



Research Article

A story of (non)compliance, bias, and conspiracies: How Google and Yandex represented Smart Voting during the 2021 parliamentary elections in Russia

On 3 September 2021, the Russian court forbade Google and Yandex to display search results for “Smart Voting,” the query referring to a tactical voting project by the jailed Russian opposition leader Alexei Navalny. To examine whether the two search engines complied with the court order, we collected top search outputs for the query from Google and Yandex. Our analysis demonstrates the lack of compliance from both engines; however, while Google continued prioritizing outputs related to the opposition’s web resources, Yandex removed links to them and, in some cases, promoted conspiratorial claims aligning with the Russian authorities’ anti-Western narrative.

Authors: Mykola Makhortykh (1), Aleksandra Urman (2), Mariëlle Wijermars (3)

Affiliations: (1) Institute of Communication and Media Studies, University of Bern, Switzerland, (2) Social Computing Group, University of Zurich, Switzerland, (3) Department of Political Science, Maastricht University, Netherlands

How to cite: Makhortykh, M., Urman, A., & Wijermars, M. (2022). A story of (non)compliance, bias, and conspiracies: How Google and Yandex represented Smart Voting during the 2021 parliamentary elections in Russia. *Harvard Kennedy School (HKS) Misinformation Review*, 3(2).

Received: October 21st, 2021. Accepted: January 28th, 2022. Published: March 7th, 2022.

Research questions

- Did Google and Yandex comply with the court order to stop displaying results for the “Smart Voting” query?
- What types of sources did Google and Yandex prioritize in response to the “Smart Voting” query?
- Were Google and Yandex search outputs politically biased, and did they include conspiratorial claims in relation to Smart Voting?

Essay summary

- Using a method that simulates human browsing behavior, we collected the top ten search results for the query “умное голосование” (“Smart Voting” in English) on Google and Yandex,² the two largest

¹ A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.

² Often referred to as “the Russian-language equivalent” (Daucé & Loveluck, 2021, para 2) of Google, Yandex is the largest Russian technical company that provides a multitude of digital services ranging from automated translation to news aggregation to online mapping. Its flagship service is the web search engine of the same name which occupies roughly 50% of the search engine market in Russia (Statscounter, 2021) and maintains a substantial share of the search market in other post-Soviet states.

search engines in Russia, on the eve of the Russian parliamentary elections in September 2021.

- Despite being ordered by the Russian court not to display results for “Smart Voting,” both search engines kept retrieving results for the query. However, Yandex did remove links to the opposition’s web resources dealing with Smart Voting and, in some cases, prioritized conspiratorial claims aligning with the anti-Western narrative of the Russian authorities. This finding corroborates earlier research which suggested that Yandex is susceptible to governmental pressure and highlights the possibility of the corporation’s algorithms enabling a less pervasive form of censorship in favor of the Kremlin.
- Our observations stress the importance of scrutinizing how platforms implement government-driven content moderation, in particular in authoritarian contexts where declared anti-disinformation efforts can facilitate censorship. The presence of political bias in Google and Yandex outputs, together with conspiratorial information in the case of Yandex, also raises concerns about platform-wide filter bubbles and the current state of information freedom in Russia. This stresses the importance of mechanisms to monitor how platforms’ algorithms distribute (politically) contentious content and how this process is affected by state regulation and other forms of political pressure.

Implications

The Smart Voting project was initiated by the currently jailed Russian opposition leader Alexei Navalny and his associates (Team Navalny) in 2018. Its aim is to undermine the political monopoly of the ruling United Russia party by preventing the dispersion of opposition votes across multiple candidates through voter coordination. The effectiveness of this strategy has been demonstrated, for instance, during the 2019 local elections in Saint Petersburg (Golosov & Turchenko, 2021). Via a dedicated website and the mobile app, Smart Voting shows voters which alternative candidate is most likely to defeat the candidate of the ruling party in their polling district. In the lead up to the parliamentary elections (September 17-19, 2021), the Russian authorities sought to limit the effect of Smart Voting: access to the app was restricted by the Russian telecommunications regulator Roskomnadzor on August 23, and the Smart Voting website was blocked on September 7.

The Russian authorities also put pressure on Western and domestic online intermediaries to undermine the effectiveness of Smart Voting. Google and Apple complied with a government request to remove the Smart Voting app from their app stores (Lokot & Wijermars, 2021). At the same time, the Moscow arbitration court requested the two leading search engines in Russia, Yandex and Google, to remove Smart Voting from their search results. The request was presented as an interim measure in a case filed by the Russian company Woolintertrade, which had registered Smart Voting as its own trademark in July. Specifically, the court ruled to “Prohibit [Google/Yandex] the use of the ‘Smart Voting’ designation in search results of a search engine owned by the defendant [Google/Yandex] as one of the search keywords” (Elektronnoe pravosudie, 2021a, p. 1; 2021b, p. 1). How exactly the search engines should have implemented this order was not specified.

The growing pressure on both Google and Yandex ahead of the 2021 elections highlights the increasing importance the Russian government attributes to expanding its influence over online intermediaries, including search engines. While some years ago the Russian authorities aimed to increase their control over online information distribution by (unsuccessfully) trying to create a state-controlled and a state-owned search engine (Sanovich et al., 2018), by now their strategy has shifted towards putting pressure on existing platforms to make them comply with the regime’s demands. In recent years, Russia’s capacity to exert control over online information distribution has been strengthened through increasingly restrictive Internet regulation and enhanced state control over internet infrastructures (Ermoshina & Musiani, 2017; Sivetc, 2021; Stadnik, 2021).

Since Yandex is a Russian domestic corporation, it is generally thought to be more responsive to

government pressure (Daucé & Loveluck, 2021) while Google is seen by many Russian NGOs as a protector of civil liberties (Bronnikova & Zaytseva, 2021). Compared with Google, Yandex was more intensively targeted by Russian regulatory mechanisms (Wijermars, 2021) and demonstrated more politically biased performance during periods of political contention (Kravets & Toepfl, 2021). However, the Russian authorities also put increasing pressure on Google as demonstrated by the growing number of requests for removing results from its search, particularly since the second half of 2020 (Figure 1).

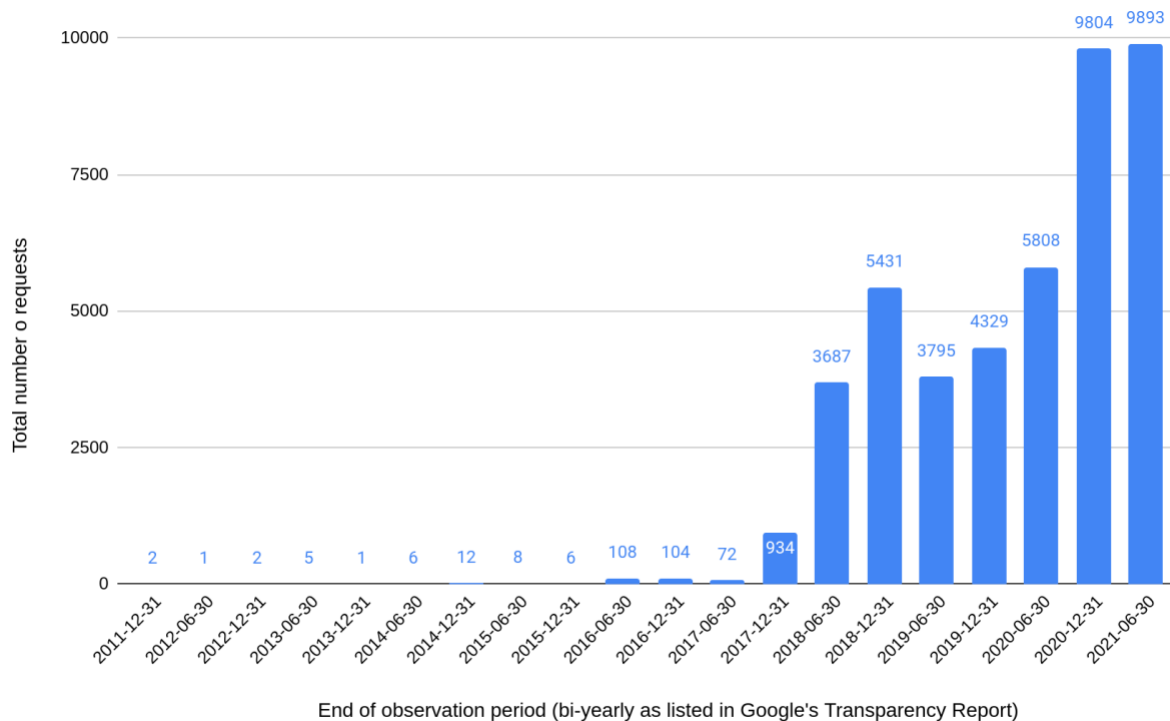


Figure 1. Number of Russian government removal requests for Google (web search). The graph is created using data from Google's Transparency Report (Google, 2021).

Across all Google services, Russia is responsible for the highest number of removal requests (Google, 2021). While more than 90% of these requests are related to copyright issues, it is impossible to estimate what share of copyright requests might be politics-related. Furthermore, Google is increasingly complying with the Russian government's requests (Figure 2). This trend seems to be specific to Russia: in the case of Turkey, for example, Google's compliance rate is around 40%, while for China it is close to 0% (Google, 2021). Given the limited information, it is difficult to ascertain whether the high compliance rate indicates a change in Google's treatment of Russian removal requests (i.e., becoming more likely to comply with them) or a qualitative change in these requests (e.g., a higher proportion of evidently illegal content which is requested to be removed).



Figure 2. Share of content removed by Google in response to the Russian government's requests (all services including web search). The graph is created using data from Google's Transparency Report (Google, 2021). Note that the beginning of 2019 is the earliest period for which the data is available in the report.

Against this backdrop of growing governmental pressure, we looked at how Google and Yandex dealt with the above-mentioned court order and found that both search engines continued to display results for the “Smart Voting” query. However, Yandex did limit access to the opposition’s web resources dealing with Smart Voting, which, presumably, was the main aim of the politically motivated court case. By contrast, Google did not remove Smart Voting from its search results and continued to link to Team Navalny resources. These differences in (non)compliance were amplified by political bias in the outputs, with Google outputs including more pro-Smart Voting content (and vice versa for Yandex), as well as Yandex tending to promote conspiratorial claims about Smart Voting coming from the Russian authorities (e.g., it being a project of the Pentagon aimed at harming the Russian people) and Google usually omitting such claims from its top results.

There are several implications of our observations. First, the example of Yandex illustrates the importance of scrutinizing how compliance with government removal requests as well as government-driven content moderation is implemented. Particularly in the case of “informational autocracies” (Guriev & Treisman, 2020, p. 1) in restricted media systems, such as Russia, where the preservation of the regime is largely contingent on successful censorship, the selective removal of links from search results can be used to facilitate dissemination of misinformation and anti-Western conspiracy theories. It can also serve as a form of “masked censorship” (Makhortykh & Bastian, 2020, p. 14), namely a less intrusive way of filtering out undesired information. The potentially damaging effects extend far beyond the removal of the links themselves as information suppression can erode democratization processes (Stoycheff et al., 2018).

Under these circumstances, the decisions of online intermediaries concerning the implementation of removal requests play a key role in mediating the impact of censorship. Depending on these decisions, intermediaries might limit the visibility of outlets spreading extremist or false claims, but also facilitate state censorship by integrating it with “private censorship” (i.e., censorship effort by private companies) (see Beazer et al., 2021; Crabtree et al., 2015) or amplifying self-censorship practices (e.g., of Russian

journalists in the aftermath of the annexation of Crimea) (Schimpfössl et al., 2020; Zeveleva, 2020). This raises questions about the feasibility of one-size-fits-all regulation of intermediaries' activities that would be applicable to their operations across democratic and non-democratic contexts, especially in the case of countries with long-standing censorship traditions (e.g., Russia and many other post-Communist states) (Ognyanova, 2019).

Second, our observations demonstrate the possibility of profound information inequalities forming between the users of Russia's two largest search engines. While the mere presence of such inequalities is not unexpected, especially in the context of the increasingly fragmented and polarized Eastern European digital ecosystems (Urman, 2019; Urman & Makhortykh, 2021a), the empirical evidence of their amplification by algorithmic mechanisms is concerning. Considering that these inequalities in the case of Smart Voting (and, potentially, other political matters in the region) are subjected to political bias, such amplification may encase Yandex and Google users into platform-wide filter bubbles (Pariser, 2011), in particular as the Russian search market is roughly split between Google and Yandex (Statscounter, 2021).

Third, while one might argue that, in the case of Smart Voting, Google search algorithms enabled a less censored selection of information, future cases may be less clear-cut. In particular, specific features of non-free media systems raise concerns about the universality of source prioritization principles. While both Google and Yandex tend to prioritize journalistic media (Zavadski & Toepfl, 2019), our observations demonstrate that their selection of specific outlets is subjected to political bias with the former prioritizing less state-dependent outlets and the latter giving priority to more Kremlin-oriented media.

It is less clear what Google will do if the number of independent media in Russia continues to decrease, considering the ongoing campaign against journalistic "foreign agents" (Roth, 2021, para 1). Will it shift towards prioritizing the state-sponsored media, following the general policies of the company which are also reflected in its algorithmic systems? Or will these policies be adapted for the case of Russia (and, potentially, other non-free media systems)? To answer these questions, it is important to keep monitoring both how platform algorithms distribute contentious political content and how such distribution is affected by regulation in both Western and non-Western contexts.

Together, our findings provide empirical evidence that may lay ground for a public debate on how and to what degree online intermediaries collaborate with authoritarian governments in enabling censorship. The importance of this debate is amplified by the ongoing transition from traditional violence-based autocracies to informational autocracies. Informational autocracies rely on the regime's ability to convince the public of its competency (Guriev & Treisman, 2020), which is a task increasingly fulfilled with the help of online intermediaries. Under these circumstances, it is essential to define what behavior is ethically desirable for the corporations working in non-democratic contexts. For instance, should Google collaborate with authoritarian states as it seems to increasingly do in Russia? Or should it non-comply as in the case of Smart Voting and risk fines and, potentially, withdraw from the regional market as in the case of China?

Answering these questions can be the first step for building up pressure from the public and policymakers to enforce the desirable behavior from the online information intermediaries. Similarly, it can help determine how intermediaries coming from authoritarian states and, potentially, serving as integral components of the informational autocracy shall be treated. Is there a point, for instance, at which Yandex search engine should be viewed as another information asset of the Kremlin, similar to Russia Today, and approached accordingly?

Findings

Finding 1: Google and Yandex kept displaying results in response to the "Smart Voting" query, despite being ordered to stop doing so.

Despite being ordered to stop displaying search results in response to the “Smart Voting” query, neither Google nor Yandex did so on the eve of the Russian elections. Both search engines kept providing the results throughout the whole period of data collection (September 10–20). This non-compliance can be primarily attributed to the technical implausibility of the initial request. Unlike the requests for Google and Apple to remove the Smart Voting app from their App Stores (Lokot & Wijermars, 2021), which occurred around the same time, the technical realization of non-displaying the results for a particular query is a non-trivial task. While modification of search outputs is not impossible, it usually involves filtering out individual websites from the outputs in general or for specific, but more narrowly defined, queries (e.g., as in the case of the implementation of the right to be forgotten, Mangini et al., 2020).

The imprecision of the request was also noted by the Yandex representatives, who remarked that “it is unclear what exactly we are requested to do [by the court] and how we can do it” (Interfax, 2021, para 2). Google did not provide any commentary on the decision, but it can be presumed that it also found the court decision difficult to implement. The fact that Yandex stopped prioritizing the links to the Team Navalny resources can be viewed as a form of compliance (albeit more with the legal requirement to stop displaying links to blacklisted websites rather than the interim measure from the trademark case).

Finding 2: Google and Yandex prioritized different types of sources.

Contrary to earlier studies which did not observe substantive differences in the selection of sources by Yandex and Google (Zavadski & Toepfl, 2017), we discovered substantial discrepancies in the content prioritized by the two engines, with the exception of journalistic media which dominated the outputs in both cases (Figure 3). Google consistently returned links to the Team Navalny resources related to Smart Voting, including the main Smart Voting website that the Russian government had blocked. Yandex did not return any outputs affiliated with the Team Navalny with the exception of September 13, when a link to the Smart Voting website appeared in the results once. We suggest that this exceptional behavior can be attributed to the output randomization used by search engines to maximize user engagement by testing different ways of ranking search results. Sometimes referred to as “Google Dance” (Battelle, 2005), this phenomenon remains largely under-studied (for some exceptions see Haim et al., 2016; Makhortykh et al., 2020), but it can be one of the factors influencing changes in the visibility of individual resources.

We also observed differences in the prevalence of blogs and social media in search outputs. On Yandex, between 20% and 40% of top links were to blogs, while on Google only a few links to blogs appeared during the selected period. For social media the situation was reversed: Google returned more links to the platforms, such as Facebook and YouTube, than Yandex. There also were differences in the affiliation of blogs and social media accounts. Whereas the blogs, in the case of Yandex, usually belonged to anonymous individuals who often promoted pro-Kremlin narratives, the social media pages prioritized by Google usually belonged to Team Navalny. The share of links to reference websites (e.g., Wikipedia) was similar across the two engines.

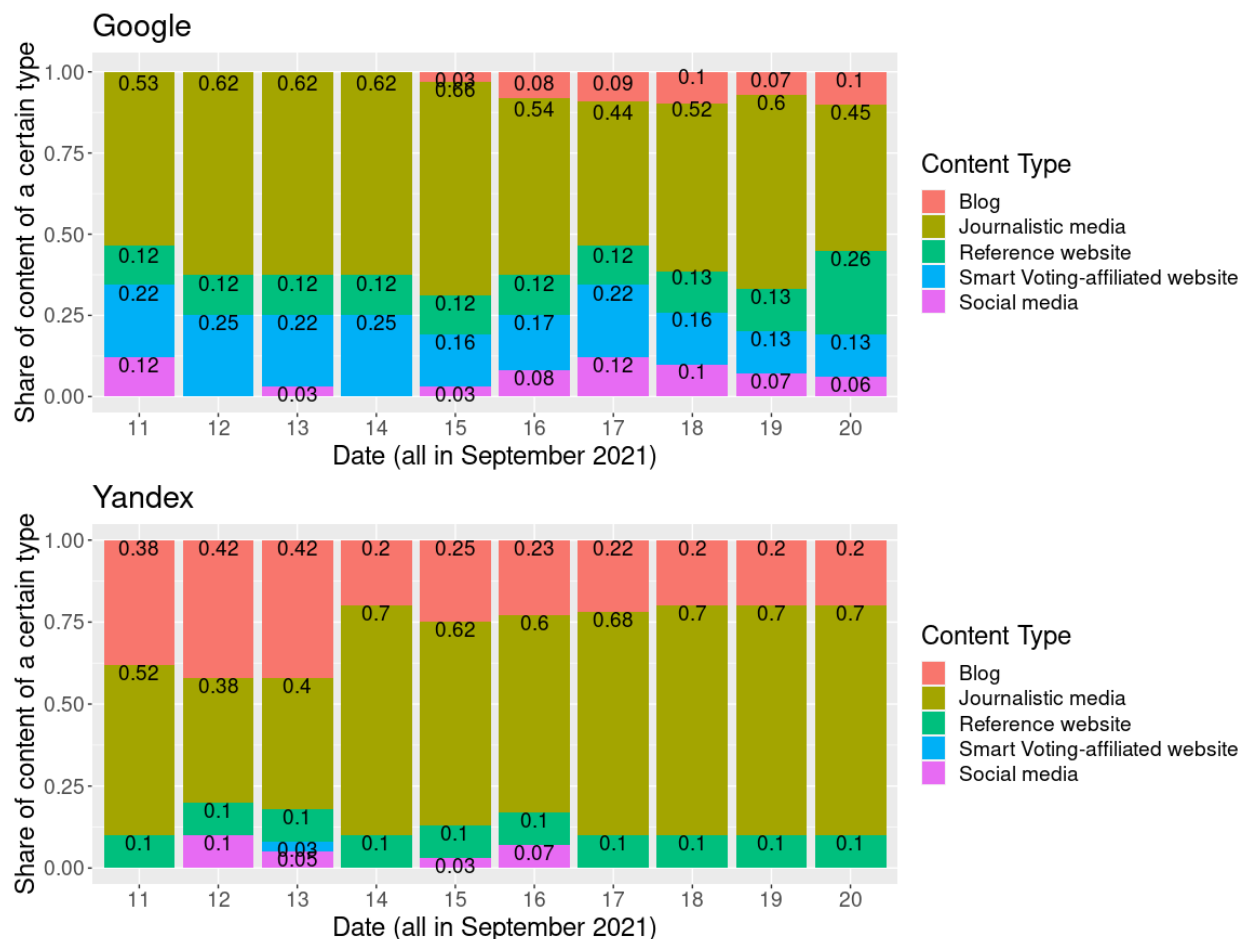


Figure 3. Shares of links to content of different types returned by Google (above) and Yandex (below), by date. The numbers on the bar segments denote the exact share³ of links of a specific type for all rounds of data collection on a particular day. For example, out of all Google search results collected on September 11, 53% linked to journalistic media, 12% to reference websites, 22% to Smart Voting-affiliated websites, and 12% to social media.

Finding 3: Content prioritized by Google and Yandex was politically biased.

We found major differences in the share of content with different stances towards Smart Voting (Figure 4). Both search engines returned a similar share of links to neutral content (hovering around 50% of all links), with Google having a higher share of neutral content at the beginning of the data collection period than Yandex. However, the non-neutral content distribution was very different: in the case of Google, almost all non-neutral content had a positive stance towards Smart Voting, while on Yandex the share of pro-Smart Voting content never went above 30% and was consistently lower than the share of anti-Smart Voting content. Thus, we observe a political bias in the outputs of the two search engines.

³ Here and in the following figures the numbers are given in the range from 0 to 1, where 0 indicates that there were no links of a specific type retrieved on this day, and 1 means that all links retrieved on this day belonged to this specific type.

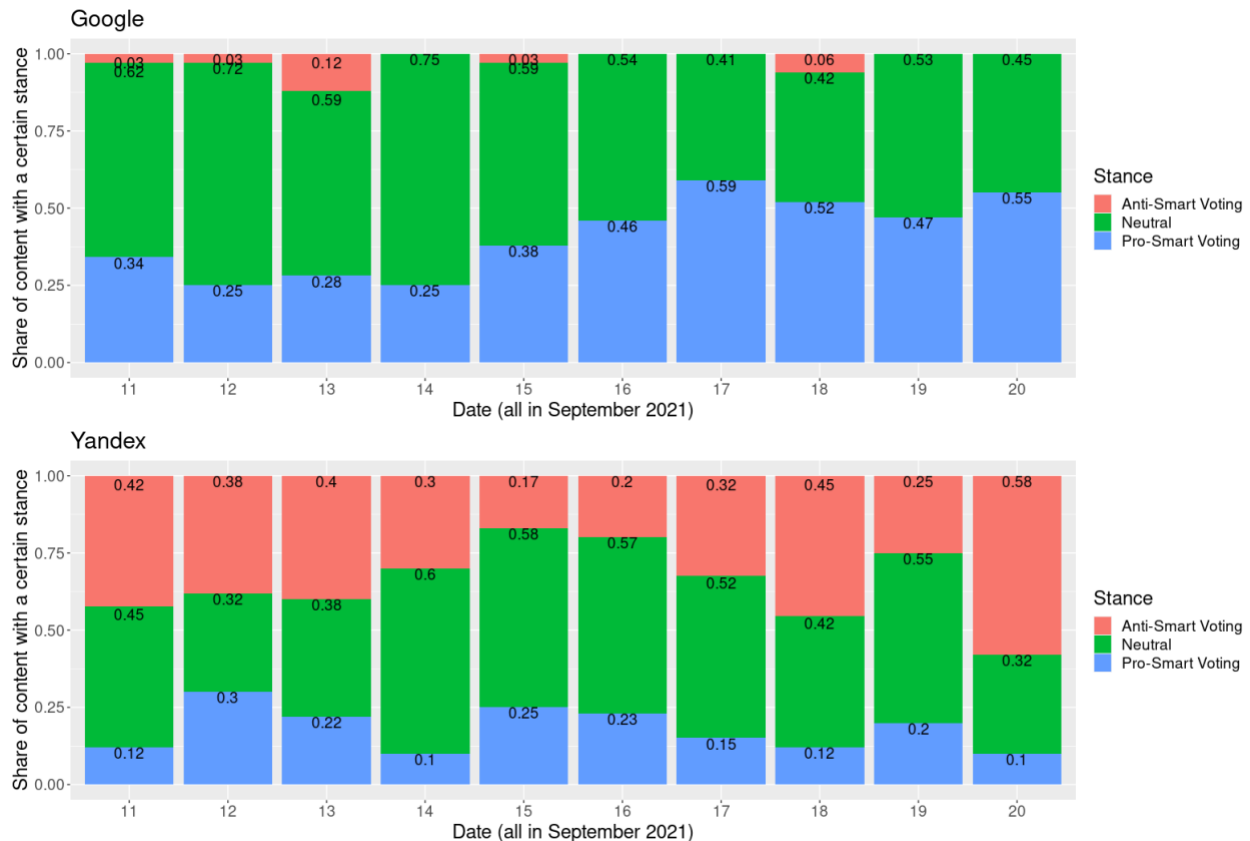


Figure 4. Shares of content with different stance towards Smart Voting returned by Google (above) and Yandex (below), by date. The numbers on the bar segments denote the exact share of links of a specific type for all rounds of data collection on a particular day. For example, out of all Google results collected on September 11, 3% linked to web pages criticizing Smart Voting, 62% to neutral web pages, and 34% to pages supporting Smart Voting.

Among positive and negative assessments of Smart Voting a broad variety of perspectives was found. In the case of anti-Smart Voting content prioritized by Yandex, we observed claims that Smart Voting is a form of manipulation aiming to “deprive the citizens of their right to vote” (Babin, 2020, para 1) or a security threat to personal data (REN TV, 2021a), as well as arguments presenting Smart Voting as ineffective or morally questionable (e.g., because it often recommended voting for the Communist Party) (Novye Izvestiia, 2021). In the case of Google, a positive stance was often expressed in the form of independent media either hosting op-eds that endorsed Smart Voting or linking to the Smart Voting candidate lists.

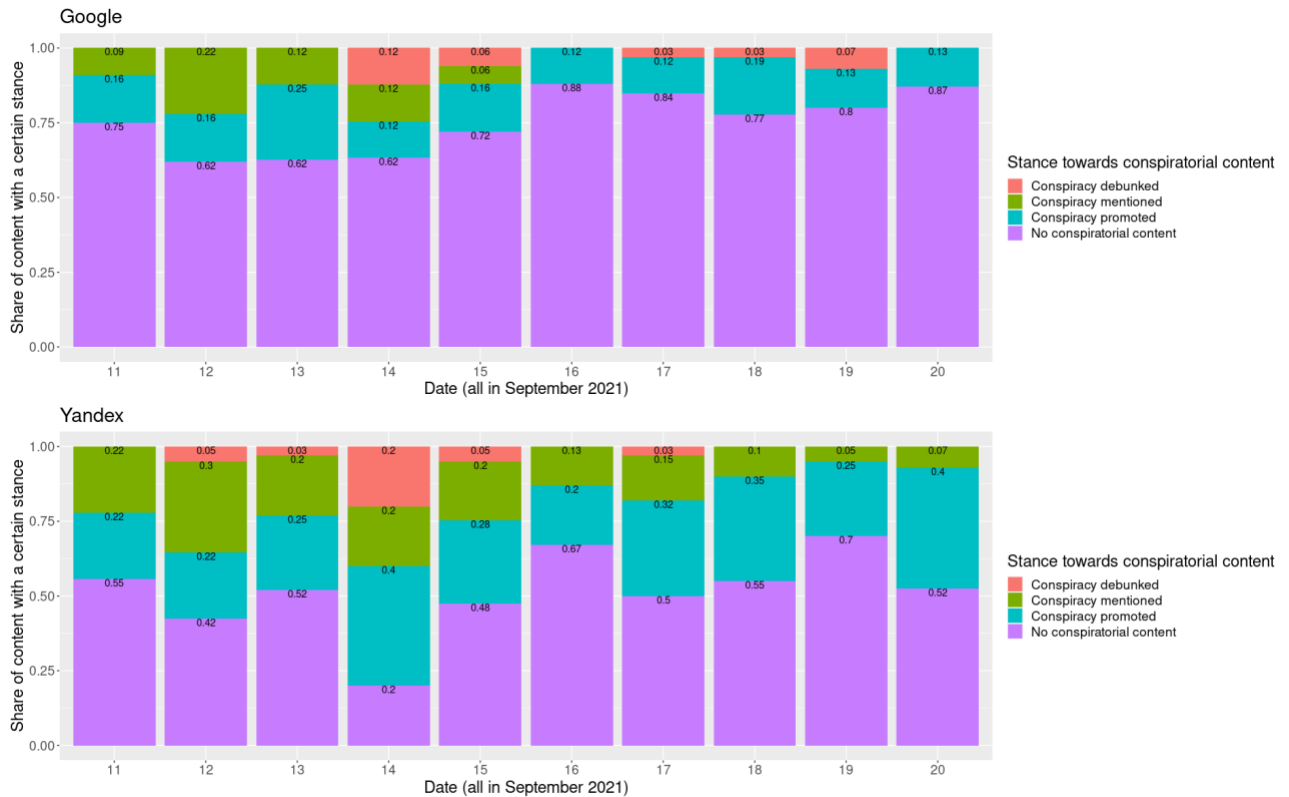


Figure 5. Shares of content with conspiratorial claims about Smart Voting returned by Google (above) and Yandex (below), by date. The numbers on the bar segments denote the exact share of links of a specific type for all rounds of data collection on a particular day. For example, out of all Google results collected on September 11, 75% linked to web pages without mentions of conspiracy theories, 16% linked to pages with conspiracy-promoting content, and 9% linked to pages that mentioned conspiracies without promoting or debunking them.

In some cases, search outputs also included conspiratorial claims related to Smart Voting (Figure 5) such as that it is aiming to harm Russian voters (Lenta, 2021) or is a U.S. Pentagon project designed to interfere in the Russian elections (REN TV, 2021b). Often, these claims originated from Russian officials such as the spokesperson of the Ministry of Foreign Affairs, Maria Zakharova, or the spokesperson of the Russian President, Dmitrii Peskov, that also resulted in them being reiterated by authoritative sources (e.g., established journalistic media). While both Google and Yandex included such claims in their top results, Google tended to prioritize content without mentions of conspiratorial information (e.g., stories about Smart Voting coming from BBC or RBC). By contrast, Yandex gave more visibility to items promoting conspiratorial claims which were usually coming from pro-Kremlin outlets (e.g., TASS or REN TV).

Methods

Data collection

We conducted an agent-based audit of the first page of search results of Yandex and Google for the query “умное голосование” (Smart Voting) (for further details on the chosen auditing approach, which seeks to simulate human browsing behavior, see Appendix, Section A). This specific query was chosen because it is the term to which the court order applies. While users may have also used other search terms, examining the content retrieved for such related (e.g., “SmartVote app”) or derived queries (e.g., “Smart Voting Navalny”) is beyond the scope of the court order and therefore this study.

Data collection started one week before the start of the voting period (September 10) and ended on the last voting day (September 20). To increase the robustness of our results, we aimed to have four rounds of the data collection each day at the same time; however, due to technical issues (e.g., a change in the underlying HTML code of Google's search page) on September 14 we had only one round of data collection, on September 19 two, and on September 16 three rounds instead of four. Notwithstanding this and a few other limitations (see Appendix, Section B), we suggest that our results are robust enough to infer how information about Smart Voting has been distributed by the two engines.

Data analysis

We extracted all hyperlinks gathered from the organic results on the first page of Yandex and Google search results. Altogether, we collected 731 hyperlinks. The odd number is due to search engines sometimes providing fewer than ten organic results on their first page because of the presence of other elements (e.g., "Knowledge Panels" or "People also ask"). Among the 731 results, there were 119 unique hyperlinks. We manually coded each hyperlinked page into the following categories:

- Whether the page contains information related to Smart Voting: a binary variable (yes/no).
- The type of the page: 1) journalistic media, 2) blog, 3) social media, 4) reference website (e.g., an online encyclopedia), or 5) political campaign (e.g., a Smart Voting-affiliated website).

Then, for pages containing information related to Smart Voting we used two additional coding categories:

- Whether the page content is politically biased: 1) arguing in favor of Smart Voting, 2) arguing against Smart Voting, 3) neutral regarding Smart Voting.
- Whether the page contains statements about Smart Voting that can be classified as conspiratorial. For example, statements that present Smart Voting as part of secret agenda pursued by domestic (e.g., the Kremlin) or foreign (e.g., the US) actors to interfere in the elections and, potentially, harm the Russian people, were coded as conspiratorial: 1) the page promotes such statements, 2) the page debunks them, 3) the page mentions such statements without adopting clear stance towards them, or 4) the page does not mention any conspiratorial information.

The coding was performed independently by two coders, both fluent in Russian and knowledgeable of the Russian political context (on intercoder reliability, see Appendix, Section C). 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.

Bibliography

- Babin, V. (2020, August 26). *Net nichego bolee idiotskogo, chem proekt 'Umnoe golosovanie'* [There is nothing dumber than the "Smart Voting project"]. 7x7. <https://7x7-journal.ru/posts/2020/08/26/net-nichego-bolee-idiotskogo-chem-proekt-umnoe-golosovanie>
- Bandy, J. (2021, April). Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the ACM on Human-Computer Interaction*, 5(1), 1–34. <https://doi.org/10.1145/3449148>
- Beazer, Q., Crabtree, C., Fariss, C., & Kern, H. (2021, September 23). When do private actors engage in censorship? Evidence from a correspondence experiment with Russian private media firms. *British Journal of Political Science*, 1–20. <https://doi.org/10.1017/S0007123421000351>

- Bronnikova, O., & Zaytseva, A. (2021, May 3). 'In Google we trust'? The Internet giant as a subject of contention and appropriation for the Russian state and civil society. *First Monday*, 26(5). <https://dx.doi.org/10.5210/fm.v26i5.11709>
- Crabtree, C., Fariss, C., & Kern, H. (2015, December 26). *Truth replaced by silence: A field experiment on private censorship in Russia*. SSRN. <https://dx.doi.org/10.2139/ssrn.2708274>
- Daucé, F., & Loveluck, B. (2021, April 22). Codes of conduct for algorithmic news recommendation: The Yandex.News controversy in Russia. *First Monday*, 26(5). <https://dx.doi.org/10.5210/fm.v26i5.11708>
- Elektronnoe pravosudie. (2021a, September 3). *Case A40-186508/21-51-1254*. https://kad.arbitr.ru/Document/Pdf/9d3235f9-445c-4ef6-9316-bce63c190854/b89d4010-c29d-4c1d-9908-ab68588c4675/A40-186508-2021_20210903_Opredelenie.pdf?isAddStamp=True
- Elektronnoe pravosudie. (2021b, September 3). *Case A40-186534/21-27-1259*. https://kad.arbitr.ru/Document/Pdf/8b1fb4a9-ebf5-4017-a8ff-bdb7bb864475/c5558759-14eb-4ff2-943a-3654734f80d4/A40-186534-2021_20210903_Opredelenie.pdf?isAddStamp=True
- Ermoshina, K., & Musiani, F. (2017, March 22). Migrating servers, elusive users: Reconfigurations of the Russian Internet in the post-Snowden era. *Media and Communication*, 5(1), 42–53. <https://doi.org/10.17645/mac.v5i1.816>
- Golosov, G. V., & Turchenko, M. (2021, June 2). Countering the “sweep effect” opposition voter coordination versus electoral system effects in authoritarian local elections. *Communist and Post-Communist Studies*, 54(1–2), 66–82. <https://doi.org/10.1525/j.postcomstud.2021.54.1-2.66>
- Google. (2021). *Government requests to remove content – Google transparency report*. <https://transparencyreport.google.com/government-removals/overview?hl=en>
- Guriev, S., & Treisman, D. (2020, June). A theory of informational autocracy. *Journal of Public Economics*, 186, 1–11. <https://doi.org/10.1016/j.jpubeco.2020.104158>
- Haim, M., Arendt, F., & Scherr, S. (2016, May 19). Abyss or shelter? On the relevance of web search engines' search results when people google for suicide. *Health Communication*, 32(2), 253–258. <https://doi.org/10.1080/10410236.2015.1113484>
- Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013, May). Measuring personalization of web search. In *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)* (pp. 527–538). Association for Computing Machinery. <https://doi.org/10.1145/2488388.2488435>
- Interfax (2020, September 6). 'landex' obzhaluet zapret sudom poiskovoi vydachi 'Umnogo golosovaniia' [Yandex will challenge the court order to forbid search outputs for “Smart Voting”]. <https://www.interfax.ru/russia/788804>
- Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., & Mislove, A. (2015, October). Location, location, location: The impact of geolocation on web search personalization. In *IMC '15: Proceedings of the 2015 Internet Measurement Conference* (pp. 121–127). Association for Computing Machinery. <https://doi.org/10.1145/2815675.2815714>
- Kravets, D., & Toepfl, F. (2021, June 7). Gauging reference and source bias over time: How Russia's partially state-controlled search engine Yandex mediated an anti-regime protest event. *Information, Communication & Society*, 1–17. <https://doi.org/10.1080/1369118X.2021.1933563>
- Lenta (2021, September 17). *V Kremle nazvali provokatsiie proekt 'Umnoe golosovanie'* [The Kremlin calls the “Smart Voting” project a provocation]. <https://lenta.ru/news/2021/09/17/provokac>
- Lokot, T., & Wijermars, M. (2021, September 30). The Kremlin forced U.S. tech firms to shut down an app some Russian voters hoped to use. Now what? *The Washington Post*. <https://www.washingtonpost.com/politics/2021/09/30/kremlin-forced-us-tech-firms-shut-down-an-app-its-opponents-were-using-now-what>

- Makhortykh, M., & Bastian, M. (2020, February 19). Personalizing the war: Perspectives for the adoption of news recommendation algorithms in the media coverage of the conflict in Eastern Ukraine. *Media, War & Conflict*, 15(1), 1–21. <https://doi.org/10.1177/1750635220906254>
- Makhortykh, M., Urman, A., & Roberto, U. (2020, May 11). How search engines disseminate information about COVID-19 and why they should do better. *Harvard Kennedy School (HKS) Misinformation Review*, 1(3). <https://doi.org/10.37016/mr-2020-017>
- Mangini, V., Tal, I., & Moldovan, A. N. (2020, August 25–28). An empirical study on the impact of GDPR and right to be forgotten - organisations and users perspective. In *ARES 2020: Proceedings of the 15th International Conference on Availability, Reliability and Security* (pp. 1–9). Association for Computing Machinery. <https://doi.org/10.1145/3407023.3407080>
- Novye Izvestiia. (2021, September 14). ‘Golosovat’ za KPRF beznравstvenno!’ Za chto oppositsiia kritikuet umnoe golosovanie [“Voting for CPRF is immoral!” Why opposition criticizes Smart Voting]. <https://newizv.ru/article/general/14-09-2021/golosovat-za-kprf-beznравstvenno-za-chto-oppozitsiya-kritikuet-umnoe-golosovanie>
- Ognyanova, K. (2019). In Putin's Russia, information has you: Media control and Internet censorship in the Russian Federation. In M. Merviö (Ed.), *Censorship, surveillance, and privacy: Concepts, methodologies, tools, and applications* (pp. 1769–1786). IGI Global. <https://doi.org/10.4018/978-1-5225-7113-1.ch088>
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- REN TV. (2021a, September 9). Pol’zovateli ‘Umnogo golosovaniia’ mogut byt’ legko identifitsirovany [The users of Smart Voting can be easily identified]. <https://ren.tv/news/v-rossii/877803-polzovateli-umnogo-golosovaniia-mogut-byt-legko-identifitsirovany>
- REN TV. (2021b, September 11). Zakharova: ‘Umnogo golosovanie’ sviazano s Pentagonom. <https://ren.tv/news/v-rossii/878505-zakharova-umnoe-golosovanie-sviazano-s-pentagonom>
- Roth, A. (2021, September 11). Putin’s crackdown: how Russia’s journalists became ‘foreign agents’. *The Guardian*. <https://www.theguardian.com/world/2021/sep/11/putins-crackdown-how-russias-journalists-became-foreign-agents>
- Sanovich, S., Stukal, D., & Tucker, J. (2018, April 1). Turning the virtual tables: Government strategies for addressing online opposition with an application to Russia. *Comparative Politics*, 50(3), 435–482. <https://doi.org/10.5129/001041518822704890>
- Schimpfössl, E., Yablokov, I., Zeveleva, O., Fedirko, T., & Bajomi-Lazar, P. (2020, January 22). Self-censorship narrated: Journalism in central and eastern Europe. *European Journal of Communication*, 35(1), 3–11. <https://doi.org/10.1177/0267323119897801>
- Schultheiß, S., Sünkler, S., & Lewandowski, D. (2018, September). We still trust in Google, but less than 10 years ago: An eye-tracking study. *Information Research: An International Electronic Journal*, 23(3). <http://informationr.net/ir/23-3/paper799.html>
- Sivetc, L. (2021, May 3). Controlling free expression “by infrastructure” in the Russian Internet: The consequences of RuNet sovereignization. *First Monday*, 26(5). <https://doi.org/10.5210/fm.v26i5.11698>
- Stadnik, I. (2021, May 3). Control by infrastructure: Political ambitions meet technical implementations in RuNet. *First Monday*, 26(5). <https://doi.org/10.5210/fm.v26i5.11693>
- Statscounter (2021). *Search engine market share: Russian Federation*. <https://gs.statcounter.com/search-engine-market-share/all/russian-federation>
- Stoycheff, E., Burgess, G. S., & Martucci, M. C. (2018, September 14). Online censorship and digital surveillance: The relationship between suppression technologies and democratization across countries. *Information, Communication & Society*, 23(4), 474–490. <https://doi.org/10.1080/1369118X.2018.1518472>

- Ulloa, R., Makhortykh, M., & Urman, A. (2021, December 21). *Algorithm auditing at a large-scale: Insights from search engine audits*. arXiv. <https://arxiv.org/abs/2106.05831>
- Urman, A. (2019). News consumption of Russian Vkontakte users: Polarization and news avoidance. *International Journal of Communication*, 13, 5158–5182. <https://ijoc.org/index.php/ijoc/article/view/11161/2838>
- Urman, A., & Makhortykh, M. (2021a, April 16). There can be only one truth: Ideological segregation and online news communities in Ukraine. *Global Media and Communication*, 17(2), 167–187. <https://doi.org/10.1177/17427665211009930>
- Urman, A., & Makhortykh, M. (2021b, May 11). *You are how (and where) you search? Comparative analysis of web search behaviour using web tracking data*. arXiv. <https://arxiv.org/abs/2105.04961>
- Urman, A., Makhortykh, M., & Ulloa, R. (2021, April 28). The matter of chance: Auditing web search results related to the 2020 US presidential primary elections across six search engines. *Social Science Computer Review*, 1–17. <https://doi.org/10.1177/08944393211006863>
- Weinberg, Z., Cho, S., Christin, N., Sekar, V., & Gill, P. (2018, October). How to catch when proxies lie: Verifying the physical locations of network proxies with active geolocation. In *Proceedings of the Internet Measurement Conference 2018* (pp. 203–217). Association for Computing Machinery. <https://doi.org/10.1145/3278532.3278551>
- Wijermars, M. (2021, February 15). Russia's law 'On news aggregators': Control the news feed, control the news? *Journalism*, 22(12), 2938–2954. <https://doi.org/10.1177/1464884921990917>
- Zavadski, A., & Toepfl, F. (2019, March 13). Querying the Internet as a mnemonic practice: How search engines mediate four types of past events in Russia. *Media, Culture & Society*, 41(1), 21–37. <https://doi.org/10.1177/0163443718764565>
- Zeveleva, O. (2020, January 11). Towards a Bourdieusian sociology of self-censorship: What we can learn from journalists adapting to rapid political change in Crimea after 2014. *European Journal of Communication*, 35(1), 46–59. <https://doi.org/10.1177/0267323119897798>

Funding

No funding has been received to conduct the research.

Competing interests

All of the authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics

Because our research did not involve data collection from human users or any interaction with human users, it was exempt from the ethical review under the guidelines of the Ethics Committee of the Faculty of Business, Economics and Informatics, University of Zurich. Further, since we ran only a few agents that did not perform actions such as clicking on specific items, we believe that our study did not interfere with the search algorithms in any potentially harmful way or affect the engagement statistics for individual results or put a heavy load on the search engines that could potentially affect their functionality.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data availability

Materials needed to replicate this study are available via the Harvard Dataverse: <https://doi.org/10.7910/DVN/C2LXTC>

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.