformation promoting videos (bubble creation) and then a series of misinformation debunking videos (bubble bursting). Meanwhile, agents logged videos recommended to them by YouTube. After manually annotating these recommendations, we were able to quantify the portion of misinformative videos among them. The results confirm the creation of filter bubbles (albeit not in all situations) and show that these bubbles can be bursted by watching credible content. Drawing a direct comparison with a previous study, we do not see improvements in overall quantities of misinformation recommended.

1 Introduction

In this extended abstract we describe the most important findings of an *auditing study*[1] originally published in [Tomlein *et al.*, 2021], discuss their impact and outline next steps towards *independent oversight of personalization behavior of large platforms*, which is a general motivation of our work. Large platforms are routinely being accused of contributing to the spread of misinformation caused by their personalized routines and, at the same time, of being reluctant to revise them [Zuboff, 2019; Vaidhyanathan, 2018]. Even when they promise some changes, there is a lack of effective public oversight that could quantitatively evaluate their fulfillment. Auditing studies, such as ours, are tools that may improve such oversight.

In the study, we investigate the *misinformation filter bubble* creation and bursting on YouTube by simulating user behavior on the platform, recording its responses (e.g., search results, recommendations) and manually annotating them for the presence of misinformative content. Then, we quantify the dynamics of misinformation filter bubble creation and also of bubble bursting, which is the main novelty of the study. While previous works examined how a user can enter a filter bubble [Abul-Fottouh *et al.*, 2020; Spinelli and Crovella, 2020; Hussein *et al.*, 2020; Papadamou *et al.*, 2020], no audits have covered *if*, *how* or with what *effort* can the user *burst* the bubble. Thus, our study extends the previous works with investigation of this important aspect.

The first contribution of this work is the analysis of YouTube's personalization behavior in a situation when a user with a history of watching misinformation promoting content and with a developed misinformation filter bubble starts to watch misinformation debunking content in an attempt to burst that misinformation filter bubble. The key finding is that watching credible content generally improves the situation, albeit with varying effects and forms, depending on a particular misinformation topic.

We align our methodology with the work [Hussein *et al.*, 2020] which also investigated the creation of misinformation filter bubbles using sockpuppeting auditing technique. We replicate parts of Hussein's study by re-using the maximum of Hussein's seed data (topics, queries, videos), using similar scenarios and the same data annotation scheme. Therefore, we are able to directly compare the outcomes of both studies. Although we expected to see less filter bubble creation behavior than Hussein et al. due to recent changes in YouTube policies [YouTube, 2020], this was generally not the case.

As the second contribution, we report changes in misinformation video occurrences on YouTube, which took place since mid-2019 when the study [Hussein *et al.*, 2020] was carried out. We observe worse situation regarding the topics of vaccination and (partially) 9/11 conspiracies and some improvements (less misinformation) for moon landing or chemtrails conspiracies.

*This is an extended abstract of a paper that won Best paper award at Fifteenth ACM Conference on Recommender Systems (RecSys '21).

[1]The implementation of the experimental infrastructure and data collected are available at https://github.com/kinit-sk/yaudit-recsys-2021

2 Background and Related Work

Misinformation filter bubbles can be defined as states of intellectual isolation in false beliefs or a manipulated perceptions of reality. They can be characterized by a high homogeneity of recommendations/search results that share the same positive stance towards misinformation. The existing studies confirmed the effects of filter bubbles in YouTube recommendations and search results. Spinelli et al. [2020] found that chains of recommendations lead away from reliable sources towards extreme and unscientific viewpoints. Similarly, Ribeiro et al. [2020] concluded that YouTube's recommendation contributes to further radicalization of users. Abul-Fottouh et al. [2020] confirmed a homophily effect; anti-vaccine videos were more likely to be recommended for anti-vaccine videos than the pro-vaccine ones and vice versa. An *algorithmic audit* is a systematic quantitative probing of an online platform, used for quantification of this proportion [Sandvig *et al.*, 2014; Hussein *et al.*, 2020].

Crowdsourcing audit studies are conducted using real user data. Silva et al. [2020] developed a browser extension to collect personalized ads with real users on Facebook. Hannak et al. [2013] recruited Mechanical Turk users to run search queries and collected their personalized results. While crowdsourcing audits cover more realistic user conditions, this also means they are noisy (e.g., user behavior is influenced by confirmation bias). Additionally, uncontrolled environment makes comparison difficult, it is challenging to retain users, and the audits also have to tackle possible privacy issues.

Sockpuppeting audits solve these problems by employing non-human bots simulating user behavior in a predefined controlled way [Sandvig *et al.*, 2014]. They, however, have their own methodological challenges [Hussein *et al.*, 2020]. First, appropriate seed data such as the initial activity of bots or search queries must be selected. Second, the experimental setup must measure the real influence of the investigated phenomena while minimizing confounding factors and noise (e.g., of name, gender or geolocation). Another challenge presents labelling of the collected data for the presence of the audited phenomena, which can be expert-based/crowdsourced [Hussein *et al.*, 2020; Silva *et al.*, 2020] or automatic [Papadamou *et al.*, 2020].

Existing auditing studies can be further distinguished by the domain they are applied on (e.g., social media [Silva *et al.*, 2020; Papadamou *et al.*, 2020; Hussein *et al.*, 2020], search engines [Metaxa *et al.*, 2019; Le *et al.*, 2019; Robertson *et al.*, 2018], e-commerce sites [Juneja and Mitra, 2021]), by adaptive systems they investigate (e.g., recommendations [Hussein *et al.*, 2020; Spinelli and Crovella, 2020; Papadamou *et al.*, 2020], search results [Papadamou *et al.*, 2020; Hussein *et al.*, 2020; Le *et al.*, 2019; Metaxa *et al.*, 2019; Robertson *et al.*, 2018], autocomplete [Robertson *et al.*, 2018]) and by phenomena they study (e.g., misinformation [Hussein *et al.*, 2020; Papadamou *et al.*, 2020], political bias [Le *et al.*, 2019; Metaxa *et al.*, 2019], political ads [Silva *et al.*, 2020]). Recently, audits also focused on creation of misinformation filter bubbles [Hussein *et al.*, 2020; Papadamou *et al.*, 2020].

3 Study Design and Methodology

In the study, we let a series of agents (bots) pose as YouTube users by performing predefined sequences of video watches and query searches. They also log the platform's responses: recommended videos and search results. The predefined actions are designed to first *create a misinformation filter bubble* by purposefully watching videos containing (or leaning towards) misinformation. Then, agents try to *burst the filter bubble* by watching misinformation debunking videos. To prevent possible carry-over effects, the agents are idle for some time between their actions. The study is a partial replication of a previous study done by Hussein et al. [2020] (denoted onwards as the *reference study*), which allows us to directly compare quantities of misinformative content encountered by the agents in ours and in the reference study. Our research questions were twofold:

RQ1 (comparison to the reference study). *Has YouTube's personalization behavior changed with regards to misinformative videos since the reference study?*

RQ2 (bubble bursting dynamics). *How does the effect of misinformation filter bubbles change, when debunking videos are watched?*

From the reference study, we reuse two metrics: *SERP-MS* and *normalized score*. Both metrics quantify misinformation prevalence in a given list of items (videos) on the $\langle -1, 1 \rangle$ interval. Videos are annotated as either *promoting* (value 1), *debunking* (value -1), or *neutral* (value 0). *Normalized score* is computed as an average of individual video annotations. Since it does not take order into account, it is suited for shorter lists, such as recommendations in our case. *SERP-MS* captures amount of misinformation and its rank, thus making it suitable for longer, ordered lists (search results in our case). It is computed as $SERP\text{-}MS = \frac{\sum_{r=1}^{n}(x_i * (n-r+1))}{\frac{n*(n+1)}{2}}$, where x_i is annotation value, r search result rank and n number of search results in the list [Hussein *et al.*, 2020].

The agents interact with YouTube following a *scenario* depicted in Figure 1. At the start of a run, the agent fetches its predefined configuration, including the YouTube user account and various controlled variables. It also fetches a topic with which it will work (e.g., "9/11"), including a list of 40 *promoting* videos, 40 *debunking* videos, and 5 search queries (e.g., "9/11 conspiracy"). The agent opens a browser in incognito mode, visits YouTube, logs in using the given user account, and accepts cookies. Then it visits YouTube, saves its homepage content and performs the *search phase*. In the *search phase*, the agent executes each predefined search query on YouTube in random order and saves the search results. To prevent any carry-over effect between the search queries, the agent waits 20 minutes after each query.

The agent then proceeds to *create a filter bubble*. It watches (for up to 30 minutes) promoting videos in random order. Immediately after watching a video, the agent saves video recommendations on that video's page and visits the YouTube homepage, saving video recommendations listed there as well. After every two videos, the agent performs another search phase. After watching all listed promoting videos, the agent follows the same procedure to *burst the filter*

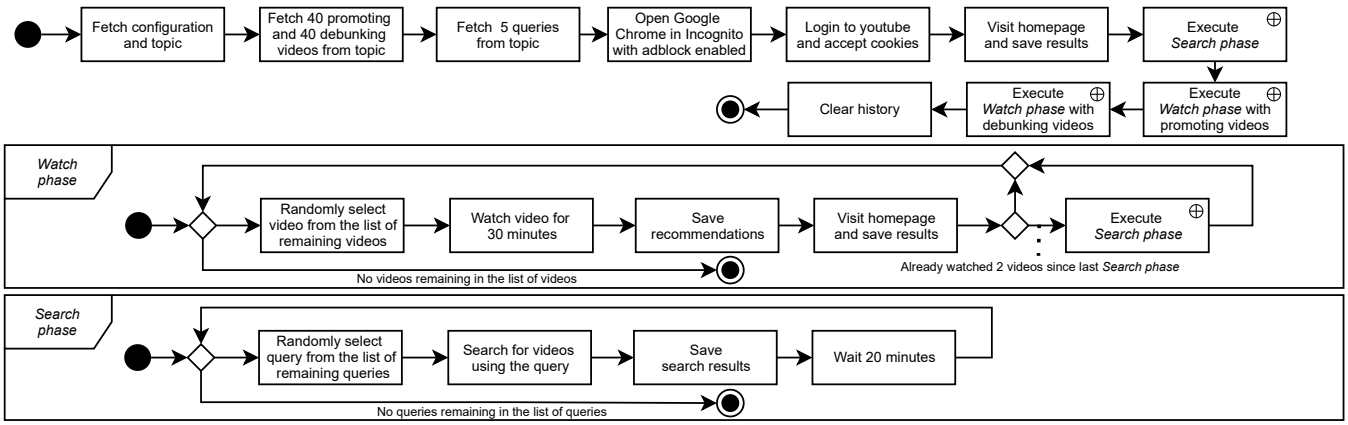

Figure 1: Scenario that agents (bots) followed to create and burst misinformation filter bubbles.

bubble using the debunking video list. After that, the agent clears YouTube history, making the used user account ready for the next run. For each selected topic, we run this scenario 10 times in parallel to reduce noise.

The agents are run from fixed *geolocation* (N. Virginia) to allow for better comparison with the reference study. The date of birth for all accounts is arbitrarily set to 6.6.1990. The gender is set as "rather not say" to prevent any personalization based on this attribute. The account names are composed randomly of the most common US surnames and unisex given names.

We use 5 topics in our study (same as the reference study): *9/11 conspiracies*, *moon landing conspiracies*, *chemtrails conspiracy*, *flat earth conspiracy*, and *vaccines conspiracy*. The narratives associated with the topics are *popular* (persistently discussed), while at the same time, *demonstrably false*, as determined by the reference study. For each topic, the experiment requires two sets of seed videos. As a basis, we use data published in the reference study. Remaining seed videos needed for our experiment (we use more seed videos than the reference study) were collected using similar methodology as in the reference study.

Agents collect videos from three main components on YouTube: 1) *recommendations* appearing next to the watched videos, 2) *home page* videos, and 3) *search results*. We collect 20 videos that YouTube normally displays next to a currently watched video. For each encountered video, the agent collects: 1) *YouTube video ID*, 2) *position* of the video in the list, and 3) *presence of a warning/clarification message* that appears with problematic topics such as COVID-19. Other metadata, such as video *title*, *channel*, or *description*, are collected using the YouTube API.

To annotate the collected videos for the presence of misinformation, we use an extended version of the methodology from the reference study. Each video was viewed and annotated by the authors of this study. The videos are annotated as *debunking*, when their narrative provides arguments against a misinformation, *neutral* when the narrative discusses the related misinformation but does not present a stance towards it, and *promoting*, when the narrative promotes the related misinformation. We also annotate, whether the topic of a video is related to the run's topic (sometimes, misinformation from different topics appeared) or whether it relates to any misinformation topic at all. We also indicate (rare) cases of non-English videos or cases where annotators' confidence about its content is uncertain. Some videos were already removed from YouTube before we managed to annotate them. We also identify videos that mock misinformation (rather than seriously debunking them). For the purpose of this study and given the constraints of manual expert annotation, we annotated only a subset of all collected videos. We annotated all search results as well as videos recommended for first 2 seed videos at the start of the run and last 2 seed videos of both phases (resulting in 6 sets of annotated videos per topic).

4 Results and Findings

We executed the study between March 2nd and March 31st, 2021. Together, we executed 50 bot runs (10 for each topic). The bots watched 3,951 videos (collected 78,763 recommendations associated with them, 8,526 of them unique), executed 10,075 queries (collected 201,404 search results, 942 of them unique), and visited homepage 3,990 times (collected 116,479 videos there, 9,977 of them unique). Overall, we recorded 17,405 unique videos originating from 6,342 channels. Five annotators annotated 2,914 unique videos (covering 255,844 appearances). In total, 244 videos were identified as promoting misinformation, 628 as debunking (including mocking videos), 184 as neutral, 1,829 as not about misinformation. Other videos numbered 29.

We report the results using SERP-MS score metrics for search results and mean normalized scores for recommendations. Since the metrics are not normally distributed with some samples of unequal sizes, we use two-sided Mann-Whitney U test to test our hypotheses. In cases where multiple comparisons by topics are performed, Bonferroni correction is applied on the significance level (in that case $\alpha = 0.05$ is divided by the number of topics (5), resulting in $\alpha = 0.01$).

Within *RQ1*, we investigated *whether YouTube's personalization behavior changed* since the reference study took place. Overall, we observe a small change towards the misinformation promoting content both for the same search queries in our and the reference data as well as for the top-6 recom-

mendations (when comparing the up-next and top-5 recommendations together using last 10 watched promoting videos in reference watch experiments and last two watched videos in our promoting phase). In case of search results, the mean SERP-MS worsened from -0.46 (std 0.42) in reference data to -0.42 (std 0.3) in our data. In case of top-6 recommendations, the mean normalized score worsened from -0.07 (std 0.27) in reference data to -0.04 (std 0.31) in our data. However, neither of these differences is statistically significant.

More considerable shifts in the data can be observed when looking at individual topics. Regarding search, improvement can be seen within certain queries for the "chemtrails" conspiracy that show a large decrease in the number of promoting videos. On the other hand, search results for "flat earth" conspiracy worsened. Within the "anti-vaccination" topic, there is an increase in neutral videos (from 12% to 35%) and a drop in debunking videos (from 85% to 61%). This may relate to new content regarding COVID-19. Regarding recommendations, only the results for the "moon landing" and "anti-vaccination" topics are statistically significantly different. In case of the "moon landing" topic, we see an improvement (more debunking videos are recommended). On the other hand, we observe a drop in debunking videos (from 29% to 9%) in the "anti-vaccination" topic and a subsequent increase in neutral (from 70% to 78%) and promoting videos (from 1% to 8%). There were also more promoting videos on the "9/11" topic (27% instead of 18%), but the distribution is not statistically significantly different.

Within *RQ2*, we investigated *what is the effect of watching debunking videos after the promoting phase*, i.e., whether we will observe the "bubble bursting" behavior. We expected the metrics would improve due to watching debunking videos, i.e., that we would observe misinformation bubble bursting. Regarding top-10 recommendations, their overall distribution is significantly different when comparing the ends of promoting and debunking phases (U=7179.5, p-value=$1.8\mathrm{e}{-9}$). Mean normalized score improves from 0.01 (std 0.31) to -0.27 (std 0.27). Except for the moon landing conspiracy, we observe significantly different distributions for individual topics as well. The 9/11 topic shows a decrease in promoting videos, while other topics show an increase in the number of debunking videos.

We also examined the differences between the very start of the experiment and its end. We expected the metrics would improve due to watching debunking videos despite watching promoting videos before that. The distributions of SERP-MS scores in search results are statistically significantly different (U=36515, p-value=0.0002). Overall, we see an improvement in mean SERP-MS score from -0.39 (std 0.28) to -0.46 (std 0.29). Moreover, all topics except " 9/11" have significantly different distributions and improved. The improvement is due to increases in debunking videos, decreases in promoting videos, or reordered search results in some search queries. Similarly, top-10 recommendations at the end of the experiment come from significantly different distributions (U=6940.5, p-value=$2.9\mathrm{e}{-7}$). Mean normalized score improves from -0.07 (std 0.24) to -0.27 (std 0.27).

5 Conclusions

We performed an audit of misinformation present in search results and recommendations on YouTube. We sought to verify, whether there is less misinformation present in both search results and recommendations after recent changes in YouTube policies [YouTube, 2020]. Unfortunately, we did not find a significantly different amount of misinformation in search results in comparison to the reference study of Hussein et al. [2020]. Only the "anti-vaccination" topic showed a statistically significant difference and that was in a worsening direction. Recommendations showed significant differences across multiple topics but were not significantly different overall. The"moon landing" topic improved normalized scores of recommendation, while the "anti-vaccination" topic worsened its scores. We suspect the changes in search results and recommendations were influenced mostly by changes in content (e.g., COVID-19 pandemic).

We can also conclude that users, even with a watch history of promoting conspiracy theories, do not get enclosed in a misinformation filter bubble *when they search* on YouTube, but they do (with varying degrees depending on the topic) in video recommendations. However, *watching debunking videos helps in practically all cases* to decrease the amount of misinformation that the users see.

Our study had several limitations. Only five topics were selected for investigation (to allow comparison with the reference study) and did not include more recent topics, e.g., "QAnon". Next, we included only a limited set of agent interactions with the platform (search and video watching).

We see several avenues for future work. First, auditing scenarios of a higher fidelity are needed, which would include a more human-like bot behavior such as liking or disliking videos, subscribing to channels, or clicking on the search results or recommendations. This should correspond to behavior of real user stereotypes present on the audited platform which could be learned automatically. Second, considering the cost of manual data annotation, automatic approaches labeling the presence of the audited phenomena should be researched. There have been first attempts, e.g., [Hou *et al.*, 2019; Papadamou *et al.*, 2020], but challenges such as robustness to concept drifts or being able to work with only limited labeled data remain to be addressed. Next, the current auditing studies are usually limited in time and scope. If they are to enable independent oversight of large platforms, they need to become continuous (longitudinal) [Simko *et al.*, 2021] and more generic to allow comparison of the same phenomena across multiple platforms and languages. Lastly, the continuous automatized audits need to assure credibility and allow human oversight. This can be achieved by continuous monitoring of the performance of the automatic ML methods (e.g., the ones used for labeling), by enabling to examine the decisions of these methods via generated explanations and through interpretability of the used models, and by allowing the humans to override or correct these decisions (in line with human-in-command as defined in [AI HLEG, 2019]).

Acknowledgments

This work was partially supported by The Ministry of Education, Science, Research and Sport of the Slovak Republic under the Contract No. 0827/2021, and by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No. 952215.

References

[Abul-Fottouh *et al.*, 2020] Deena Abul-Fottouh, Melodie Yunju Song, and Anatoliy Gruzd. Examining algorithmic biases in youtube's recommendations of vaccine videos. *Int. Journal of Medical Informatics*, 140:104175, 2020.

[AI HLEG, 2019] AI HLEG. Ethics Guidelines for Trustworthy AI. Technical report, European Commission, 2019.

[Hannak *et al.*, 2013] Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. Measuring personalization of web search. In *Proc. of the 22nd International Conference on World Wide Web (WWW '13)*, page 527–538, New York, NY, USA, 2013. ACM.

[Hou *et al.*, 2019] Rui Hou, Veronica Perez-Rosas, Stacy Loeb, and Rada Mihalcea. Towards automatic detection of misinformation in online medical videos. In *2019 International Conference on Multimodal Interaction*, ICMI '19, page 235–243, New York, NY, USA, 2019. Association for Computing Machinery.

[Hussein *et al.*, 2020] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. Measuring misinformation in video search platforms: An audit study on youtube. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW1), May 2020.

[Juneja and Mitra, 2021] Prerna Juneja and Tanushree Mitra. Auditing E-Commerce Platforms for Algorithmically Curated Vaccine Misinformation. In *Proc. of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, 2021.

[Le *et al.*, 2019] Huyen Le, Andrew High, Raven Maragh, Timothy Havens, Brian Ekdale, and Zubair Shafiq. Measuring political personalization of Google news search. In *Proc. of the World Wide Web Conference (WWW '19)*, pages 2957–2963, 2019.

[Metaxa *et al.*, 2019] Danaë Metaxa, Joon Sung Park, James A. Landay, and Jeff Hancock. Search media and elections: A longitudinal investigation of political search results. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

[Papadamou *et al.*, 2020] Kostantinos Papadamou, Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Michael Sirivianos. "it is just a flu": Assessing the effect of watch history on youtube's pseudoscientific video recommendations. In *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM-2022)*, Palo Alto, California, USA, 2020. AAAI Press. [to appear].

[Ribeiro *et al.*, 2020] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. Auditing radicalization pathways on youtube. In *Proc. of the 2020 Conf. on Fairness, Accountability, and Transparency*, page 131–141, New York, NY, USA, 2020. ACM.

[Robertson *et al.*, 2018] Ronald E. Robertson, David Lazer, and Christo Wilson. Auditing the personalization and composition of politically-related search engine results pages. In *Proc. of the 2018 World Wide Web Conf.*, WWW '18, page 955–965, Republic and Canton of Geneva, CHE, 2018. Int. WWW Conferences Steering Committee.

[Sandvig *et al.*, 2014] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, 22:4349–4357, 2014.

[Silva *et al.*, 2020] Márcio Silva, Lucas Santos de Oliveira, Athanasios Andreou, Pedro Olmo Vaz de Melo, Oana Goga, and Fabricio Benevenuto. Facebook ads monitor: An independent auditing system for political ads on facebook. In *Proc. of The Web Conference (WWW '20)*, page 224–234, New York, NY, USA, 2020. ACM.

[Simko *et al.*, 2021] Jakub Simko, Matus Tomlein, Branislav Pecher, Robert Moro, Ivan Srba, Elena Stefancova, Andrea Hrckova, Michal Kompan, Juraj Podrouzek, and Maria Bielikova. Towards continuous automatic audits of social media adaptive behavior and its role in misinformation spreading. In *Adjunct Proc. of the 29th ACM Conf. on User Modeling, Adaptation and Personalization (UMAP '21)*, page 411–414, New York, NY, USA, 2021. ACM.

[Spinelli and Crovella, 2020] Larissa Spinelli and Mark Crovella. How youtube leads privacy-seeking users away from reliable information. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, page 244–251, New York, NY, USA, 2020. ACM.

[Tomlein *et al.*, 2021] Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, and Maria Bielikova. An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes. In *Fifteenth ACM Conference on Recommender Systems*, pages 1–11, New York, NY, USA, September 2021. ACM.

[Vaidhyanathan, 2018] Siva Vaidhyanathan. *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press, 2018.

[YouTube, 2020] YouTube. Managing harmful conspiracy theories on youtube. https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/, 2020. Accessed: 2022-06-03.

[Zuboff, 2019] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books, 2019.