Title: Appendix for "A story of (non)compliance, bias, and conspiracies: How Google and Yandex represented Smart Voting during the 2021 parliamentary elections in Russia"
Authors: Mykola Makhortykh (1), Aleksandra Urman (2), Mariëlle Wijermars (3)
Date: March 7th, 2022
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix

A. Agent-based auditing methodology

Agent-based auditing uses automated agents, namely software simulating human browsing behavior (e.g., page scrolling), to collect information online (Ulloa et al., 2021). Unlike other auditing approaches (for the review, see Bandy, 2021), agent-based auditing allows collecting data under controlled conditions (e.g., by initiating queries at the same time or modeling earlier search history) to control for search personalization (Hannak et al., 2013) and randomization (Makhortykh et al., 2020).

We deployed our agents via the Firefox browser using Selenium WebDriver, a programming interface used for browser testing. The choice of Firefox was motivated by two considerations: first, it was the default browser option within our infrastructure, and it showed more stable performance compared with Google Chrome when running Selenium. Second, we preferred to use a browser that was not affiliated with the company behind the engine audited (i.e., not Google Chrome or Yandex Browser). For every round of search, a new browser instance with clean cookies and history was initialized, thus allowing us to avoid search personalization stemming from the agents' previous behavior. Following initialization, an agent would navigate to Yandex and Google, enter "умное голосование" ("Smart Voting" in Russian) into the search console, collect the first page of results and then close the browser. The first page was collected because existing research (e.g., Schultheiss et al., 2018; Urman & Makhortykh, 2021b) suggests that in the absolute majority of cases people look at and click on results from the first page. The outputs displayed there are, therefore, the most relevant for analyzing information distribution through web search.

In addition to accounting for the effect of the time at which the search was conducted by starting the agent routines at the same times of the day during the whole period of data collection, we also took into consideration location-based search personalization (Kliman-Silver et al., 2015). In order to generate results as if they were seen by the Russian users, we used a commercial VPN provider (RedShield VPN) that provides Russia-based VPN services. We used the Russia-based VPN for all our agents and double-checked using IP-to-location services that our agents were consistently identified as being located in Saint Petersburg during both the testing and the data collection periods.

B. Limitations

The conducted research has several limitations. The use of remote vantage points (e.g., VPNs) has been criticized for being not reliable (e.g., not being located in the advertised countries) (see Weinberg et al., 2018). While we verified that the vantage point was, indeed, located in Saint Petersburg, future studies can benefit from alternative approaches such as recruiting crowdworkers from the respective region to run either the search queries directly or the scripts for powering the agents. An additional benefit of using crowdworkers can be the possibility to examine actual search behavior (e.g., by asking crowdworkers to use not a pre-fixed set of queries but come up with their own search suggestions) and its interactions with search censorship and not just the behavior of engines in response to a query.

Another limitation concerns the implementation of the auditing method used to conduct the study. The current study relies on a single search query and uses agents deployed via a single browser. While we assume that Google and Yandex did not censor other queries related to Smart Voting because of not being explicitly requested to do so, it would be worthwhile to empirically check the validity of our assumption. Similarly, further research can examine whether there are cross-browser differences in retrieved results, which is a phenomenon observed by some earlier studies (e.g., Makhortykh et al., 2020; Urman et al., 2021).

C. Intercoder reliability

To evaluate the intercoder reliability, we compared the results produced by the two coders in the course of the original coding. There were no disagreements between the coders with regard to whether a page was related to Smart Voting as well as the page types. For the last two categories - political bias and conspiratorial information - the coders agreed in 80% (24 disagreements out of 119 hyperlinks) and 78% (26 disagreements out of 119 hyperlinks) of cases respectively. These disagreements were resolved through consensus coding. Finally, we matched the links collected for each of the search engines with their classifications and computed the shares of different types of content distributed by each search engine.