DISCUSSION PAPER SERIES

DP21789

PLATFORM POWER IN NEWS AGGREGATION

Andrey Simonov, Daniil Mikhailov, Ruben Enikolopov and Ruben Durante

INDUSTRIAL ORGANIZATION AND POLITICAL ECONOMY

ISSN 0265-8003

PLATFORM POWER IN NEWS AGGREGATION

Andrey Simonov, Daniil Mikhailov, Ruben Enikolopov and Ruben Durante

Discussion Paper DP21789
Published 26 July 2026
Submitted 22 July 2026

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

PLATFORM POWER IN NEWS AGGREGATION

Abstract

News aggregators are among the primary gatekeepers of online news, shaping both consumption and production through the algorithms that rank stories and sources. We provide the first large scale estimates of this algorithmic power using the example of Yandex News, Russia's largest aggregator. Combining archived front pages with 12.3 million published articles and outlet-level traffic data, we estimate that occupying the entire top-5 news block for a day raises an outlet's traffic by roughly 380,000 visitors. After a 2016 law made aggregators liable for cited content, Yandex halved its references to independent outlets, yet consumers did not penalize the manipulation: its referrals converted readers at the same rate, and its market share did not fall. Removed outlets shifted toward longer articles and stopped optimizing headlines for the algorithm, showing that aggregator algorithms shape not only news consumption but also production.

JEL Classification: L82, L86, D72, L51, L15, D83

Keywords: News aggregators, Search engines, Digital platforms, Recommender systems, Media capture, Platform power

Andrey Simonov - as5443@gsb.columbia.edu
Columbia University and CEPR

Daniil Mikhailov - daniilmikhailov11@gmail.com
Stanford University

Ruben Enikolopov - Ruben.Enikolopov@upf.edu
ICREA-Universitat Pompeu Fabra, IPEG and CEPR

Ruben Durante - ruben.durante@upf.edu
National University of Singapore, ICREA-UPF, CESifo, IZA and CEPR

Acknowledgements
Authors are listed in reverse alphabetical order and contributed equally. We thank Scott Gehlbach, Quan Le, Alexey Makarin, Karine Van Der Straeten, seminar participants at Carnegie Mellon University, Columbia University, Free University of Bozen-Bolzano, HEC Lausanne, LUISS, National University of Singapore, Stigler Center at the University of Chicago, UC Riverside, University of Colorado Boulder, Stanford GSB, University of Warwick, SITE, Digital Competition and Tech Regulation at Harvard Business School, Conference on AI, ML, and Business Analytics at Yale SOM, Polarize Conference at Bergen, 17th Digital Economics Conference at Toulouse School of Economics, MPWZ-CEPR, Barcelona Summer Forum, Economics of Advertising Conference, and Geoeconomics and Political Economy in the Alps conference for helpful comments and suggestions. This project is supported by the Chazen Institute for Global Business at Columbia Business School and by the Harriman Institute at Columbia University. Ruben Enikolopov acknowledges financial support from the Plan Nacional project PID2023-150768NB-I00 from the Spanish Ministerio de Ciencia e Innovación, and the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), funded by MCIN/AEI/10.13039/501100011033. Ruben Durante gratefully acknowledges financial support from the European Research Council (ERC) under the European Union's Horizon Europe research

and innovation program (Grant Agreement No. 101125953).

Platform Power in News Aggregation*

Andrey Simonov† Daniil Mikhailov‡ Ruben Enikolopov§
Ruben Durante¶

July 2026

Abstract

News aggregators are among the primary gatekeepers of online news, shaping both consumption and production through the algorithms that rank stories and sources. We provide the first large-scale estimates of this algorithmic power using the example of Yandex News, Russia's largest aggregator. Combining archived front pages with 12.3 million published articles and outlet-level traffic data, we estimate that occupying the entire top-5 news block for a day raises an outlet's traffic by roughly 380,000 visitors. After a 2016 law made aggregators liable for cited content, Yandex halved its references to independent outlets, yet consumers did not penalize the manipulation: its referrals converted readers at the same rate, and its market share did not fall. Removed outlets shifted toward longer articles and stopped optimizing headlines for the algorithm, showing that aggregator algorithms shape not only news consumption but also production.

*Authors are listed in reverse alphabetical order and contributed equally. We thank Scott Gehlbach, Quan Le, Alexey Makarin, Karine Van Der Straeten, seminar participants at Carnegie Mellon University, Columbia University, Free University of Bozen-Bolzano, HEC Lausanne, LUISS, National University of Singapore, Stigler Center at the University of Chicago, UC Riverside, University of Colorado Boulder, Stanford GSB, University of Warwick, SITE, Digital Competition and Tech Regulation at Harvard Business School, Conference on AI, ML, and Business Analytics at Yale SOM, Polarize Conference at Bergen, 17th Digital Economics Conference at Toulouse School of Economics, MPWZ-CEPR, Barcelona Summer Forum, Economics of Advertising Conference, and Geoeconomics And Political Economy In the Alps conference for helpful comments and suggestions. This project is supported by the Chazen Institute for Global Business at Columbia Business School and by the Harriman Institute at Columbia University. Ruben Enikolopov acknowledges financial support from the Plan Nacional project PID2023-150768NB-I00 from the Spanish Ministerio de Ciencia e Innovación, and the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), funded by MCIN/AEI/10.13039/501100011033. Ruben Durante gratefully acknowledges financial support from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (Grant Agreement No. 101125953).

†Columbia University and CEPR. Email: as5443@gsb.columbia.edu.
‡Stanford University. Email: daniilmikhailov11@gmail.com
§ICREA-Universitat Pompeu Fabra, IPEG and CEPR. Email: Ruben.Enikolopov@upf.edu
¶National University of Singapore, ICREA-UPF, CESifo, IZA, and CEPR. Email: rubendurante@gmail.com

1 Introduction

Digital platforms that mediate access to information—such as search engines, AI chatbots, social media, and news aggregators—have become the primary gatekeepers of online news consumption, responsible for the majority of news readership (Reuters, 2024). These platforms do not produce news but design algorithms that decide which stories are surfaced, in what order, and from which sources. News aggregators are the starkest example: their only product is a ranking of news events and sources, yet that ranking shapes how much news is consumed, which stories are read, and which outlets receive the readers.

How much can news aggregators impact the news ecosystem through their control of the algorithm that ranks news content? On the one hand, platforms can adjust the algorithm and manipulate news content shown to consumers. If consumers are swayed by these rankings, the platform can determine which news events are read and which outlets get traffic, changing news consumption and the incentives for news production. Aggregators would then hold market and media power (Prat, 2018) of a distinct kind: power over the distribution of content that others produce, rather than over production itself.[1] On the other hand, consumers might see through deviations from truthful aggregation—a ranking of events and outlets by their importance and relevance—and choose alternative channels for news consumption. In that way, demand can constrain the power of aggregators in the same way it constrains the media power of traditional media (Durante and Knight, 2012; Knight and Tribin, 2019).

We study the ability of news aggregators to influence online news consumption and production using the example of the major Russian news aggregator, Yandex News. We focus on the period between 2013 and 2017, which combines two useful features for our analysis. First, during this period, the front page of the Yandex search engine prominently featured the top-5 news stories from Yandex News, exposing tens of millions of daily users to the selected stories and referring these users to selected outlets. This allows us to collect historical data on Yandex's recommendations and gives statistical power to measure the impact of Yandex News's ranking on upstream news outlets' traffic. Second, during this period, Yandex made major changes to its news aggregation algorithm in response to changes in Russian legislation and pressure from the Russian government (Wijermars, 2021); we use this variation to examine whether discretion in changing the algorithm is constrained by demand factors and identify the impact of algorithms on news production.

In addition to providing a convenient empirical context to study general effects of news aggregators, the case study of Yandex News in Russia is important in itself, since it helps to assess the role Yandex played in the gradual government capture of the Russian online news market (Simonov and Rao, 2022). At the same time, the design we study is not idiosyncratic:

[1] Similarly, social media affects news consumption primarily through its effect on the personalized distribution of news content created by other online outlets (Braghieri et al., 2025).

nearly every major search engine other than Google—together accounting for roughly 30% of worldwide search activity—curates a similar news block on its front page, and Google itself has recently begun adding its Discover news feed to its mobile and desktop homepages (Section 2.2).

To understand the impact of Yandex's news aggregation algorithm on news consumption and production, we combine three new datasets. First, we use the Wayback Machine to scrape 29,841 snapshots of top-5 news stories on Yandex's homepage in 2013-2017. This allows us to track Yandex's news citations and outlet referrals over time. We augment these data with the universe of publication records (12.3 million articles) covering 50 major Russian online news websites, helping us to understand the underlying distribution of news articles and events in Russian media and measure changes in news production over time. Finally, for a subset of news outlets, we collect data from an internet tracker on daily traffic broken down by types of visitors, traffic sources, and other features. This data allows us to measure the effect of Yandex referrals on news consumption and outlets' market shares.

We start by estimating the impact of Yandex's top-5 referrals on the news outlets' traffic. Our identification strategy relies on the sharp difference in visibility between articles that appear in the top-5 section on Yandex's front page and those that do not. We exploit day-to-day variation in the exact share of top-5 slots that an outlet occupies over the course of a day, after absorbing outlet-by-week, outlet-by-day-of-the-week, and day fixed effects. As an additional check, we control for whether the outlet appears in the less visible positions just below the top-5, confirming that our estimates capture the effect of prominent front-page placement rather than the general newsworthiness of the outlet's coverage on that day.

Our results indicate that Yandex referrals have a large impact on traffic. A one standard deviation increase in how often an outlet's link appears in at least one top-5 slot during a day increases outlet traffic by around 27,600 visitors. If all slots in Yandex's top-5 positions were allocated to a single outlet for an entire day, its traffic would grow by nearly 380,000 visitors. Because the effect is pinned down in levels by the number of visitors to Yandex's front page rather than by outlet size, smaller outlets benefit disproportionately: the same referral traffic represents a 26% increase for the largest outlet in our sample but a 1,132% increase for the smallest one. The effect is larger for higher positions, ranging from 58,000 additional visitors for the fifth slot to 97,000 for the first, and becomes small and imprecise for positions 6–10, which do not appear on Yandex's front page.

We validate the causal nature of our estimates in a series of robustness checks. First, we show that referrals in Yandex's top-5 slots impact only outlets' traffic from Yandex but not direct traffic or traffic from other sources (e.g., social media and other aggregators), providing additional evidence that the effect is not driven by the general newsworthiness of the articles featured in Yandex's top-5. Second, appearing in Yandex's top-5 slots only affects an outlet's traffic on the day it is referred to but has no detectable effect on future traffic, ruling out strong intertemporal spillovers. On that day, however, incremental referrals generate 1.8 page views per visit, implying that referred users also browse articles beyond the one directly linked by Yandex. Finally, Yandex's referrals disproportionately affect casual consumers—who do not regularly visit the outlet and make fewer page visits—rather than regular ones.

Having measured the impact of Yandex's algorithm on outlets' traffic, we check whether Yandex's ability to direct traffic to outlets is constrained by the types of cited topics or the composition of visitors. We find little evidence for such constraints. First, the effects of Yandex's referrals on an outlet's traffic are similar across major topic areas (e.g., politics), for government-sensitive coverage (e.g., related to political opposition or government corruption), and for both high- and low-prominence events, suggesting that Yandex's capacity to drive traffic is not meaningfully constrained by the news events it cites. Second, we compare the effectiveness of Yandex across user characteristics. We find stronger effects for visitors identified by the traffic tracker as older and male, but no difference across operating systems (e.g., iOS, Windows) or device types (e.g., mobile, desktop).

To understand whether Yandex can change its algorithm at will and if it is constrained by demand factors, we turn to the analysis of the effect of the adoption of the major law "on news aggregators" in June 2016. The law made news aggregators liable for the content they cite, with some exceptions based on the source of news (Wijermars, 2021, see Section 2.3).

We observe a major change in Yandex's algorithm shortly after the signing of the law. In a sudden and dramatic change, Yandex stopped referencing 9 out of 24 outlets that used to regularly appear in the "top-5" news slots before the change, primarily targeting independent outlets. As a result, the share of top-5 references to independent outlets dropped from 20% to 10%, with all independent outlets other than two business-focused ones essentially excluded from the top-5 slots. At the same time, Yandex maintained broadly the same share of citations of sensitive events, indicating that the algorithm change operated through outlet removal rather than direct topic suppression.

Despite this drastic change, Yandex's ability to direct traffic to news outlets survived the law. We see no evidence of users losing confidence in the aggregator and discounting its recommendations. We also see no detectable change in Yandex's market share. Taken together, we find no evidence that the ability of the news aggregator to change its algorithm under political pressure is constrained by demand factors.

To understand the mechanism through which Yandex responded to the law, and through which it can exercise control over traffic allocation more broadly, we characterize its algorithm, guided by Yandex's public description of its design principles. The algorithm proceeds in two steps: first, it ranks news events—clusters of recent news articles—by prominence; second, it selects a specific article from a particular outlet for each event. We estimate both steps using our data on news events, constructed via a topic detection and tracking algorithm (Allan et al., 1998; Cagé et al., 2020), and Yandex's observed citations.

We find that in the first step, events covered by more outlets, with more articles, and developing faster are more likely to be cited. Notably, Yandex's algorithm under-reports events that government-controlled outlets systematically under-cover, even before the law.

In the second step, holding the event fixed, Yandex favors outlets that cover it early, mention more facts, write shorter titles, and do not copy earlier articles about the event; within an outlet, heavier coverage of the event also raises the chance of citation. However, much of the variation in referrals is explained by outlet weights, captured as fixed effects, which can be adjusted seamlessly over time. We show that the removal of independent outlets after the 2016 law was driven by a direct change in these weights, rather than adjustments to how article features are scored.

Using these estimates, we evaluate counterfactual scenarios that quantify Yandex's potential power over traffic allocation. If Yandex could direct all referrals to a single outlet with no restrictions on which events to cite, it could increase that outlet's traffic by between 26% and 1,132%, depending on outlet size. Even when constrained to cite the same events, Yandex could increase outlets' traffic by 15%–270% by adjusting only outlet weights. Evaluating the actual 2016 algorithm change, we find traffic losses of up to 18% for removed outlets and gains of up to 25% for favored ones, with aggregate traffic shifting from independent to government-controlled and potentially influenced outlets.

Finally, we document upstream production effects of the algorithm change. Comparing the 9 outlets removed from Yandex's top-5 slots to the remaining 15, we show that removed outlets shifted toward longer articles, reduced their coverage of aggregator-cited events, and stopped optimizing headline features for the algorithm, with adjustments taking roughly 10 months to materialize. This highlights the ability of Yandex to impact even inframarginal readers of news outlets: news producers respond to incentives and adjust their offering, changing the media diet of their readers.

To the best of our knowledge, this paper presents the first large-scale empirical evidence on the ability of news aggregators to impact news consumption and production through the control of their citation and referral algorithms. Existing work has shown that the general presence of aggregators and social media platforms increases news consumption from upstream producers (Sismeiro and Mahmood, 2018; Calzada and Gil, 2020; Athey et al., 2021) and has used natural experiments to show that adding or removing outlets from news aggregators impacts the outlets' traffic (Chiou and Tucker, 2017; George and Hogendorn, 2020). The related work used case studies of changes in the algorithms of social media and search engines to show their effect on news readership and engagement (Calzada et al., 2023; Dujeancourt and Garz, 2023; Moehring, 2024; Freimane, 2025). Our contribution to this stream of literature is in characterizing the algorithm that aggregators use and finding no evidence that consumers respond to the algorithm adjustments of the aggregator, which gives aggregators the power to preference particular outlets in their ranking and direct traffic to these outlets. We estimate Yandex's algorithm and show that it prioritizes particular outlets and down-weights government-sensitive news stories, a behavior similar to platform bias in other contexts (Reimers and Waldfogel, 2023).

We also show that news aggregators impact upstream news production due to the incentives of editors to produce and package news content in a way that will be referenced by the aggregator; these results relate to the theoretical work on product differentiation of outlets due to news aggregators (Dellarocas et al., 2013; Jeon and Nasr, 2016) and to the empirical work on the incentives behind news production (e.g. Sen and Yildirim, 2016; Cagé et al., 2022; Boxell and Conway, 2022; Sen et al., 2023; Garz and Szucs, 2023). Our work also speaks to the growing literature that examines the effects of adjusting algorithms or recommendation systems of platforms on user engagement, both in the context of news (Peukert et al., 2023) and more broadly (e.g. Anderson et al., 2020; Holtz et al., 2020; Aridor et al., 2022).

Finally, we contribute to the literature on the political economy of mass media (Prat and Strömberg, 2013), and especially to the study of media capture—the control of information by governments and other interested parties (Besley and Prat, 2006; Gehlbach and Sonin, 2014), a central source of media power (Prat, 2018). Such control shapes political knowledge, participation, and votes (DellaVigna and Gentzkow, 2010; Enikolopov et al., 2011), and the internet and social media have become central to this contest, especially in autocracies (Enikolopov et al., 2020; Zhuravskaya et al., 2020; Guriev and Treisman, 2019), with Russia's online news market being a leading case (Simonov and Rao, 2022). This literature has largely located capture at the level of news *producers*—who owns outlets and what they report (Djankov et al., 2003; Enikolopov and Petrova, 2015). Our contribution is to document capture one layer up, over the *distribution* of others' content: under government pressure, the algorithmic gatekeeper that ranks the news reshapes which outlets readers reach without altering what any outlet produces—and does so without prompting consumers to switch away.

The rest of the paper is organized as follows. Section 2 provides background information. Section 3 describes our data sources. Section 4 explains how we classify news articles into events, as well as describes the events. Section 5 estimates the effect of referencing outlets in Yandex's top-5 section on the outlets' traffic. Section 6 looks at the changes after the adoption of the law regulating news aggregators. Section 7 concludes.

2 Background Information

2.1 Yandex News

Yandex News (*Yandex.Novosti*) is an automated news aggregation and recommendation service operated by Yandex, a major Russian search engine. Like other aggregators, it does not produce original reporting; instead, it continuously ingests articles from a large pool of news outlets, detects underlying news events, ranks news events by prominence, and displays the most prominent events. The prominence of news events is determined by an algorithm, which takes into account available data on news production (e.g., volume of articles, recency) and consumption (e.g., clicks, shares on social media) (e.g., see Chiou and Tucker, 2017; Athey et al., 2021).

Our analysis focuses on Yandex News in 2013–2017, when Yandex was one of the two dominant search engines in Russia and Yandex News was the country's largest news aggregator.[2] The webpage of Yandex News `news.yandex.ru` displayed the five most important stories across all categories as well as the most important stories by category (politics, international, sports, etc.). A defining feature of this period is that Yandex prominently embedded the list of the five most important stories across categories on the Yandex homepage `yandex.ru`, immediately above the search box (see Figure 1). The top-5 news stories were global (similar for all visitors) and were frequently updated. A click on a news title directed a user to the referenced news outlet.[3] This design choice exposed a very large daily audience to the ranked headlines and generated substantial referral traffic to the selected outlets, making Yandex News a plausible "gatekeeper" for both news consumption and upstream production incentives.

2.2 Front-Page News Curation In Other Countries

This design in which top stories from a news aggregator are displayed on the front page of the corresponding search engine is common well beyond Yandex's Russian front page. Yandex itself operated localized versions of the same front page in other countries where it was, and remains, a major search engine: as of May 2026, its market share was 70% in Russia, 34% in Belarus, 29% in Kazakhstan, and between 10% and 24% in six other countries (StatCounter, 2026).

More generally, virtually all the most popular news aggregators in the world are linked to search engines. These search engines curate front-page news for a large share of the world's internet users. Nearly every major engine other than Google carries a news block on its homepage, including Bing, Yahoo, Baidu and Haosou in China, Naver and Daum in South Korea, and Seznam in the Czech Republic. Weighting each country's search-engine market shares by its number of internet users, Google accounts for roughly 70% of worldwide search activity, and engines that curate front-page news account for nearly all of the remaining 30%.[4]

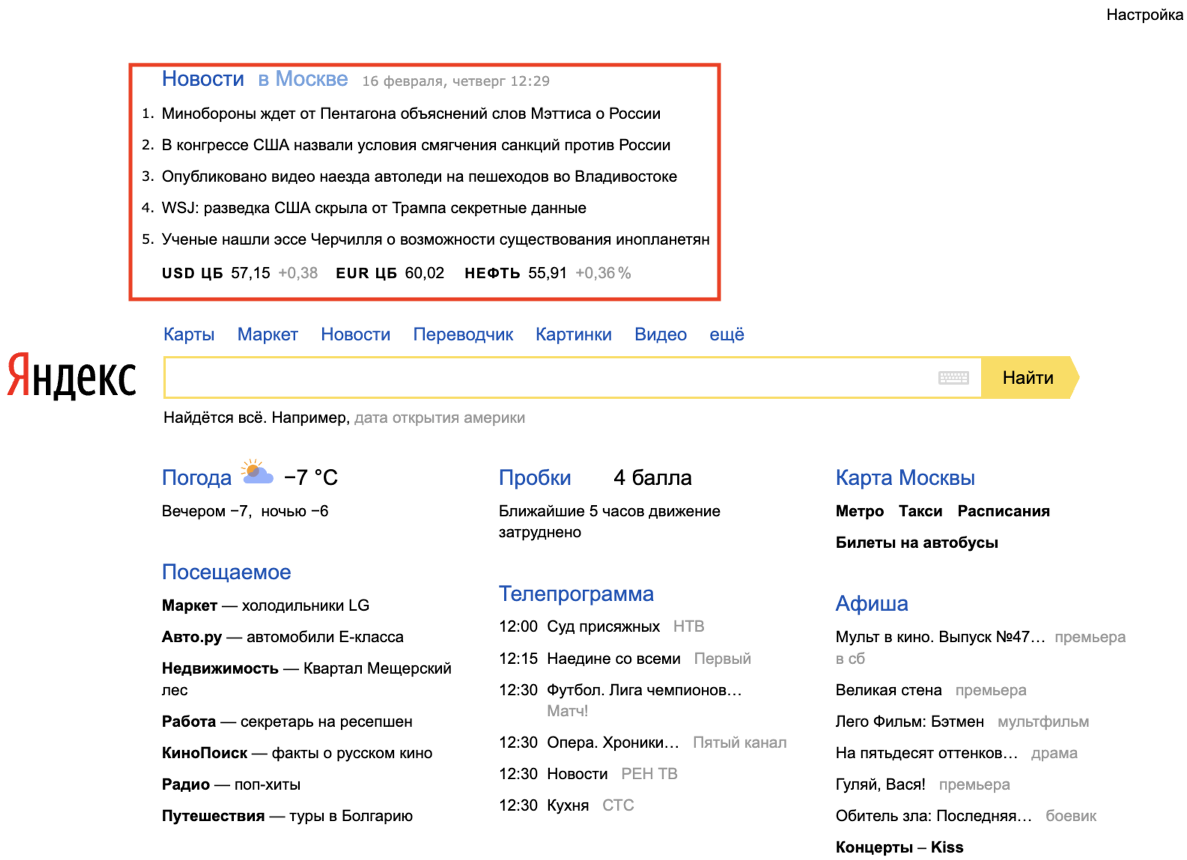
Google long kept its own front page bare, but has been moving in the same direction: since

[2]Yandex's market share in search was similar to Google. In early 2023, Yandex still held leadership in terms of referral traffic, responsible for 19% of referrals to news outlets (Statista, 2023). In September 2022, Yandex sold its news service – together with its main domain, `yandex.ru` – to VK.com (Yandex, 2022); the main page with the news service was rebranded as `dzen.ru`.

[3]This design changed in April 2018, when Yandex started to direct users to their subdomain that describes the news story. We focus on the period before this change.

[4]Appendix A1 describes the data and the weighting behind these figures.

Figure 1: An Example of the Front Page of Yandex



Notes: The snapshot depicts the front page of `yandex.ru` on February 16, 2017, at 8:55 am GMT+3. Links in the red rectangle are the global top-5 news of the moment, similar across all locations. These links refer a user directly to the news website that published the referred article, as shown at the bottom of the image (the cursor hovers over the first link, making it red).

2018, its mobile homepage and app have carried Discover, an algorithmically curated news feed below the search box (Google, 2018), and in 2025 Google began extending Discover to its desktop homepage (Schwartz, 2025). Unlike Yandex's top-5, though, Discover is personalized to each user rather than a single public ranking. Nor is this curation confined to users who open a homepage: the Google mobile app pushes notifications about news stories and topics selected for each user, delivering the algorithm's picks directly to the phone's screen (Google, 2026). Browsers have adopted the same design—Mozilla's Firefox fills its default new-tab page with stories chosen by its recommendation algorithms (Mozilla, 2026). The gatekeeping we quantify for Yandex thus operates, in some form, across the search engines, browsers, and mobile devices used by a large share of the world's internet users.

2.3 The Law "On News Aggregators"

The 2013–2018 period is especially important because it spans a major regulatory and political inflection point for Russian news intermediaries. On 23 June 2016, President Vladimir Putin signed a new federal law targeting news aggregators with large daily audiences (Wijermars, 2021), which became effective on 1 January 2017. The law covered only aggregators with more than one million daily users—impacting Yandex News but not Google News, whose Russian audience was below the threshold (Vedomosti, 2016; Hartog, 2017). Before the law, news aggregators were not held liable for the information they aggregated and distributed, with the legal responsibility carried by news publishers.[5] The law changed this, making aggregators liable for spreading "fake news" unless the information came from the media registered with the government.

We use this discrete policy-driven shift in the legal environment to study how it affected algorithmic rankings and how, in turn, this change in algorithmic rankings affected both downstream consumption (traffic to outlets) and upstream news production.

Political pressure on all media companies, including Yandex News, significantly increased after the start of the full-scale invasion of Ukraine in February 2022. As a result, Yandex sold the news aggregator to government-controlled VK group in September 2022 and renamed it to Dzen News. A 2026 VK presentation describes a draft bill, slated for the State Duma in May 2026, that would require every large platform—marketplaces, social networks, search engines, and video services with audiences above five million—to embed a state news widget built on Dzen News (Rudkovskiy, 2026). The widget would show a government-curated "top-5" news block whose permitted sources are chosen by the authorities rather than by an algorithm or by editors. What we study as one aggregator's discretionary ranking of a single webpage would thus become a mandatory, state-defined ranking embedded across the entire Russian internet.

3 Data

We combine three novel data sources to characterize the Yandex News algorithm and measure its effect on news consumption and production: the citations of Yandex News, publication records from major news outlets in Russia, and records of news consumption.

3.1 Yandex News

The first dataset that we create contains historical citations of news by Yandex News in Russia. We construct these citations using the Internet Archive, a digital library containing historical snapshots of various web pages. Our primary interest is in the five most important stories of the

[5] This is similar to the legislation in other jurisdictions, e.g., Section 230 of the Communications Decency Act (CDA) in the United States that provides legal immunity to online platforms for distributing third-party content.

moment that are shown on the main page of `yandex.ru`.

Since Yandex's front page is visited by tens of millions of users daily, the Internet Archive has a very detailed collection of its snapshots (the number of snapshots is proportional to the popularity of the web page). Our sample contains 29,841 snapshots of the front page covering the period between June 2013 and December 2017. The median time between two snapshots is 32 minutes. From each snapshot, we extract the top-5 news titles and the underlying URL links. We augment the data with snapshots of the news rankings from the Yandex News (`news.yandex.ru`) page, which, in addition to the top-5 news titles, also contains the articles that were close to entering the top-5 by being among the top news stories in different categories, but did not quite make it to appear on the front page of Yandex.

From these snapshots, we construct our key regressor: the share of Yandex's top-5 slots occupied by a particular outlet over a day, which equals one if the outlet held all five slots for the whole day. We also construct a disaggregated measure for each of the five slots separately. From the `news.yandex.ru` rankings, we additionally construct indicators for "near-miss" positions "top-6" through "top-10": stories that led the topic subcategory rankings without reaching the front page. These positions carry comparable newsworthiness but no front-page visibility, and they play two roles in the analysis that we describe in Section 5.2: controlling for them separates the effect of front-page placement from the newsworthiness of the outlet's coverage that day, and their own null effects on traffic serve as a placebo test of the research design. Because `news.yandex.ru` attracts less traffic than `yandex.ru`, its Wayback Machine coverage is sparser, and these variables are available for fewer days. Appendix A2 details the construction.

3.2 Publication Records

We complement the Yandex citation records with the publication records of 50 major online news outlets in Russia at the time. Data were provided by Medialogia, a public relations company that collects the universe of social media posts and media publications in Russia (`mlg.ru`). The publication records include outlet names, article titles and texts, the publication time, and URLs. The data contains 12,336,080 news articles published during the period from 23 June 2013 to 4 December 2017.

Both government-controlled and independent news outlets coexisted in the Russian media space at the time. We follow the approach of Simonov and Rao (2022) and use outlets' ownership structure, location, and information on the independence of their editorial boards to classify news outlets into four groups: government-controlled, which are directly owned and controlled by the government; independent but potentially-influenced, which are privately owned but have close ties to the government, or for which there is evidence that the editorial board is not independent; independent, which are privately owned and have independent editorial boards; and international, which are based outside of Russia but focus their news on the Russian audience.

We match the publication records to the news that Yandex cites in two steps. First, for Yandex, we can match published and cited news articles directly by the article URLs, since they are observed on the front page of Yandex and in the publication records. [6] Second, we cluster publication records into news events and match Yandex-cited news to the corresponding events using the distance between the titles of the cited articles and the center of the cluster. See Section 4 for more details.

3.3 Consumption Data

The third dataset that we create measures the daily news consumption of a subset of major news websites in our sample. We assemble this dataset using a traffic tracker of `top.mail.ru`. Websites install this tracker to help with marketing analysis, and can choose to make different types of data public. Eleven out of fifty outlets in our sample keep their overall traffic open for the public; this includes daily records of the total visitors, visitors split by the intensity and frequency of visits (e.g., no visits in the last 31 days), visitor demographics (age, gender, region), total pageviews, desktop and mobile visitors, among other metrics. For a subset of these websites, we observe more detailed statistics, such as daily visits to particular webpages, the location of visitors, their browser, and their operating system. In particular, four websites provide data on the source of their traffic, allowing us to separate traffic from Yandex, other news aggregators, or other referral domains.

We confirm the quality of data from `top.mail.ru` using other public data trackers in Russia and abroad. In particular, we use `liveinternet.ru`, `rambler.ru`, and `semrush.com`; while these domains do not allow us to pull historical data from 2013-2017, we examine the correlation in the traffic records from these trackers and `top.mail.ru` in the more modern data. Appendix A3 provides details of this validation exercise. We also use traffic data for all 50 news outlets since 2017, provided by `semrush.com` to estimate the effect of changes in the aggregator's algorithm on overall news consumption.

We supplement the traffic data with data on political preferences, beliefs, and demographic characteristics from the MegaFOM opinion poll conducted by the Public Opinion Foundation (Fond Obschestvennogo Mneniya, or FOM) in October–November 2011. This is a regionally representative survey of 56,900 respondents from 79 regions. We use these data to characterize the effectiveness of Yandex in directing consumers of different political beliefs to news articles and topics. See Appendix A4 for details about the dataset.

[6] Yandex stopped directing users to source outlets in April 2018. Four months earlier, in December 2017, Yandex redesigned its main page to display a "dynamically updated" top-5 ranking that changed while users remained on the page. Although the HTML code contains entries for all articles that appeared in the top-5 positions, we restrict our analysis to the period prior to this redesign to ensure that each ranked position was actually observed by users.

4 Classification of News Events

Because Yandex's algorithm operates on news events rather than individual articles (see Section 6.2), we need to detect and characterize events in the data both to characterize the effect of Yandex's referrals on outlet traffic and the algorithm itself.

We classify articles into news events using the topic detection and tracking algorithm (TDT) developed by Allan et al. (1998) and used in previous work (e.g. Cagé et al., 2020). The method represents articles as TF-IDF vectors and clusters those published within overlapping three-day windows based on cosine similarity; clusters that share articles across windows are merged into a single event. We select hyperparameters—including the vectorization method, clustering algorithm, and radius—using the fact that our dataset includes 19 Ukrainian Russian-language outlets alongside the 50 Russian ones. Since Ukrainian outlets should cover largely distinct events from the Russian ones, we select the specification that best separates Russian and Ukrainian articles while maintaining adequate clustering coverage. Our preferred specification uses Agglomerative Clustering with a radius of 0.6; we provide details and verify robustness to alternative settings in Appendix A5.

The resulting classification produces 1.88 million news events (defined as clusters of at least two articles), covering 71% of the 12.3 million articles in our data. Panel A of Table 1 presents summary statistics: a median event has three articles from two outlets and lasts 5.5 hours. The distribution is heavily right-skewed, with the largest event containing 1,709 articles from 49 outlets. We match these events to Yandex's citations using the article URLs observed on Yandex's front page. Events cited by Yandex (Panel B) are substantially larger—a median of 14 articles from 12 outlets lasting 12.3 hours—as expected given that the algorithm prioritizes prominent events.

We classify events as government-sensitive based on whether they are systematically underreported by government-controlled outlets relative to independent ones, following the procedure of Simonov and Rao (2022) (see Appendix A6 for details). Panel C of Table 1 summarizes these events. They are larger and longer-lasting than typical events, and predominantly involve political opposition and government corruption—for example, the activities of Alexey Navalny or changes in the editorial leadership of independent media.

Finally, we group news articles into general news topics. For this, we estimate a structural topic model (Roberts et al., 2013) based on article texts and detect 30 different topics. We group these topics into 16 broader categories.[7]

[7]A set of 16 categories closely corresponds to the topic groups employed by Yandex News. These categories include Foreign Politics, Military, Economy, Culture, Incidents, Local, Society, Education, Sports, Domestic Politics, Science, Crime, Weather, Health, Technology, and Environment.

Table 1: Summary Statistics of News Events

	Min	25%	Median	Mean	75%	Max
Panel A: All Events						
articles per topic	2	2	3	4.69	5	1709
outlets per topic	1	2	2	3.57	4	49
topic length (hours)	0	1.68	5.52	15.08	20.52	359.95
Panel B: Yandex Top-5 Events						
articles per topic	2	7	14	30.77	28	1709
outlets per topic	1	6	12	14.34	21	49
topic length (hours)	0	4.32	12.30	33.90	40.47	357.90
Panel C: Sensitive Events						
articles per topic	10	14	19	34.06	31	1543
outlets per topic	10	12	15	18.46	22	49
topic length (hours)	0.37	10.84	22.68	40.24	50.76	356.93

Notes: Panel A reports statistics for all 1.88 million detected news events. Panel B reports statistics for 34 thousand events cited in Yandex Top-5 news at least once. Panel C reports statistics for 8,734 sensitive events, i.e., events systematically under-reported by government-controlled outlets relative to independent ones (Section 4).

5 Effects on News Consumption

5.1 Descriptive Evidence

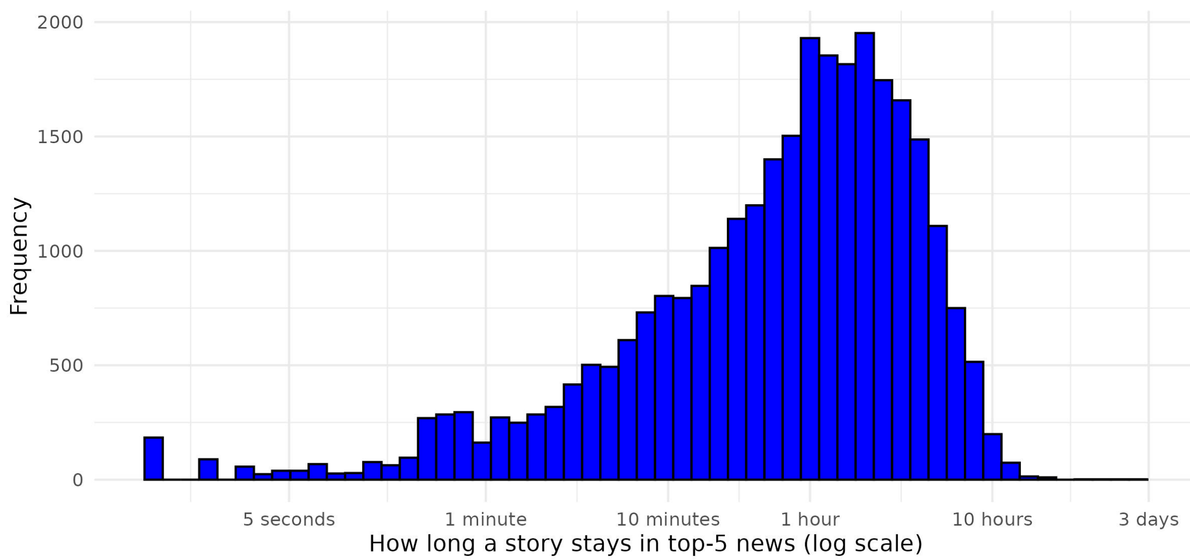
We begin by documenting two features of how Yandex's top-5 section operates. The first supplies the variation that our research design exploits; the second shows how much front-page attention the section can concentrate on a single outlet.

First, articles rotate rapidly through the top-5. On a typical day, 60–70 different articles pass through the section, and the median story stays for only 56 minutes (Figure 2).[8] Almost no article remains for more than 10 hours. This rapid turnover makes an outlet's share of top-5 slots swing sharply from one day to the next, and it is this day-to-day variation that we exploit for identification below.

Second, a single outlet can hold several slots at once: the most-cited outlet of the moment occupies 1.84 slots on average. The section can therefore concentrate a large share of front-page attention on one outlet, which is what gives the algorithm its scope to redirect traffic. This concentration rose further after the 2016 law on news aggregators, as we show in Section 6.

[8]This is based on articles observed in at least two snapshots, so the actual rotation is even faster.

Figure 2: Duration of News Articles in the Top-5 Yandex News



Notes: A histogram of the duration of stories in the top-5 section of Yandex. We exclude articles that we observe only on one snapshot of Yandex.

5.2 Effects on Outlet Traffic

We estimate the causal effect of top-5 placement on outlet traffic by exploiting the day-to-day variation in outlets' top-5 exposure:

$$y_{jd} = \alpha X(j \text{ in top-5})_d + \rho_{jw} + \delta_{j,\text{wd}} + \theta_d + \xi_{jd} \tag{1}$$

where y_{jd} is the total number of visitors (in logs or levels) to outlet j on day d, and $X(j \text{ in top-5})_d$ is the share of top-5 slots occupied by j over day d (it equals 1 if j held all five slots for the whole day). The terms ρ_{jw}, $\delta_{j,\text{wd}}$, and θ_d are outlet-by-week, outlet-by-day-of-the-week, and day fixed effects. The error ξ_{jd} is an idiosyncratic shock to outlet j's traffic on day d.

Our identifying assumption is that, conditional on these fixed effects, an outlet's share of top-5 slots is uncorrelated with the other determinants of its traffic that day (ξ_{jd}). The outlet-by-week effects ρ_{jw} flexibly absorb changes in outlets' algorithmic weights over time, so the regression uses only residual within-week variation in top-5 exposure.

The main threat to this identifying assumption is that the effect is driven by unobserved newsworthiness of the stories: Yandex places an outlet in the top-5 partly *because* its coverage that day is unusually newsworthy, and that newsworthiness may draw traffic on its own, so α could conflate the effect of placement with the effect of the underlying coverage. We address this in three ways. First, some specifications control for $I(j \text{ in top-6})_d$, an indicator for whether outlet j appears in the "top-6" slot on day d.[9] Articles in the top-6 slot are comparably newsworthy—

[9] We have fewer observations that capture $I(j \text{ in top-6})_d$, so the sample shrinks when we include it.

they nearly entered the front page but are not displayed on it—so the top-5 coefficient is identified from the difference in visibility, holding newsworthiness roughly fixed. This approach is close in spirit to a regression discontinuity design based on the relative ranking of the stories, with a discontinuous jump in prominence for the stories that make it to top-5. Second, for the subset of outlets with article texts, other specifications control for $X(\text{coverage of top-5 events by } j)_d \in [0,1]$, the share of that day's top-5 events that outlet j covered, weighted by each event's presence in the top-5, which directly absorbs the newsworthiness of the outlet's own coverage. Third, using the four outlets with traffic-by-source data, we split the effect by channel and find it confined to Yandex traffic; a newsworthiness confound would have lifted their direct and other-referral traffic too, so its absence there points to placement rather than newsworthiness.

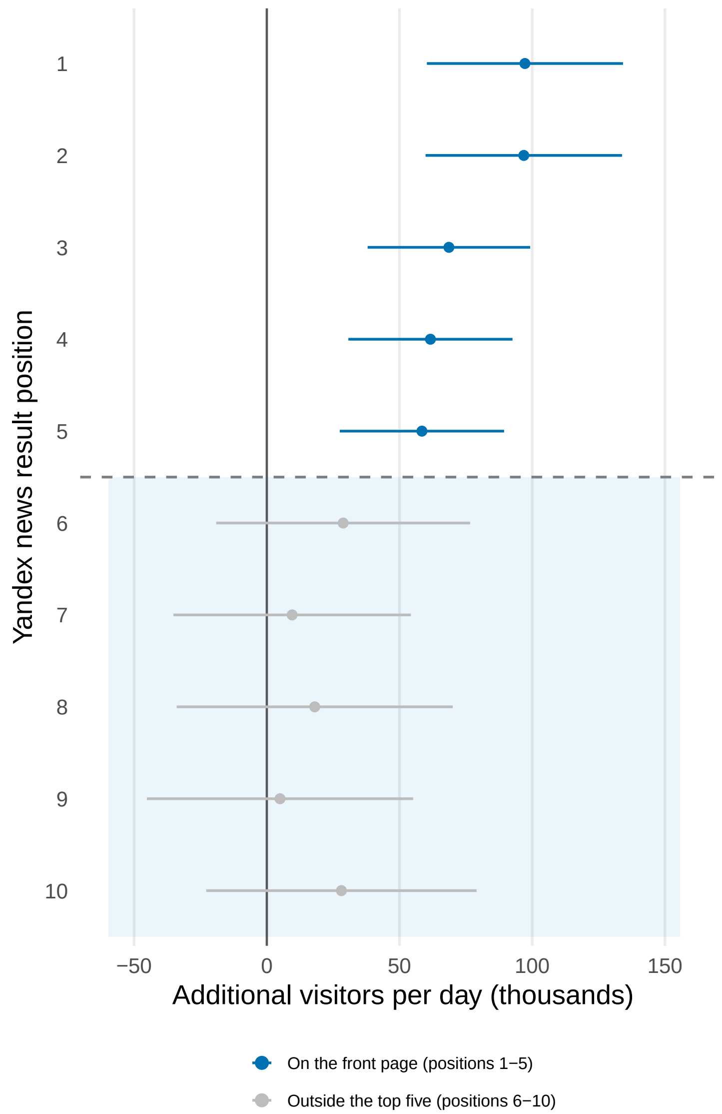
Table 2: Effect of Outlet References in Yandex's Top-5: Overall Traffic

	Dependent variable:					
	$log(y_{jd})$			y_{jd}		
	(1)	(2)	(3)	(4)	(5)	(6)
$X(j \text{ in top-5})_d$	0.46***	0.45***	0.40***	341.52***	374.03***	366.48***
	(0.03)	(0.04)	(0.04)	(32.53)	(38.13)	(38.93)
$I(j \text{ in top-6})_d$		0.04***	0.04***		10.64***	8.43**
		(0.01)	(0.01)		(3.43)	(3.58)
$X(\text{coverage of top-5 events by } j)_d$			0.10***			18.09**
			(0.01)			(8.48)
Fixed Effects:						
- Week x Outlet	Y	Y	Y	Y	Y	Y
- Week Day x Outlet	Y	Y	Y	Y	Y	Y
- Date	Y	Y	Y	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks	Weeks	Weeks	Weeks
Observations	22,971	20,084	16,057	22,971	20,084	16,057
Adjusted R^2	0.98	0.98	0.98	0.95	0.95	0.96

Notes: The dependent variable y_{jd} is the number of daily visitors to outlet j, in thousands; coefficients in columns 4–6 are therefore in thousands of visitors (374.03 $\approx$ 374,000 visitors for full-day occupancy of all five slots). Standard errors, clustered by week, in parentheses. *p<0.1; **p<0.05; ***p<0.01.

The results indicate that top-5 placement generates large traffic gains. Table 2 reports the estimates of α in logs and levels. A full-day occupancy of all five slots raises an outlet's traffic by 341,000–374,000 visitors (columns 4–6), or by 49.2%–58.4% (columns 1–3). Controlling for top-6 placement and event coverage leaves the estimate essentially unchanged, so the effect does not reflect the general newsworthiness of the day's stories. Our preferred specification is in levels (columns 4–6): the size of Yandex's front-page audience pins down the magnitude of a referral in absolute terms, not in proportion to outlet size. The same referral therefore delivers a far larger proportional gain to a small outlet than to a large one—a gap we quantify in the counterfactuals of Section 6.

Figure 3: Effect of Outlet References in Yandex's Top-5: By Position



Notes: Points correspond to the estimates of regression (1) with $X(j \text{ in top-5})_d$ broken into ten position-specific shares; effects are in thousands of visitors for a full-day occupancy of the given position. Positions 1–5, above the dashed red line, appear on Yandex's front page; positions 6–10 do not. Error bars correspond to the 95% confidence intervals, with standard errors clustered by week. Table A10 in the Appendix reports the underlying estimates.

The results show that higher slots within the top-5 news deliver more traffic. Breaking X into ten position-specific shares, Figure 3 plots the effect of a full-day occupancy of each position: position 1 adds 97,000 visitors, against 58,400 for position 5. Summed across the five slots, the implied effect is 383,000 visitors, consistent with the pooled estimate of 374,000 in Table 2. Positions 6–10, which are not displayed on the front page, have small and statistically insignificant effects: because these positions reflect comparable newsworthiness without front-page visibility, their null effects confirm that placement—not the underlying interest of an outlet's coverage—drives the traffic. The underlying estimates are reported in Table A10 in the Appendix and are robust to alternative controls for the "top-6"–"top-10" positions.

Five further checks support the causal interpretation and characterize the referred traffic. First, the effect is driven by the traffic coming from the Yandex website. For the four outlets with traffic-by-source data, we split total visitors by referral source and re-estimate equation (1) on each component (Table 3). Column 1 first replicates the main estimate on this subsample: an increase of 461,000 visitors, less precise (a standard error of 67,000 rather than 38,000) but statistically similar to the baseline. All of these visitors are navigating from Yandex (422,000, column 2) while the direct traffic (column 3) and non-Yandex traffic (column 4) remain unchanged. This points to placement effects rather than newsworthiness: had the underlying coverage drawn readers, their direct traffic and other referrals would have risen too.

Second, using disaggregated data on the type of visitors we show that referrals bring casual readers, not the regular readers of the corresponding outlets. We re-estimate equation (1) separately for "new" visitors (no visit in the prior three months, by `top.mail.ru`'s definition) and "regular" visitors (at least 7 visits in the last 28 days). A full-day top-5 occupancy adds 157,000 new visitors and zero regular ones (Table A11, columns 2–3); the remaining incremental visitors are returning but infrequent. The referrals are incremental and do not convert into loyal readership.

Third, the marginal visitor browses less than the inframarginal one. The 374,000 additional visitors generate 681,000 additional pageviews (Table A11, column 4), or 1.8 pageviews per marginal visit—against roughly 4 for a typical visit to an average outlet.[10] Pageviews matter for these outlets because they monetize through display advertising rather than subscriptions.[11]

Fourth, the effect does not persist. Re-estimating equation (1) with outcomes from days $d+1$ through $d+14$, no later-day estimate is statistically distinguishable from zero, and we can reject an effect as large as 100,000 visitors against the day-of effect of 374,000 (Figure 4). Thus, the results are not driven by temporary swings in the popularity of different outlets.

Fifth, measurement error in our exposure variable, if anything, biases the estimates downward.

[10]Pageviews per marginal visitor are 681,000/374,000 $\approx$ 1.8. For inframarginal visitors, we combine the estimates in Tables 2 and A11 with the result that the 374,000 additional visitors correspond to 56.2% of an average outlet's daily visitors.

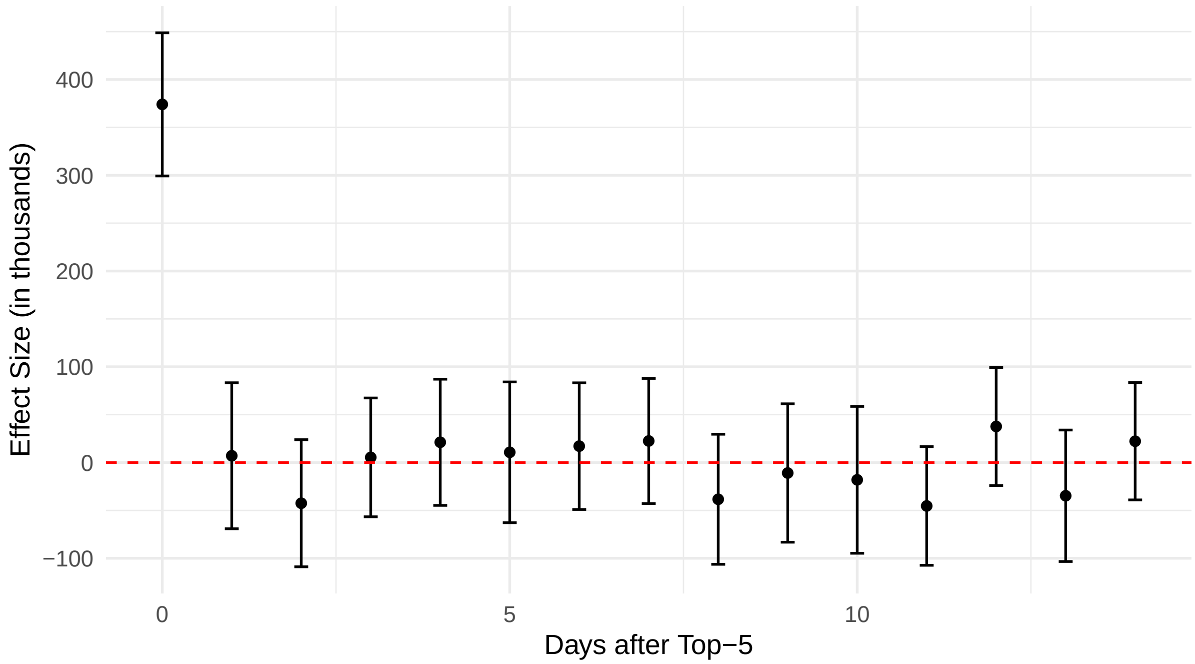
[11]The only exception is the small independent outlet `slon.ru`, which introduced subscriptions in mid-2014 and was renamed `republic.ru` in 2016.

Table 3: Effect of Outlet References in Yandex's Top-5: Traffic by Source

	Dependent variable:			
	y_{jd}	y_{jd}^{Y}	y_{jd}^{-Y}	$y_{jd}^{\text{ext}-Y}$
	(1)	(2)	(3)	(4)
$X(j \text{ in top-5})_d$	461.61***	421.91***	39.70	30.42
	(66.82)	(56.00)	(33.69)	(27.79)
$I(j \text{ in top-6})_d$	7.96	7.59**	0.37	-0.56
	(6.20)	(3.81)	(4.19)	(2.42)
Fixed Effects:				
- Week x Outlet	Y	Y	Y	Y
- Week Day x Outlet	Y	Y	Y	Y
- Date	Y	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks	Weeks
Observations	5,716	5,716	5,716	5,716
Adjusted R^2	0.94	0.86	0.96	0.94

Notes: The sample is the four outlets with traffic-by-source data. Dependent variables, in thousands of daily visitors: total visitors (y_{jd}), visitors arriving from Yandex (y_{jd}^{Y}), direct traffic (y_{jd}^{-Y}), and traffic from other external referrers ($y_{jd}^{\text{ext}-Y}$). Standard errors, clustered by week, in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Figure 4: Effect of Yandex's Top-5 Citations: Next Two Weeks



Notes: Points correspond to α estimates from regression (1) with $y_{jd'}$, $d' \in \{d, d+1, ..., d+14\}$ as an outcome variable. Error bars correspond to the 95% confidence intervals.

We reconstruct an outlet's daily share of top-5 slots from archived snapshots of the front page—about 18 per day on average, a coarse sample of a section through which 60–70 stories rotate daily, each lasting a median of 56 minutes (Section 5.1). The resulting measure of top-5 exposure is noisier on days with fewer snapshots, and such classical measurement error should attenuate α toward zero. Re-estimating equation (1) with observations weighted by the number of available snapshots per day—giving more weight to better-measured days—raises the level estimate from 374,000 to 437,000 visitors (Table A12), consistent with attenuation in the baseline.

5.3 Heterogeneity of Effects

We examine the heterogeneity of the effect of Yandex references to outlets to answer two further questions. First, examining whether the effect varies with the *events* Yandex cites tells us if its influence is constrained by the news it must cover—the binding question for the counterfactual algorithms we study in Section 6. Second, examining whether it varies across readers tells us *who* is more likely to be influenced by Yandex's choice of the top-5 news—which determines the value of controlling the algorithm to an actor seeking to move particular audiences. We examine variation across news types, regions, visitor demographics, and political views.

For these comparisons we normalize the outcome by Yandex's own audience, dividing y_{jd} by the number of users navigating from Yandex search that day, y_{jd}^{Yandex}.[12] The normalized coefficient is the share of Yandex users who follow a top-5 link—a rate that is comparable across topics, regions, and demographic groups and that nets out growth in Yandex's front-page audience over time. We re-estimate equation (1) with this normalized outcome.

Effects by News Type. The effect varies little with the topic of news cited. Interacting $X(j \text{ in top-5})_d$ with each of the 16 topic categories defined in Section 4, only the Sports interaction is significant at the 5% level, and it is negative (Table A13 in the Appendix). Yandex is equally effective at directing traffic to events under-reported by government-controlled outlets (Table A14), so it could raise or lower readers' exposure to sensitive coverage without losing reach. Ranking events by their newsworthiness, as measured by the predicted top-5 citation probability—from the first step of Yandex's algorithm, which we estimate in Section 6.2—the top and bottom quartiles of newsworthiness generate statistically indistinguishable traffic (Table A15). Overall, the results indicate that the influence of Yandex is not affected by the type of news it cites.

Effects on Local News. Although Yandex visitors look indifferent across topics, they have strong tastes for local news. We re-estimate effectiveness across 83 Russian federal subjects, stacking observations across regions and interacting every control in equation (1) with region fixed effects, and then classify each top-5 story by whether it concerns the visitor's home region (for example, a hockey game in Saint Petersburg shown to a Saint Petersburg visitor). For these region-relevant stories, effectiveness rises from 2.7 to 17.5 percentage points—a more than sixfold increase (Table A16). The results suggest that readers have stronger tastes for local news than for particular topics or political sensitivity.

Effects by Visitor Characteristics. The effect does not differ by operating system or device type (Table A17), but it does differ by demographics. Effectiveness rises with age—0.034 for visitors over 50, roughly two and a half times the 0.014 for visitors aged 19–24 (Table A18)—and is 60 percent higher for men compared to women (0.029 against 0.018; see Table A19).

Effects by Political Views. Finally, we examine whether influence of Yandex differs for groups of population with different political views. We do not have information at the individual level, but we can exploit the fact that we have information on traffic by demographic (age and

[12]We use the number of unique people navigating from Yandex to websites tracked by `top.mail.ru` to measure the number of Yandex users on a given day. This measure counts active users of Yandex but does not capture users who came to Yandex without searching; we assume these inactive users were also uninterested in the front-page news. On average there are 1.85 times more visitors to the front page than there are users who search on Yandex, so in order to account for the less active users all effectiveness measures must be divided by 1.85. See Appendix A3.1 for details and validation of the Yandex visitors' measure.

Table 4: Factors Correlated with Yandex Top-5 Effectiveness by Political Views

	Dependent variable:					
	Effectiveness of Yandex Top-5					
	(1)	(2)	(3)	(4)	(5)	(6)
Gov. party supporters	-0.036^{***}					-0.030^{***}
	(0.002)					(0.002)
Low income		-0.014^{***}				-0.008^{***}
		(0.002)				(0.002)
Likely protesters			0.010^{***}			0.001
			(0.002)			(0.002)
See corruption as an issue				0.025^{***}		0.017^{***}
				(0.002)		(0.002)
Support democracy					0.023^{***}	0.010^{*}
					(0.007)	(0.006)
Constant	0.040^{***}	0.028^{***}	0.022^{***}	0.013^{***}	0.022^{***}	0.030^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Weights:	# Respondents	# Respondents	# Respondents	# Respondents	# Respondents	# Respondents
Observations	1,229	1,229	1,229	1,229	1,229	1,229
R^2	0.271	0.033	0.014	0.137	0.010	0.345
Adjusted R^2	0.270	0.032	0.013	0.136	0.009	0.342

Notes: The unit of observation is one demographic group (Region x Age x Gender; 1,229 cells). The dependent variable, effectiveness of the Yandex top-5, is the cell-specific estimate of α from equation (1) with traffic normalized by Yandex's daily audience: the share of Yandex users in that cell who follow a top-5 link under full-day occupancy of all five slots. Each X variable is the share of the cell's survey respondents choosing the corresponding answer (Appendix A4); only respondents who report frequent internet use are included. Observations are weighted by the number of survey respondents in the cell. *p<0.1; **p<0.05; ***p<0.01.

gender) groups, as well as by regions. We re-estimate α separately for each (age, gender, region) cell—1,229 cells in all—and merge the estimates with cell-level shares of government support, income, protest activity, and attitudes toward corruption and democracy from the MegaFOM survey (Section 3).[13] Yandex is more effective at directing traffic to demographic cells with fewer supporters of the incumbent party, higher incomes, and greater concern about corruption; these three correlations survive jointly, while protest propensity and support for democracy correlate with effectiveness only one at a time (Table 4). Two caveats apply: the survey predates our traffic data, and the correlations are estimated across demographic cells rather than individuals. With those caveats, top-5 placement affects the politically active, opposition-leaning part of the audience disproportionately—a pattern that, all else equal, makes control of the algorithm more valuable to a government seeking to manage those readers.

6 Effects of the Law

Section 5.2 measured how much traffic the top-5 moves while taking Yandex's algorithm as constant. The algorithm, however, is Yandex's to change. Whether that discretion amounts to genuine power depends on the answers to two questions: can Yandex change the algorithm without losing the audience, and to what extent can such a change reshape which outlets readers reach? We address both by looking at the effect of the 2016 law "on news aggregators"—which made aggregators liable for the content they cite (Section 2). We document how Yandex's citations changed after the passing of the law, characterize the two-step algorithm that selects top-5 references, test whether consumers constrained the change, quantify how far it could shift traffic across outlets, and ask whether it altered the news that outlets produce.

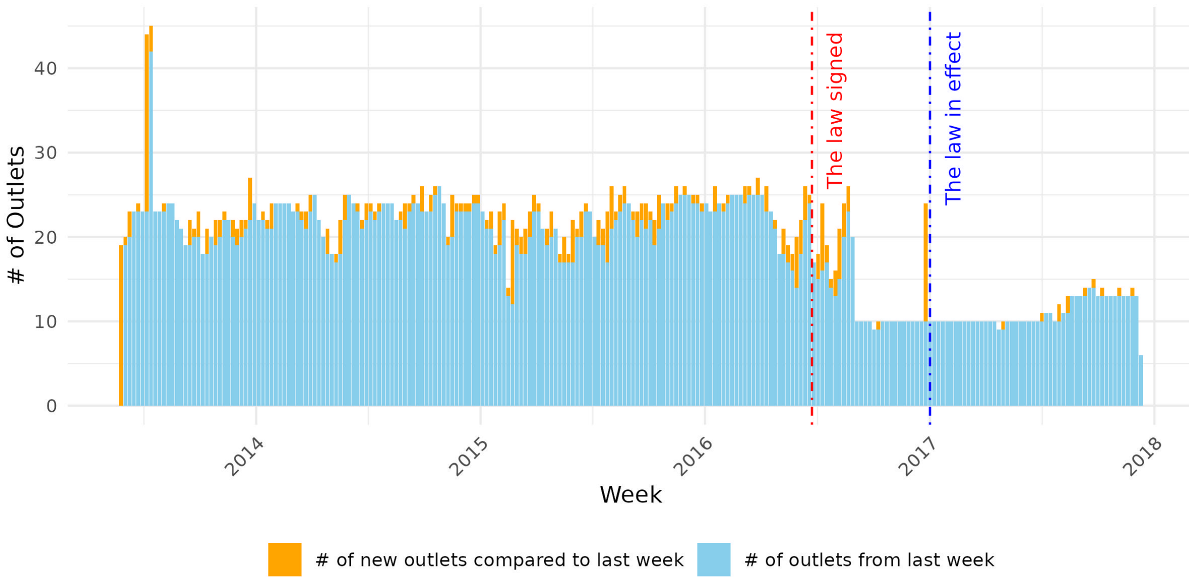
6.1 Descriptive Evidence

The data demonstrate that Yandex changed its citations abruptly once the law was signed. Figure 5 plots the number of distinct outlets that appeared in the top-5 each week. In a typical week before the passing of the law, around 24 outlets reached the top-5; within weeks of the signing of the law in late June 2016, that number fell to about 10 and stayed near 10 for almost a year, recovering only to just below 15 by the end of 2017—still well under its pre-law level.[14] On net, 9 of the 24 pre-law outlets were dropped for good. In the analysis we date the algorithm change to September 2016 throughout, ignoring the partial readjustment in the end of 2017.

[13] We restrict the sample to respondents who report frequent internet use, so that both the survey shares and the traffic estimates reflect online users; results are qualitatively similar using all respondents.

[14] A 2022 journalistic investigation reported that President Putin's administration had agreed with Yandex on a list of around 15 outlets cleared to appear in the top-5 news as early as 2015 (Reiter, 2022). The post-law count settled at this number, and the surviving outlets align with the sources reported in the article.

Figure 5: The Number of Unique Outlets Cited in Top-5 Slots



Notes: Bars represent the number of unique outlets Yandex cites in top-5 slots that week. Blue color counts the outlets that were also cited last week; yellow color counts the outlets that were not cited last week. The red vertical line corresponds to June 23, 2016, when the law on news aggregators was signed. The blue vertical line corresponds to January 1, 2017, when the law on news aggregators came into effect.

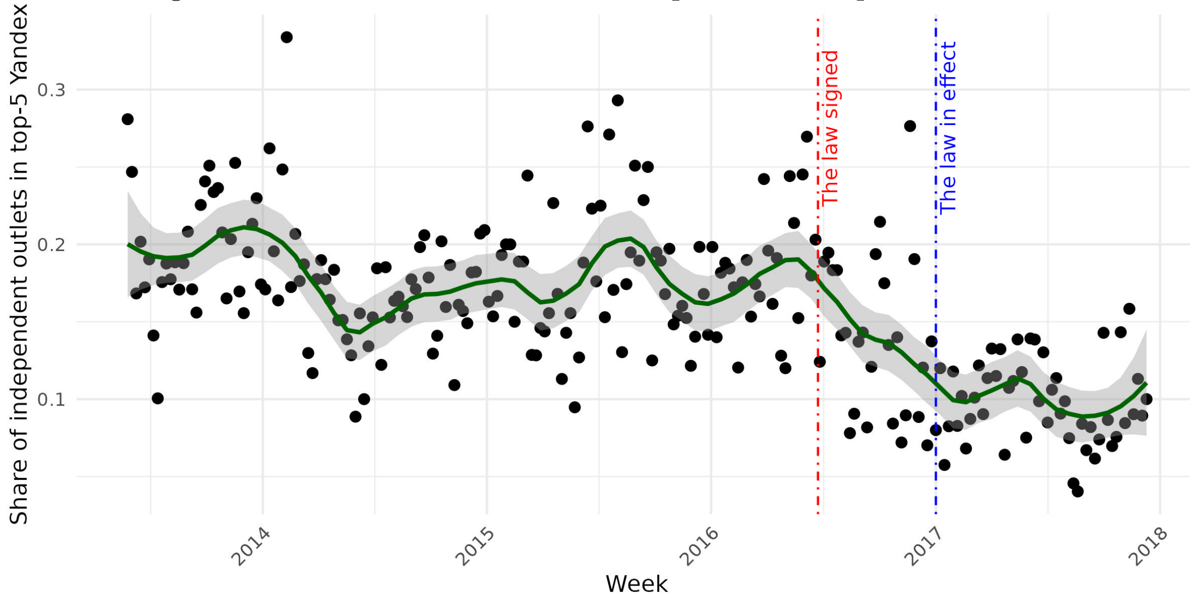
Citing fewer outlets resulted in citing the same outlet more often. The average number of slots held by the most-cited outlet at any moment rose to 2.13 after the law, from the 1.84 documented in Section 5.1—more than two of the five front-page slots now routinely went to a single outlet (Appendix Figure A6).

The reduction in the number of cited outlets mostly affected independent outlets. Yandex stopped citing prominent independents that had appeared regularly before the change, including Forbes Russia, Novaya Gazeta (headed by the future Nobel Prize winner Dmitry Muratov), and TV Rain; only two of the ten remaining outlets were independent, and both specialized in business news (RBC and Vedomosti).[15] The share of top-5 citations going to independent outlets accordingly fell from around 20% before the law—close to the independents' share of all news produced—to around 10% after it (Figure 6).

The removal of independent outlets did not, however, change which events Yandex covered. Figure 7 plots the share of sensitive events—those systematically under-reported by government-controlled outlets—in both the overall news flow and Yandex's top-5. Throughout the sample, including after the law, the share of sensitive events among Yandex's citations tracks the market

[15] Our classification is based on the situation at the beginning of the sample period. Coincidentally, RBC changed the main editors, who were considered too independent, in May 2016 and changed ownership in 2017; Vedomosti changed ownership and the editorial team in 2020.

Figure 6: Share of news in Yandex's top-5 from independent outlets



Notes: Each dot corresponds to an average share of independent news in Yandex's top-5 news on a given week. The red vertical line corresponds to June 23, 2016, when the law on news aggregators was signed. The blue vertical line corresponds to January 1, 2017, when the law on news aggregators came into effect.

average. Thus, the results indicate that the change affected the outlets Yandex referenced, but not the events it selected: it suppressed independent *outlets* without visibly suppressing sensitive *topics*.

6.2 Changes in the Algorithm

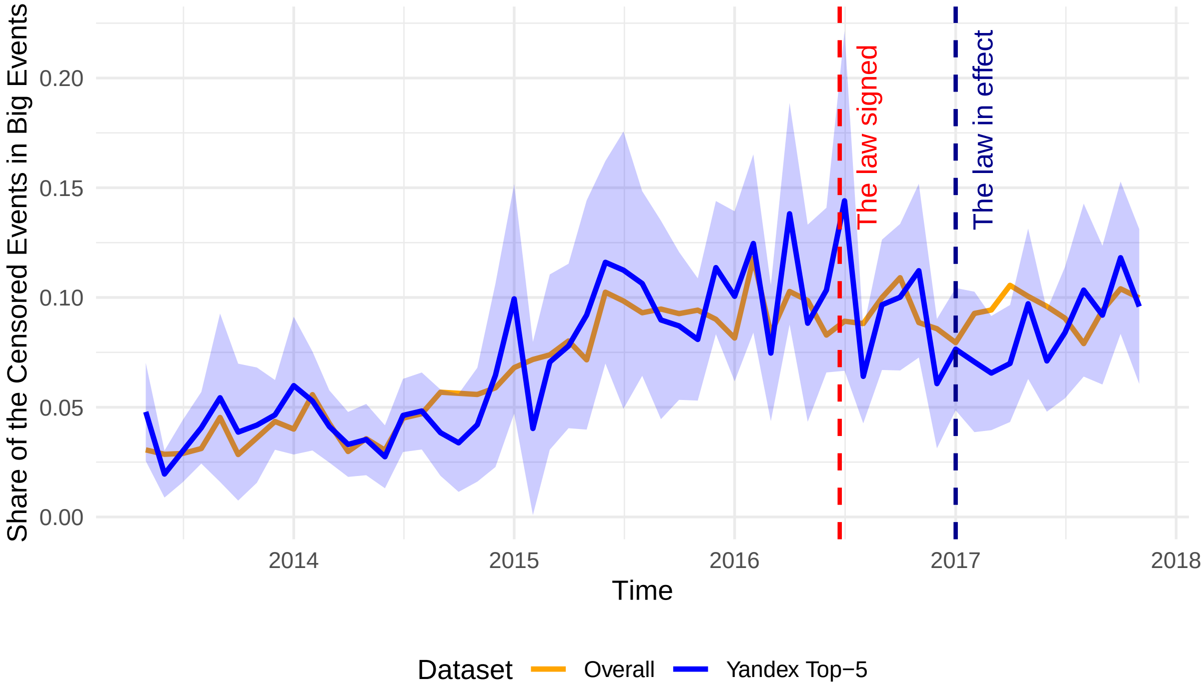
Yandex's discretion over which outlets it cites suggests it can steer traffic to outlets of its choosing. To pin down that discretion and to ground the counterfactuals that follow, we characterize Yandex's news-aggregation algorithm and trace how it changed after the law. Yandex has not published the exact algorithm but has disclosed its general design and main inputs.[16] It proceeds in two steps. The algorithm first ranks news events—inferred from clusters of arriving articles—by their newsworthiness, then chooses one outlet's article to cite for each of the top events.

6.2.1 First Step: Ranking of News Events

In each period t, a set of news events E_t has been realized, and each event e has a newsworthiness n_e that the researcher does not observe. A news outlet j chooses which events to report, given its preferences over topics and the events' newsworthiness. Yandex tries to select the five events

[16]The description can be found on the blog of Yandex (Yandex, 2014a). In particular, we use the same parser that Yandex uses to analyze texts (Yandex, 2014b).

Figure 7: Share of Sensitive Events: Overall News Flow vs. Yandex's Top-5



Notes: Each point is a monthly average share of *sensitive* events—those systematically under-reported by government-controlled outlets relative to independent ones—computed for the overall news flow and for the events cited in Yandex's top-5. Shaded areas are 95% confidence intervals, computed from the variation of daily shares in that month. We restrict the sample to events covered by at least 10 outlets. The red vertical line corresponds to June 23, 2016, when the law on news aggregators was signed; the blue vertical line to January 1, 2017, when it came into effect.

with the highest n_e, but it cannot see n_e directly; instead it infers an event's importance from how outlets cover it. We model the inferred score as

$$s_{et} = \boldsymbol{X'_{et}}\boldsymbol{\beta} + \epsilon_{et} \tag{2}$$

where X_{et} collects characteristics of event e's coverage at t—the number of articles and outlets covering it, the breadth of that coverage, and the speed with which it develops—and ϵ_{et} is an idiosyncratic shock to the event's score.

We estimate $\boldsymbol{\beta}$ from the realized top-5 rankings and the underlying set of detected events. Table 5 reports the results. Events covered by more articles, by more outlets, and with a larger share of their coverage in the first hour (a proxy for speed) are more likely to be cited; these three variables alone explain 16.8% of the variation in Yandex's citation decisions, while day (column 2) and topic (column 3) fixed effects add little.

The results demonstrate that the algorithm treated sensitive events differently, and did so even before the law (columns 4–5 of Table 5). Holding coverage constant, a sensitive event was

Table 5: Factors Correlated with Event Citations by Yandex

Dependent Variable:	I(Event in Yandex Top-5)				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Constant	-0.0404***				
	(0.0008)				
# of outlets	0.0134***	0.0133***	0.0133***	0.0136***	
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	
# of articles	0.0013***	0.0013***	0.0013***	0.0013***	0.0011***
	(7.33×10^{-5})	(7.33×10^{-5})	(7.32×10^{-5})	(7.33×10^{-5})	(6.49×10^{-5})
Share of coverage in first hour	0.0148***	0.0146***	0.0127***	0.0129***	0.0138***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Sensitive event				-0.0584***	-0.0691***
				(0.0057)	(0.0056)
Sensitive event x After law				-0.0267***	-0.0035
				(0.0097)	(0.0088)
Fixed-effects					
Day		Yes	Yes	Yes	Yes
Topic			Yes	Yes	Yes
Outlet					Yes
Fit statistics					
Observations	1,794,010	1,794,010	1,794,010	1,794,010	1,794,010
R^2	0.16835	0.17719	0.17861	0.17965	0.20354
Within R^2		0.16850	0.16479	0.16586	0.19015

Notes: The unit of observation is a news event; linear probability model. Clustered (week) standard-errors in parentheses. News topics correspond to thirty news topics. A sensitive event is one systematically under-reported by government-controlled outlets relative to independent ones (Section 4). Column 5 includes fixed effects for each outlet that published articles on the event; the number of covering outlets is absorbed by these fixed effects. The estimation sample includes all detected events starting on days with at least ten Yandex top-5 placements in our snapshots. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1.

5.8 percentage points less likely to reach the top-5; after the law this gap widened by a further 2.7 percentage points (column 4). That post-law widening of the gap, however, is entirely absorbed by outlet fixed effects (column 5). The pattern matches Figure 7: Yandex already under-weighted sensitive events before the law, and the additional decline afterward came from dropping independent outlets, not from filtering sensitive topics at the event stage.

6.2.2 Second Step: References of News Outlets

Having fixed the top-5 events, Yandex chooses one article—and hence one outlet—to cite for each. For an event e, a subset of outlets J_e have covered it, possibly with several articles each, and the algorithm selects among their articles using features of the coverage. Yandex's public description and our interviews with editors who worked at outlets at the time point to the same features: how heavily the outlet covered the event, how quickly it published relative to the first article on the story, how many distinct facts the article contains, properties of its title (length, punctuation, and fit to the event), and how much it copies the original source, with heavy copying penalized. On top of these features, Yandex can directly change the weights it assigns to different outlets, raising or lowering their chance of being cited.

We estimate the role of these features with

$$c_{je} = \tilde{\boldsymbol{X}}'_{\boldsymbol{ej}}\tilde{\boldsymbol{\beta}} + \gamma_e + \alpha_j + \tilde{\epsilon}_{ej} \tag{3}$$

where c_{je} indicates whether outlet j's article on event e reached the top-5, $\tilde{\boldsymbol{X}}_{\boldsymbol{ej}}$ collects the coverage features above, γ_e and α_j are event and outlet fixed effects, and $\tilde{\epsilon}_{ej}$ is an idiosyncratic shock.

The estimates in Table 6 match the practitioners' account. With event fixed effects absorbing newsworthiness, an outlet is more likely to be cited when it covers the event early, packs in more named entities (a proxy for the number of facts),[17] writes a shorter title with less punctuation, keeps the title close to the core of the event, and differentiates its text from the first article on the story.[18] Editors told us that Yandex penalizes articles copying 40% or more of the first article's text; splitting text distance into quintiles (column 2) confirms it, as the most-similar quintile carries a large negative penalty. Adding outlet fixed effects (column 3) leaves these relationships intact—if anything they sharpen—but raises the R^2 from 8.5% to 16.4%. Coverage features matter, but outlet weights matter at least as much.

To isolate those weights and track how they changed after the adoption of the law, we estimate a multinomial logit of the citation decision, which also accounts directly for the fact that each

[17] Beyond the count of named entities, article length itself does not raise citation probability; conditional on the number of named entities, longer articles are slightly less likely to be cited, with a precisely estimated but economically small coefficient.

[18] We score each article from 0 to 1 as one minus the cosine similarity of its text to that of the first article about the event, weighting words by their TF-IDF scores; the first article scores 0. Higher values mean less copying.

Table 6: Factors Correlated with Outlet References by Yandex

Dependent Variable:	I(Outlet j in Yandex Top-5)		
Model:	(1)	(2)	(3)
Variables			
Share of articles on this event by j	-0.1611*	-0.1606*	0.0895***
	(0.0890)	(0.0889)	(0.0323)
log(hours published after first event article)	-0.0165***	-0.0160***	-0.0039**
	(0.0040)	(0.0040)	(0.0015)
# of words in event articles by j	$-2.43 \times 10^{-5**}$	$-2.45 \times 10^{-5**}$	$-2.02 \times 10^{-5***}$
	(1.07×10^{-5})	(1.07×10^{-5})	(5.61×10^{-6})
# of named entities in event articles by j	0.0019***	0.0019***	0.0004***
	(0.0006)	(0.0006)	(0.0001)
Title length (symbols)	-0.0006**	-0.0006**	-0.0007***
	(0.0003)	(0.0003)	(0.0002)
# of punctuation marks in the title	-0.0049**	-0.0049**	-0.0025**
	(0.0019)	(0.0019)	(0.0010)
Title-to-text distance	-0.0964***	-0.0944***	-0.0247***
	(0.0260)	(0.0261)	(0.0078)
I(j published the earliest article)	0.0828***	0.0776***	0.0424***
	(0.0295)	(0.0272)	(0.0085)
Difference from the first publication	0.0724***		
	(0.0162)		
Difference from the first publication: 0-20%		-0.0265***	-0.0090*
		(0.0068)	(0.0046)
Difference from the first publication: 20-40%		-0.0120***	-0.0034***
		(0.0032)	(0.0011)
Difference from the first publication: 60-80%		0.0103***	0.0025*
		(0.0032)	(0.0013)
Difference from the first publication: 80-100%		0.0140***	0.0035
		(0.0048)	(0.0021)
Fixed-effects			
Event	Yes	Yes	Yes
Outlet			Yes
Fit statistics			
Observations	778,348	778,348	778,348
R^2	0.08466	0.08452	0.16423
Within R^2	0.01682	0.01667	0.00396

Notes: Linear probability model. Observations correspond to news articles about the events that were cited by Yandex in top-5 slots at least once. "Difference from the first publication" is the textual distance of the article from the first article on the event (see footnote in Section 6.2); the omitted category in columns 2–3 is the 40–60% quintile. Title length is in symbols. Standard errors clustered on the outlet level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1.

outlet covers only a subset of events:

$$Pr(\text{Outlet } j \text{ in Yandex Top-5}_e \mid j \in J_e) = \frac{\exp\Big(\alpha_j + \alpha'_j I(\text{post-law})_e + \tilde{X}'_{ej}\tilde{\beta}\Big)}{\sum_{j' \in J_e} \exp\Big(\alpha_{j'} + \alpha'_{j'} I(\text{post-law})_e + \tilde{X}'_{ej'}\tilde{\beta}\Big)} \tag{4}$$

where α_j is outlet j's weight, α'_j is the change in that weight after the law, and $\tilde{\beta}$ are the controls from equation (3).

Table 7: Implied Probabilities of Citation Based on Outlet Weights: By Outlet Type

	Probabilities of Citation (in %)			
Type	Equal Weights	Before the Law	After the Law	Difference
International	10.20	0.04 (0.03)	0.01 (0.00)	-0.03 (0.03)
Independent	26.53	28.89 (0.43)	15.90 (0.36)	-12.99 (0.57)
Potentially-Influenced	42.86	30.85 (0.37)	37.11 (0.47)	6.26 (0.65)
Government-Controlled	20.41	40.23 (0.35)	46.99 (0.50)	6.76 (0.64)

Notes: Implied probabilities of references based on outlet weights α_j and α'_j from equation (4). The probabilities are first computed for each outlet and then summed up by outlet type. The "equal weights" column is proportional to the number of outlets of each type in the sample. One of the 50 outlets is excluded because it has no pre-law coverage in the sample.

Table 7 translates the estimated weights into implied citation probabilities by outlet type, holding coverage features constant; Table A20 in Appendix A7 provides the corresponding outlet-level figures. Column 1 is the benchmark in which Yandex weights all outlets equally. Before the law (column 2), independent outlets carried weights roughly proportional to their market share, potentially-influenced and international outlets were cited below their share, and government-controlled outlets carried the highest weights, with a 40% implied citation probability.

The weights moved sharply after the law. Independent outlets, unpenalized before, changed the most: their implied citation probability fell from 28.9% to 15.9%, as several independents received strongly negative post-law weights that drove their implied probability to zero.[19] These estimates confirm the descriptive pattern in Figures 5 and 6: independents lost citations because Yandex lowered their second-stage weights, not because of the events they chose to cover or anything else in their coverage.[20] An aggregator can re-weight outlets seamlessly—and we now ask how much that discretion is worth.

6.3 Changes in Effectiveness

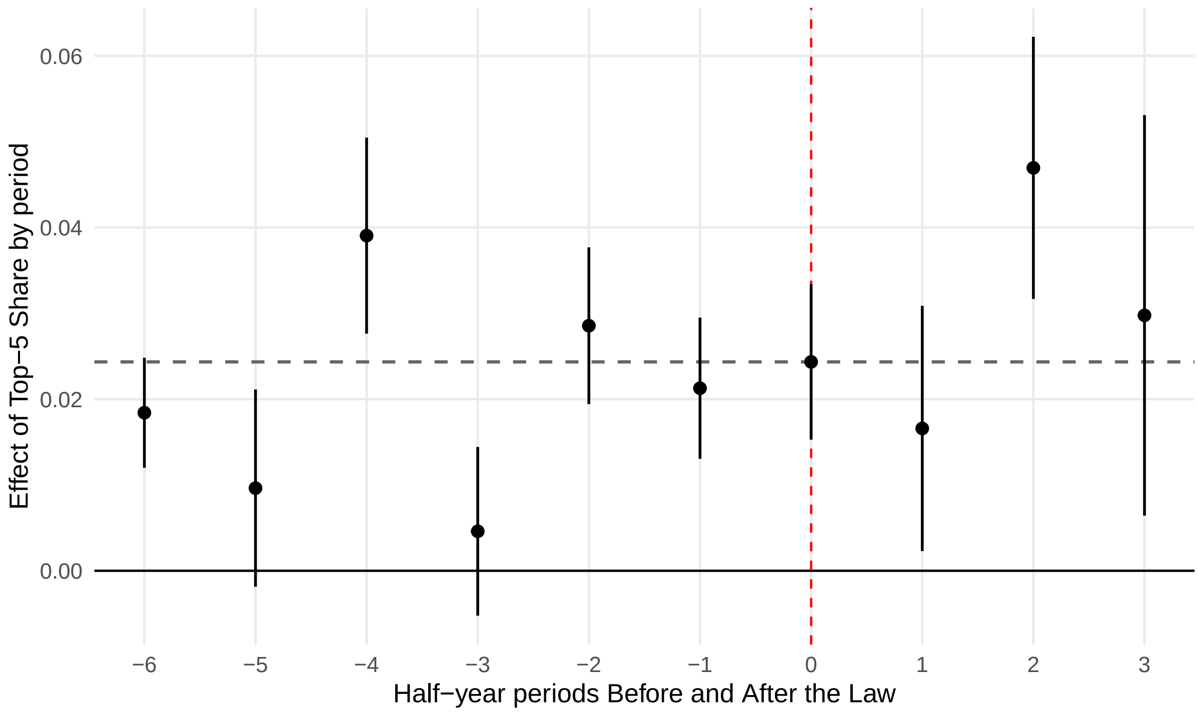
If consumers punished Yandex for a less informative ranking, the narrowing of outlets should have eroded the top-5's pull: users would learn to discount its recommendations, and a given

[19]See Table A20 in Appendix A7 for outlet-level results.

[20]Consistent with such a binary, outlet-level switch, a journalistic investigation of leaked Yandex News code reported a per-source flag governing whether an outlet could appear on the front page (The-Village, 2022).

referral would influence fewer of them. We test this by estimating Yandex's effectiveness—the normalized effect of equation (1), introduced in Section 5.2—over six-month windows around the law, during which Yandex cut the number of cited outlets from around 24 to 10 (and later to 15).

Figure 8: Effectiveness of Outlet References in Yandex's Top-5: Before and After the Law



Notes: Points correspond to the α estimates from regression (1) with y_{jd} normalized by the volume of traffic arriving on `yandex.ru` on day d, estimated over six-month windows. Error bars correspond to the 95% confidence intervals.

Figure 8 shows that there was no evidence that a change in the algorithm reduced the influence of Yandex. A full day in all five top-5 slots sends an outlet about 2% of Yandex's front-page visitors throughout, and if anything, effectiveness is slightly higher in the six months after the law than before. Cutting the number of cited outlets from 24 to 10 did not weaken Yandex's ability to direct traffic. There is also no evidence that users were leaving the platform: Yandex's search market share was flat or slightly higher after the law on both desktop and mobile (Appendix Figures A2 and A3). Overall, we see no evidence that the demand side poses any constraint on the ability of Yandex to change its algorithm.

6.4 Yandex's Power over Outlet Traffic

Because demand does not constrain Yandex's choice of outlets or events, we can ask how far it could move outlets' traffic by manipulating the weights it assigns to different outlets. We use the level estimates of equation (1) to simulate the traffic each outlet would receive under counterfactual citation rules, for the 11 outlets with traffic data.

We start with the most permissive rule: Yandex faces no restriction on which events to cite or which outlets to reference, and devotes all five top-5 slots to a single outlet. The left side of Figure 9 splits each outlet's traffic into what it earns without Yandex, what it earns from Yandex today, and what it would earn under full preferencing. Today the top-5 supplies between almost nothing (`fontanka.ru`) and 5.89% (`gazeta.ru`) of an outlet's traffic; under full preferencing Yandex could raise an outlet's traffic by between 26% (for `ria.ru`, the largest outlet here) and 1,132% (for `trud.ru`, the smallest).

The opposite rule does not allow Yandex to change the first stage of its algorithm: it must cite the same events, dictated by their newsworthiness, and can only choose which covering outlet to reference. Its power to steer traffic is now bounded by which outlets cover the required events. Re-simulating under this constraint (the right side of Figure 9), Yandex could still raise an outlet's traffic by between 15% (`ria.ru`) and 270% (`trud.ru`). The constraint binds hardest on specialized outlets such as `echo.msk.ru`, which cover fewer of the events that reach the top-5.

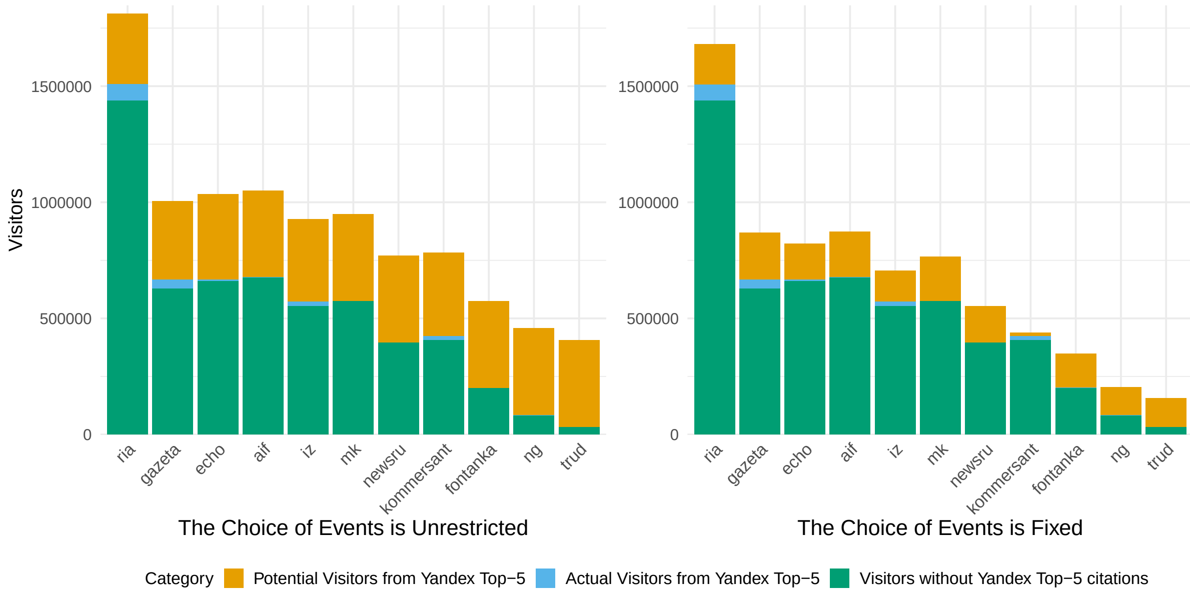
Two assumptions underlie these counterfactuals. First, they extrapolate a linear per-slot effect to occupancy levels far above those observed: the most-cited outlet holds 1.84–2.13 of the five slots on average, so occupying all five slots for the whole day lies well outside the support of the data. Second, the effect is identified from transitory within-week variation; the absence of dynamic effects over the following two weeks (Figure 4) supports treating daily referrals as additive, but sustained preferencing over months could change reader behavior in ways short-run variation cannot reveal. The counterfactuals are therefore best read as bounds on Yandex's ability to redirect attention under its 2013–2017 design, not as forecasts of a permanent regime. In addition, because the five slots are a fixed resource, traffic directed to one outlet is traffic withheld from others; the outlet-level counterfactuals should be read one at a time, not summed across outlets.

Finally, we evaluate the actual 2016 change rather than a hypothetical one. Because our traffic tracker does not cover all 50 outlets before 2017, we apply the treatment estimates from Section 5.2 to 2017 traffic levels from SEMRush.[21] The individual reallocations are large (Figure 10). The independent outlet `forbes.ru` lost the most, around 18% of its traffic, while the potentially-influenced `iz.ru` gained around 25%; in levels (Subfigure (b) of Figure 10), the biggest winner was the government-owned agency `ria.ru`, with roughly 64,000 additional daily visitors.

In the aggregate, the results indicate that the change in the algorithm moved readership from independent toward government-aligned sources, although the magnitude of the effect was moderate. Summing across all 50 outlets, it cut traffic to independent outlets by 0.9% and raised traffic to government-controlled and potentially-influenced outlets by 0.5% and 1.4%, respectively.

[21] As mentioned in Section 3, the SEMRush dataset does not contain traffic measures before 2017.

Figure 9: Counterfactual Traffic to News Outlets



Notes: Each bar decomposes an outlet's predicted daily traffic into three parts: the traffic it would receive with no Yandex top-5 referrals, the additional traffic it receives from Yandex under the current algorithm, and the further traffic it would receive if Yandex devoted all five top-5 slots to it. The left panel shows the unrestricted counterfactual, in which Yandex may choose both events and outlets. The right panel shows the fixed-event counterfactual, in which Yandex must cite the same events and may choose only among outlets covering each event. Predictions use the level estimates of regression (1).

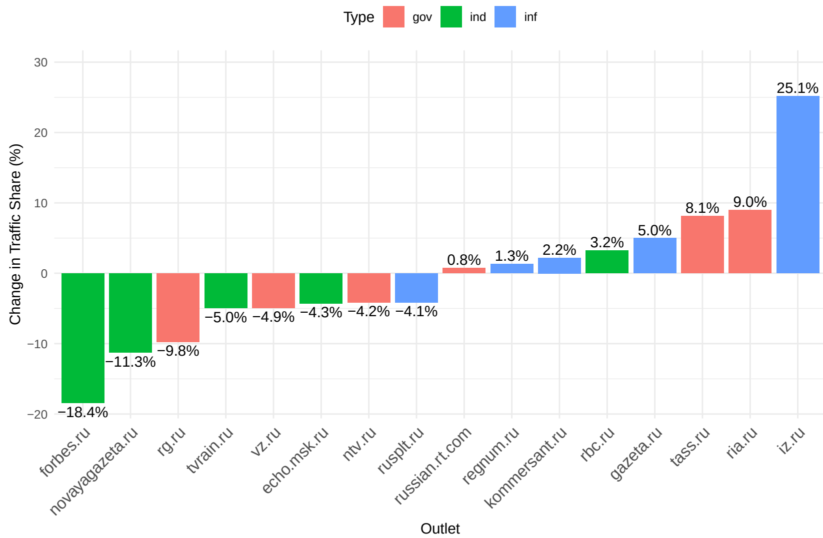
6.5 Effects on News Production

Given how much traffic the top-5 delivers, outlets have a clear incentive to shape their coverage to be cited, and they openly admit they do. One editor at a major Russian news wire put it bluntly: "We're not writing news for the readers. We are writing them for Yandex" (Kovalev, 2021). In our interviews, editors working at the time described—accurately—how Yandex weighs title structure, the number of distinct facts, and the degree to which an article copies its source, and explained how they adjusted their output accordingly.

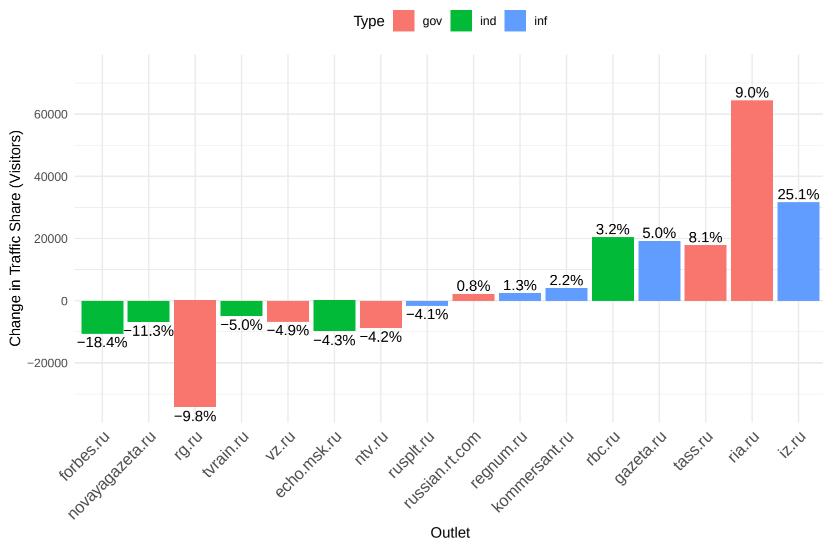
If outlets optimize toward the algorithm, a change in the algorithm should change what they produce, since they are likely to target a different audience. Instead of going for casual readers that are attracted by Yandex citations, they may focus on loyal regular readers who visit their webpage directly. We test this using the September 2016 change, which dropped 9 of the 24 regularly cited outlets for good (Section 6.1), including frequently cited independents such as Forbes and Novaya Gazeta. We compare the production decisions of these 9 removed outlets to those of the 15 Yandex kept, in an event study around the change, weighting outlets by their pre-change citation probability.

We look at two consequential margins and two cosmetic ones. The most consequential margin is whether an outlet writes short "spot" news on widely covered events—which stand a chance

Figure 10: Changes in Traffic to News Outlets due to Algorithm Change



Subfigure (a): Traffic in Shares (%)

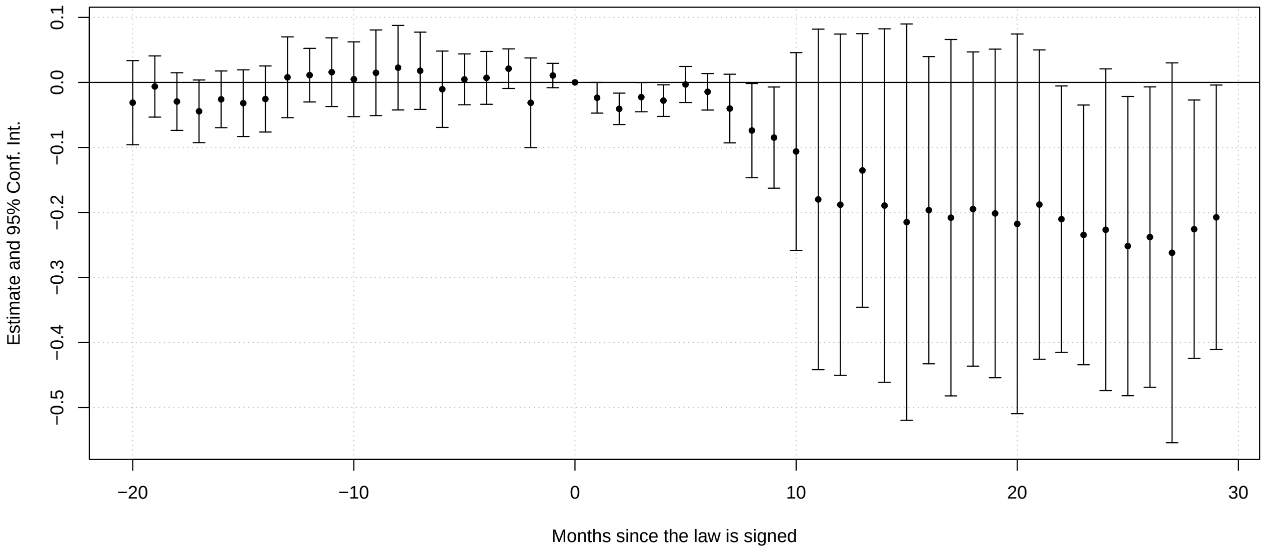


Subfigure (b): Traffic in Levels

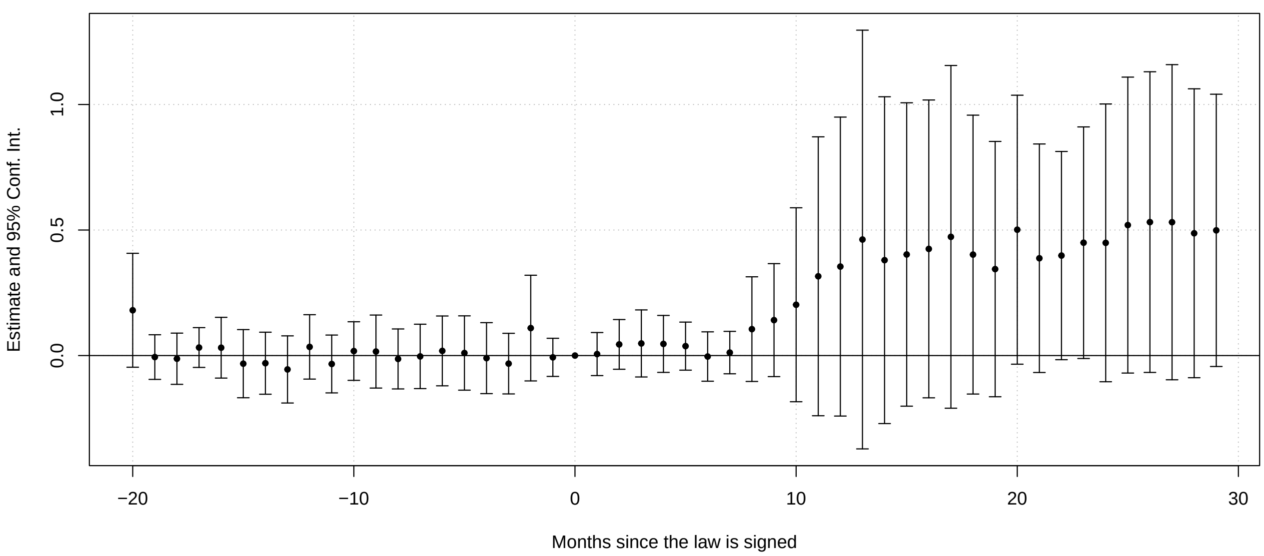
Notes: The figure shows the eight outlets with the largest predicted traffic gains and the eight with the largest losses from the 2016 algorithm change. Subfigure (a) reports the change as a share of each outlet's traffic; subfigure (b), in levels (daily visitors). Predicted changes apply the estimates of regression (1) to baseline 2017 traffic from the SEMRush dataset.

of citation—or differentiates toward longer "feature" articles on narrower topics. Another consequential margin is if outlets start writing longer articles. The cosmetic margins are title length and title punctuation, which matter only for the algorithm. Relative to the outlets still cited, the removed outlets became 20% less likely to write articles that cluster into events and wrote 50% longer articles (Figure 11), consistent with a shift toward feature writing. They also stopped

Figure 11: Impact of Yandex's Algorithm Change on News Production: Style of News



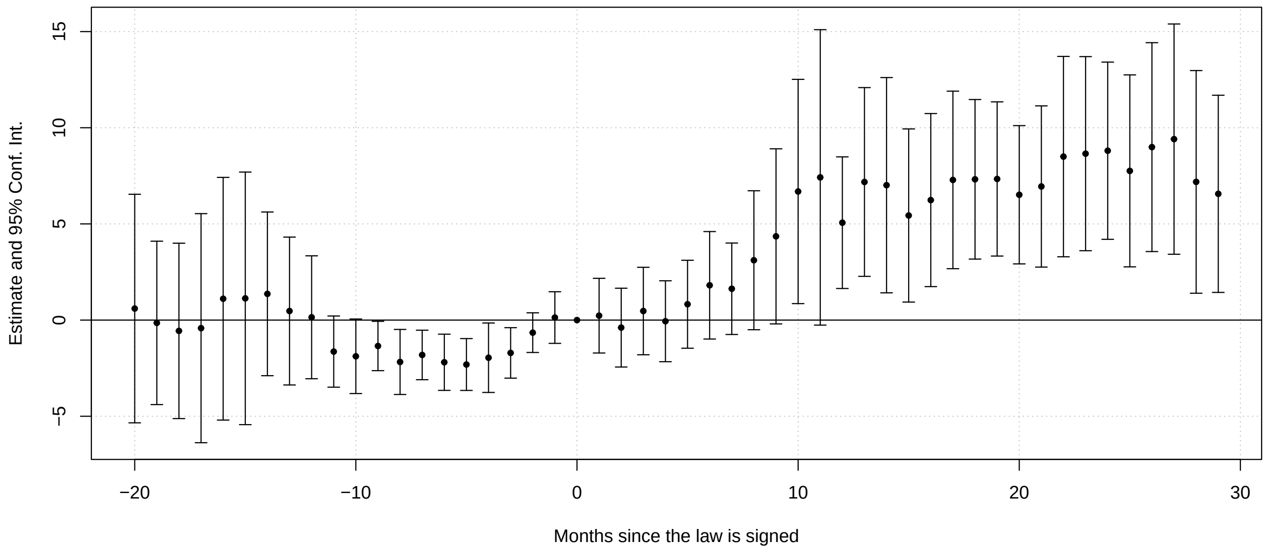
(a) Share of News Classified into Events



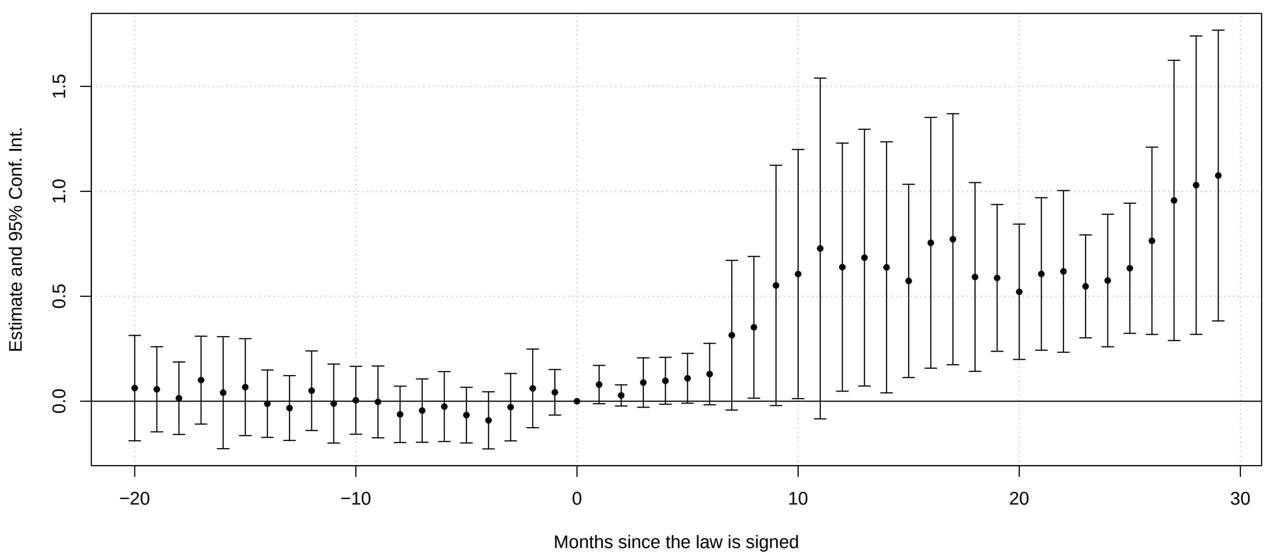
(b) Average Number of Words

Notes: Estimates from an event study, where a change in algorithm serves as an event, news outlets that are removed from the top-5 news serve as a treatment group, and the rest of outlets serve as a control group. We control for outlet-by-day-of-the-week and day fixed effects. Outlets are weighted by the pre-treatment probability of being referenced by Yandex in the top-5 section. Error bars correspond to the 95% confidence intervals. Standard errors are clustered at the outlet level.

Figure 12: Impact of Yandex's Algorithm Change on News Production: Features of Titles



(a) # of Symbols in a Title



(b) # of Punctuation Symbols in a Title

Notes: Estimates from an event study, where a change in algorithm serves as an event, news outlets that are removed from the top-5 news serve as a treatment group, and the rest of outlets serve as a control group. We control for outlet-by-day-of-the-week and day fixed effects. Outlets are weighted by the pre-treatment probability of being referenced by Yandex in the top-5 section. Error bars correspond to the 95% confidence intervals. Standard errors are clustered at the outlet level.

grooming their titles: title length rose by 7–8 symbols and punctuation by 0.5–1 marks (Figure 12). The adjustment took about 10 months to complete.[22]

These shifts show that an aggregator's algorithm affects not only the marginal users who are directed to the outlets through Yandex, but has important indirect spillover effects. When a removed outlet starts writing differently, the change affects all of its readers—including the inframarginal ones who never arrived through Yandex and now receive different coverage from a source they already preferred.

7 Conclusion

We provide the first large-scale estimates of how much a news aggregator can shape the flow of news through the algorithm that ranks stories and sources. Using the example of the leading news aggregator in Russia, Yandex News, in 2013–2017, and treating the 2016 liability law as a natural experiment, we produce three major findings. First, news aggregators have a very strong effect on traffic to outlets. A day in all slots atop Yandex's front page sends an outlet roughly 380,000 additional visitors, an effect that varies little with the topics or events cited. Second, the algorithm that determines citations by the news aggregator can be changed at will and its effectiveness is not constrained by demand factors. When the law pushed Yandex to drastically reduce its referrals to independent outlets, readers kept following its recommendations as before the algorithm change. Finally, the news aggregator has an upstream effect on the content produced by media outlets. The outlets that were excluded entirely from the news aggregator adjusted their news production, differentiating towards longer articles and optimizing less on features that the algorithm had rewarded. That change reached even the outlets' loyal readers, who never arrived through Yandex at all.

One of the most important findings is that the demand side that disciplines traditional media—viewers switch away in response to changes in content (Durante and Knight, 2012; Simonov and Rao, 2022)—did not discipline the news aggregator in our context. Freed from that discipline, control over distribution of news—even without any control over production—can become an instrument of media capture (Prat, 2018). Although the extreme result of no demand constraint may not hold in more competitive environments, we suspect that this constraint is likely to be quite loose in other contexts as well.

More broadly, we expect similar forces to give other platforms—social media, search engines, and, increasingly, AI assistants—comparable power to shape what people read without producing

[22] One removed outlet, `forbes.ru`, also changed owners about a year before the algorithm change, in September 2015; the new owner favored an editorial line less focused on political news (e.g., Surganova, 2015), which Forbes eventually adopted after being censored by Yandex. Given that Forbes was an outlet that lost the most because of Yandex's algorithm change (Figure 10), we find it plausible that the eventual decision to change the editorial line was impacted by the loss of Yandex's referral traffic.

any content, often by personalizing what each user sees rather than ranking a single public list (e.g. Braghieri et al., 2025).[23] Through that control over the distribution of news, these platforms ultimately influence not only what their users read but what they come to believe, which may have far-reaching political and social consequences (DellaVigna and Gentzkow, 2010; Enikolopov and Petrova, 2015).

References

Allan, J., Carbonell, J., Doddington, G., Yamron, J., Yang, Y., et al. (1998). Topic detection and tracking pilot study: Final report. In *Proceedings of the DARPA broadcast news transcription and understanding workshop*, volume 1998, pages 194–218. Citeseer.

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., and Lalmas, M. (2020). Algorithmic effects on the diversity of consumption on spotify. In *Proceedings of the web conference 2020*, pages 2155–2165.

Aridor, G., Gonçalves, D., Kluver, D., Kong, R., and Konstan, J. (2022). The economics of recommender systems: Evidence from a field experiment on movielens. *arXiv preprint arXiv:2211.14219*.

Athey, S., Mobius, M., and Pal, J. (2021). The impact of aggregators on internet news consumption. Technical report, National Bureau of Economic Research.

Besley, T. and Prat, A. (2006). Handcuffs for the grabbing hand? media capture and government accountability. *The American Economic Review*, 96(3):720–736.

Boxell, L. and Conway, J. (2022). Journalist ideology and the production of news: Evidence from movers. *Available at SSRN 4295587*.

Braghieri, L., Eichmeyer, S., Levy, R., Mobius, M., Steinhardt, J., and Zhong, R. (2025). Article level slant and polarization of news consumption on social media. *Working Paper*.

Cagé, J., Hervé, N., and Mazoyer, B. (2022). Social media and newsroom production decisions. *Working Paper*.

Cagé, J., Hervé, N., and Viaud, M.-L. (2020). The production of information in an online world. *The Review of economic studies*, 87(5):2126–2164.

[23] Nor is the power of algorithmic distribution confined to authoritarian settings: in the United States, the Trump White House has published government content that appears in Google News alongside independent reporting, an arrangement that can lend official messaging the credibility of journalism (Mehta and Robinson, 2025).

Calzada, J., Duch-Brown, N., and Gil, R. (2023). Do search engines increase concentration in media markets? *CESifo Working Paper*.

Calzada, J. and Gil, R. (2020). What do news aggregators do? evidence from google news in spain and germany. *Marketing Science*, 39(1):134–167.

Chiou, L. and Tucker, C. (2017). Content aggregation by platforms: The case of the news media. *Journal of Economics & Management Strategy*, 26(4):782–805.

Dellarocas, C., Katona, Z., and Rand, W. (2013). Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management science*, 59(10):2360–2379.

DellaVigna, S. and Gentzkow, M. (2010). Persuasion: empirical evidence. *Annu. Rev. Econ.*, 2(1):643–669.

Djankov, S., McLiesh, C., Nenova, T., and Shleifer, A. (2003). Who owns the media? *The Journal of Law and Economics*, 46(2):341–382.

Dujeancourt, E. and Garz, M. (2023). The effects of algorithmic content selection on user engagement with news on twitter. *The Information Society*, 39(5):263–281.

Durante, R. and Knight, B. (2012). Partisan control, media bias, and viewer responses: Evidence from berlusconi's italy. *Journal of the European Economic Association*, 10(3):451–481.

Enikolopov, R., Makarin, A., and Petrova, M. (2020). Social media and protest participation: Evidence from russia. *Econometrica*, 88(4):1479–1514.

Enikolopov, R. and Petrova, M. (2015). Media capture: Empirical evidence. In *Handbook of media economics*, volume 1, pages 687–700. Elsevier.

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and political persuasion: Evidence from russia. *American economic review*, 101(7):3253–3285.

Freimane, M. (2025). Substituting away? the effect of platform bargaining regulation on content display. *Woriking Paper*.

Garz, M. and Szucs, F. (2023). Algorithmic selection and supply of political news on facebook. *Information Economics and Policy*, 62:101020.

Gehlbach, S. and Sonin, K. (2014). Government control of the media. *Journal of Public Economics*, 118:163–171.

Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.

George, L. M. and Hogendorn, C. (2020). Local news online: Aggregators, geo-targeting and the market for local news. *The Journal of Industrial Economics*, 68(4):780–818.

Google (2018). Introducing Google Discover: New information and inspiration with Search, no query required. https://blog.google/products/search/introducing-google-discover/ [Accessed: July 7, 2026].

Google (2026). Change your Google notifications. https://support.google.com/websearch/answer/9919653 [Accessed: July 18, 2026].

Guriev, S. and Treisman, D. (2019). Informational autocrats. *Journal of Economic Perspectives*, 33(4):100–127.

Hartog, E. (2017). How a new law is making it difficult for russia's news aggregators to tell what's going on. https://www.themoscowtimes.com/2017/04/07/how-a-new-law-is-making-it-difficult-for-russias-news-aggregators-to-tell-whats-going-on-a57657 [Accessed: May 30, 2026].

Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., and Aral, S. (2020). The engagement-diversity connection: Evidence from a field experiment on spotify. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 75–76.

Jeon, D.-S. and Nasr, N. (2016). News aggregators and competition among newspapers on the internet. *American Economic Journal: Microeconomics*, 8(4):91–114.

Knight, B. and Tribin, A. (2019). The limits of propaganda: Evidence from chavez's venezuela. *Journal of the European Economic Association*, 17(2):567–605.

Kovalev, A. (2021). The political economics of news making in russian media: Ownership, clickbait and censorship. *Journalism*, 22(12):2906–2918.

Mehta, D. and Robinson, C. (2025). The white house is its own media outlet. https://www.cjr.org/analysis/the-white-house-is-its-own-media-outlet.php [Accessed: May 30, 2026].

Moehring, A. (2024). Personalization, engagement, and content quality on social media: An evaluation of reddit's news feed. *Working Paper*.

Mozilla (2026). Recommended stories on the Firefox New Tab page. https://support.mozilla.org/en-US/kb/recommendations-firefox-new-tab [Accessed: July 18, 2026].

Our World in Data (2021). Number of internet users, by country. https://ourworldindata.org/grapher/number-of-internet-users [Accessed: July 7, 2026].

Peukert, C., Sen, A., and Claussen, J. (2023). The editor and the algorithm: Recommendation technology in online news. *Management science*.

Prat, A. (2018). Media power. *Journal of Political Economy*, 126(4):1747–1783.

Prat, A. and Strömberg, D. (2013). The political economy of mass media. *Advances in Economics and Econometrics*, Volume 2, Applied Economics.

Reimers, I. and Waldfogel, J. (2023). A framework for detection, measurement, and welfare analysis of platform bias. Technical report, National Bureau of Economic Research.

Reiter, S. (2022). "we are tired of fighting": How russia's invasion of ukraine split yandex. https://meduza.io/feature/2022/05/05/my-zamuchilis-borotsya [Accessed: May 31, 2026].

Reuters (2024). Overview and key findings of the 2024 digital news report. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/dnr-executive-summary [Accessed: December 10, 2025].

Roberts, M. E., Stewart, B. M., Tingley, D., Airoldi, E. M., et al. (2013). The structural topic model and applied social science. In *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*, volume 4, pages 1–20. Harrahs and Harveys, Lake Tahoe.

Rudkovskiy, A. (2026). National news platform based on "zen": What vk and the government have planned. https://www1.ru/en/news/2026/04/14/nacionalnaia-novostnaia-platforma-na-baze-dzena-cto-zadumali-vk-i-pravitelstvo.html [Accessed: May 30, 2026].

Schwartz, B. (2025). Google to expand Discover to Google's desktop homepage. Search Engine Land, April 9, 2025. https://searchengineland.com/google-to-expand-discover-googles-desktop-homepage-454155 [Accessed: July 7, 2026].

Sen, A., Grad, T., Ferreira, P., and Claussen, J. (2023). (how) does user-generated content impact content generated by professionals? evidence from local news. *Management Science*.

Sen, A. and Yildirim, P. (2016). Clicks bias in editorial decisions: How does popularity shape online news coverage? *Working paper*.

Simonov, A. and Rao, J. (2022). Demand for online news under government control: Evidence from russia. *Journal of Political Economy*, 130(2):259–309.

Sismeiro, C. and Mahmood, A. (2018). Competitive vs. complementary effects in online social networks and news consumption: A natural experiment. *Management Science*, 64(11):5014–5037.

StatCounter (2026). Search engine market share by country, may 2026. https://gs.statcounter.com/search-engine-market-share [Accessed: July 7, 2026].

Statista (2023). From which search engines do you usually read news, information reports? https://www.statista.com/statistics/1102121/russia-most-popular-search-engines-for-news/ [Accessed: March 14, 2023].

Surganova, E. (2015). The new owner of forbes to rbc: “we will try not to get involved in politics”. https://www.rbc.ru/interview/technology_and_media/16/10/2015/561fd53e9a794751f652b1d5 [Accessed: May 31, 2026].

The-Village (2022). How we hacked “yandex”, to find censorship in it [in russian]. https://www.the-village.ru/city/eksklyuziv/rassledovanie-yandex [Accessed: March 16, 2023].

Tokar, N. (2015). Yandex from the inside: How many people use yandex, or how the company lost a third of its audience in two years. https://tokar.ua/read/9412/iandeks-yznutry-statystyka-yly-kak-kom/ [Accessed: May 31, 2026].

Vedomosti (2016). Google news will work in russia without restrictions [in russian]. https://www.vedomosti.ru/technology/articles/2016/06/15/645319-google-news [Accessed: May 30, 2026].

Wijermars, M. (2021). Russia’s law ’on news aggregators’: Control the news feed, control the news? *Journalism*, 22(12):2938–2954.

Yandex (2014a). How yandex.news works. https://yandex.ru/blog/company/76641 [Accessed: May 31, 2026].

Yandex (2014b). Tomita parser. https://github.com/yandex/tomita-parser/ [Accessed: May 31, 2026].

Yandex (2022). Yandex will have a new main page. what will change [in russian]. https://yandex.ru/blog/company/u-yandeksa-budet-novaya-glavnaya-chto-izmenitsya [Accessed: March 16, 2023].

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12:415–438.

A Appendix

A1 Worldwide Market Shares of Search Engines with Front-Page News

This appendix details the calculation behind the worldwide search-engine shares reported in Section 2.2.

Country-level search-engine market shares are from StatCounter (StatCounter, 2026) for May 2026. StatCounter measures search-engine referrals: clicks from a results page onto one of the more than one million websites in its tracking network. It therefore captures neither the number of searches nor the number of unique users directly; country-level shares are shares of tracked search referrals within that country. The number of internet users per country is from Our World in Data (Our World in Data, 2021), using the latest available year for each country (2021 for most). We match 208 countries and territories, covering roughly 4.8 billion internet users.

We compute the worldwide share of each search engine as the average of its country-level shares, weighted by each country's number of internet users. This weighting answers the question relevant for our purposes—what search environment the average internet user faces—rather than the composition of page views on StatCounter's tracked websites. Under this weighting, Google accounts for 69.6% of worldwide search activity. Among all engines that reach at least 1% of the market in at least one country, we check whether the engine's homepage carries a curated news block. Nine do—Baidu (10.2%), Bing (8.4%), Yandex (5.8%), Haosou (3.3%), Yahoo (1.0%), Naver (0.4%), Seznam, Mail.ru, and Daum (below 0.1% each)—and together they account for 29.3% of worldwide search activity, nearly all of the non-Google remainder. Engines without a news homepage (e.g., DuckDuckGo, Ecosia, CocCoc, Qwant, Sogou) account for the rest.

StatCounter's own headline worldwide statistic puts Google at about 90%. That figure is a deliberately unweighted aggregate of raw page views (StatCounter, 2026), so it over-represents regions where StatCounter-tracked websites are concentrated. Most importantly, StatCounter's network barely covers China—home to more than a fifth of the world's internet users, where Google is unavailable—so the unweighted aggregate effectively excludes it. Repeating our user-weighted calculation without China yields roughly 88%, consistent with the headline figure. The 70% and 90% figures thus measure different populations: Google's share including China at its user weight, and Google's share of the non-China internet, respectively.

A2 Construction of the Top-6 to Top-10 Positions

The main page of Yandex displayed only five news stories, so no "top-6" position exists there. We infer which stories nearly entered the top-5 from the Yandex News webpage (`news.yandex.ru`), which, in addition to the overall top-5, ranked the top stories within each topic subcategory.[24] We assign the leading stories of the subcategory rankings that do not appear on the front page to the "top-6" position; stories in the second subcategory slots to the "top-7" position; stories in the third slots to the "top-8" position; and so on through "top-10." Because `news.yandex.ru` attracts less traffic than `yandex.ru`, its snapshots on the Wayback Machine are sparser, and these variables are available for fewer days than the top-5 measures.

A3 Traffic Data Validation

There are four sources of daily traffic data of the Russian media market:

- `top.mail.ru` – a traffic counter which provides public data for 13 relevant outlets at some point since 2010; eleven of them keep their daily traffic open throughout our estimation window and form the traffic sample in the main text. We use it as the main source for our estimates.

- `semrush.com` – a service that approximates websites' daily audience using third-party data and its own algorithms. It provides data on all 50 outlets studied, available since 2017. We use this source to estimate the average number of visitors for all outlets.

- liveinternet.ru – a traffic counter from which we have collected data on a small subset of outlets since 2022.

- `top100.rambler.ru` – another traffic counter from which we have collected data since 2020.

We use `top.mail.ru` as the primary data source for our estimates, because it is the only one that provides data for periods earlier than 2017. In this section we use overlapping time periods and websites across different data sources to validate the quality of the data provided by `top.mail.ru`.

Table A1 presents the number of overlapping websites across different data sources, while Table A2 provides detailed information for each pair of sources.

[24] The set of subcategories varied somewhat during this period, but the main ones included Politics, Society, Economics, World News, Sports, Crime, Culture, Science, Technology, and Auto.

Table A1: Overlap of Websites Across Traffic Data Sources

	mail	semrush	liveinternet	rambler
mail	34	13	13	20
semrush	13	50	23	28
liveinternet	13	23	84	33
rambler	20	28	33	895

Table A2: Websites Shared Between Pairs of Data Sources

Data Sources	**Overlapping Websites**
mail & semrush (13)	aif.ru, echo.msk.ru, fontanka.ru, gazeta.ru, iz.ru, kommersant.ru, lenta.ru, mk.ru, newsru.com, ng.ru, rg.ru, ria.ru, trud.ru
mail & rambler (20)	9tv.co.il, aif.ru, argumentiru.com, echo.msk.ru, fedpress.ru, fontanka.ru, gazeta.ru, kommersant.ru, lenta.ru, m24.ru, newsru.co.il, newsru.com, ng.ru, pln-pskov.ru, portal-kultura.ru, ren.tv, rg.ru, ria.ru, tvzvezda.ru, ura.ru
mail & liveinternet (13)	aif.ru, echo.msk.ru, fontanka.ru, gazeta.ru, kommersant.ru, lenta.ru, m24.ru, mk.ru, ren.tv, rg.ru, ria.ru, tvzvezda.ru, ura.ru
semrush & rambler (28)	aif.ru, bfm.ru, echo.msk.ru, fontanka.ru, forbes.ru, gazeta.ru, kommersant.ru, kp.ru, lenta.ru, life.ru, newsru.com, newtimes.ru, ng.ru, ntv.ru, rbc.ru, regnum.ru, rg.ru, ria.ru, rosbalt.ru, rusplt.ru, russian.rt.com, snob.ru, sobesednik.ru, svpressa.ru, tass.ru, tvrain.ru, utro.ru, vesti.ru
semrush & liveinternet (23)	aif.ru, bfm.ru, dni.ru, echo.msk.ru, fontanka.ru, gazeta.ru, kommersant.ru, kp.ru, lenta.ru, mk.ru, novayagazeta.ru, rbc.ru, regnum.ru, rg.ru, ria.ru, ridus.ru, rosbalt.ru, svpressa.ru, tass.ru, tvrain.ru, utro.ru, vedomosti.ru, znak.com
liveinternet & rambler (33)	aif.ru, bfm.ru, echo.msk.ru, expert.ru, finmarket.ru, fontanka.ru, gazeta.ru, govoritmoskva.ru, inosmi.ru, kazanfirst.ru, kommersant.ru, kp.ru, lenta.ru, m24.ru, mir24.tv, moslenta.ru, rbc.ru, regnum.ru, ren.tv, rg.ru, ria.ru, rosbalt.ru, russian7.ru, sm-news.ru, smi2.ru, svpressa.ru, tass.ru, tvrain.ru, tvzvezda.ru, ura.ru, utro.ru, vestinn.ru, wek.ru

Thus, for some websites and days, we observe the number of unique visitors reported by more than one source. We remove observations for which the number of visitors is close to zero, because this usually indicates that the data source does not have reliable information on the traffic for that time period. We then build regressions predicting the number of visitors to an outlet j on day d from data source S, based on the number of visitors reported by `top.mail.ru` and the fixed effects included in all our models.

$$y_{jd}^{S} = a y_{jd}^{top.mail.ru} + \rho_{jw} + \delta_{j,\mathrm{wd}} + \theta_d + \xi_{jd} \tag{A1}$$

where y_{jd}^{S} is the total number of visitors to outlet j on day d reported by data source S. All regressions include ρ_{jw}, $\delta_{j,\mathrm{wd}}$, and θ_d, which are outlet-week, outlet-day-of-the-week, and day fixed effects, respectively. The term ξ_{jd} is an idiosyncratic shock that impacts outlet traffic reported by source S.

Table A3 presents the model estimates. The R-squared of the projected model using traffic data from `semrush.com` is relatively low, as `semrush.com` is not a traffic counter that directly observes and reports real visitors numbers, but rather a service that estimates traffic using proprietary algorithms. In other cases – particularly with liveinternet.ru – we observe a high correlation between reported traffic, which supports the reliability of data from `top.mail.ru`.

We nevertheless use data from `semrush.com` to approximate the magnitude of daily traffic for each outlet. Table A4 compares average daily traffic estimates from two sources.

A3.1 Total Yandex Visitors

To measure the effectiveness of Yandex recommendations, we use the daily number of Yandex search engine users as a proxy for the number of visitors to Yandex's main page. We have two data sources for these statistics:

- `top.mail.ru` – a source that does not track the total number of search engine users, but aggregates data from many websites that use it as a traffic counter. It provides data since 2011.

- `radar.yandex.ru` – the official Yandex platform, which provides data since 2015. Currently, the service reports only daily visit counts for the Yandex search engine, However, it used to publish statistics on visits to the Yandex main page before 2016. Archived versions of these metrics can still be found online, for example in web articles such as (Tokar, 2015).

We use `top.mail.ru` as our primary source, as it allows us to analyze earlier time periods and different audience segments. We assess the robustness of the daily variation in our measure using the `radar.yandex.ru` search engine visits statistics. In addition, we use archived data on

Table A3: Validation of top.mail.ru Data using Other Data Sources

	Dependent variable:		
	$y_{jd}^{semrush}$	$y_{jd}^{liveinternet}$	$y_{jd}^{rambler}$
	(1)	(2)	(3)
$y_{jd}^{top.mail.ru}$	0.336***	0.979***	0.578***
	(0.019)	(0.013)	(0.036)
Fixed Effects:			
- Week x Outlet	Y	Y	Y
- Week Day x Outlet	Y	Y	Y
- Date	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks
Observations	29,002	8,650	16,109
R^2	0.982	0.994	0.985
Adjusted R^2	0.976	0.992	0.981
R^2 (Proj. model):	0.23	0.728	0.458
Notes:	*p<0.1; **p<0.05; ***p<0.01		

visits to the Yandex main page to quantify the extent to which our proxy understates total traffic and, consequently, overstates the effectiveness of the Yandex top-5.

Table A5 presents model estimates that validate the data from `top.mail.ru`. We observe that the values reported by `radar.yandex.ru` are systematically higher. This is mostly explained by the fact that `top.mail.ru` captures visitors, while `radar.yandex.ru` provides statistics on the number of visits. Another reason is that `top.mail.ru` doesn't capture all Yandex users – only those visiting websites that have chosen `top.mail.ru` as their traffic tracker. The most important aspect for our analysis is the correlation between the two data sources, and we find that the R-squared values are very high.

Figure A1 compares the daily number of Yandex main page visitors reported by `radar.yandex.ru` with our proxy based on `top.mail.ru`. Two series exhibit a correlation of 0.82, and the numbers from `radar.yandex.ru` are on average 1.85 times higher. This gap arises because the proxy constructed from `top.mail.ru` captures only those Yandex users who visited at least one site tracked by `top.mail.ru` on a given day – that is, a more active subset of users. Using this comparison, we adjust our estimates of the Yandex top-5 effectiveness to account for less active users as well. As shown in Figure 8, the Yandex top-5 section directs approximately 2% of "more active" main-page visitors, or about 1.1% of all visitors.

Table A4: Outlets Average Daily Visitors in 2018

Outlet	top.mail.ru		semrush.com	
Outlet	Avg Daily Visitors	Ranking	Avg Daily Visitors	Ranking
ria.ru	3426205	1	3205213	1
mk.ru	1641932	2	1317274	2
rg.ru	1034156	3	740947	4
aif.ru	897314	4	1018081	3
echo.msk.ru	804749	5	715925	5
kommersant.ru	761327	6	474322	7
iz.ru	648264	7	561116	6
newsru.com	294035	8	446458	8
fontanka.ru	281809	9	191924	9
ng.ru	103508	10	76333	10
trud.ru	17655	11	14764	11

Notes: We compare the numbers for 2018, as `semrush.com` only began accounting for mobile traffic that year. Outlets `gazeta.ru` and `lenta.ru` are omitted from the table because they stopped using `top.mail.ru` as a traffic counter before 2018.

Table A5: Validation of `top.mail.ru` Yandex Users with `radar.yandex.ru` Data

	Dependent variable:
	$y_{jd}^{radar.yandex.ru}$
$y_{jd}^{top.mail.ru}$	8.287***
	(0.149)
Fixed Effects:	
- Week	Y
- Week Day	Y
Clustered S.E.	Weeks
Observations	15,327
R^2	0.990
Adjusted R^2	0.990
R^2 (Proj. model):	0.897
Notes:	*p<0.1; **p<0.05; ***p<0.01

Figures A2 and A3 show that Yandex's market share remains stable over the study period and, if anything, increases slightly after the introduction of the law on both desktop and mobile devices. Thus, we find no evidence that users substituted away from Yandex toward other platforms. Market shares are measured as the proportion of visits and are obtained from `radar.yandex.ru`. Google, Yandex's main competitor, accounts for approximately 25% of the desktop market and 55% of the mobile market.

Figure A1: Yandex Main Page Visitors Validation

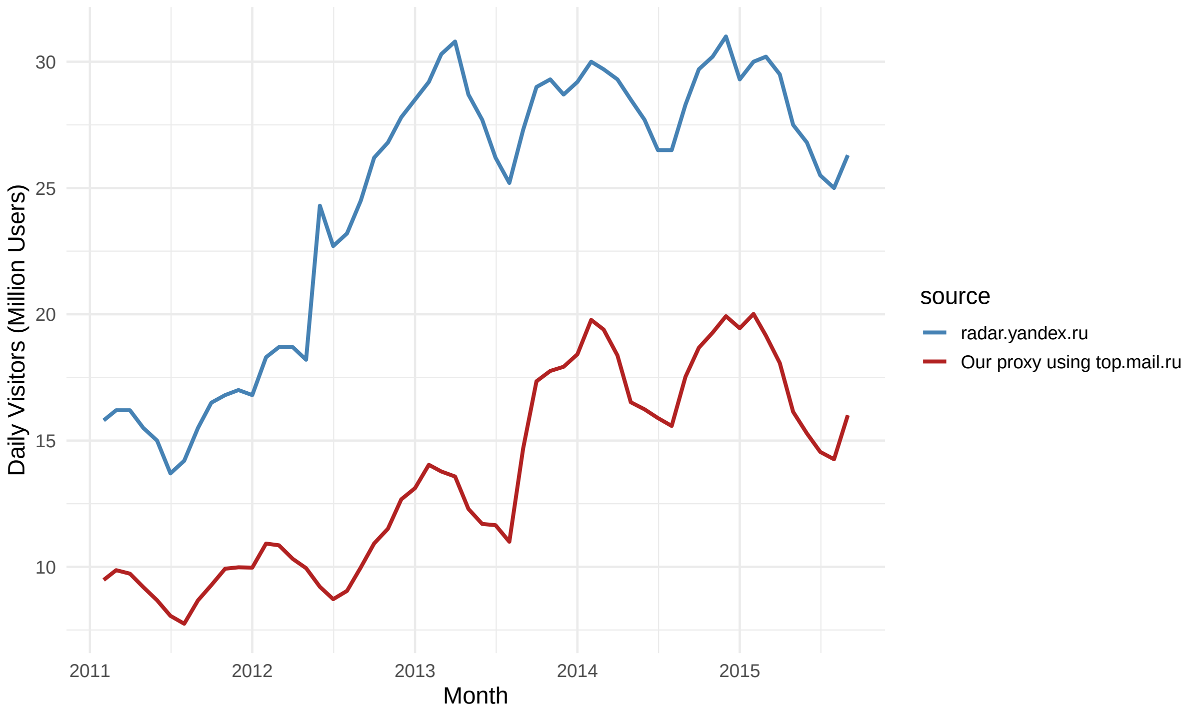


Figure A2: Yandex Search Engine Market Share – Desktop

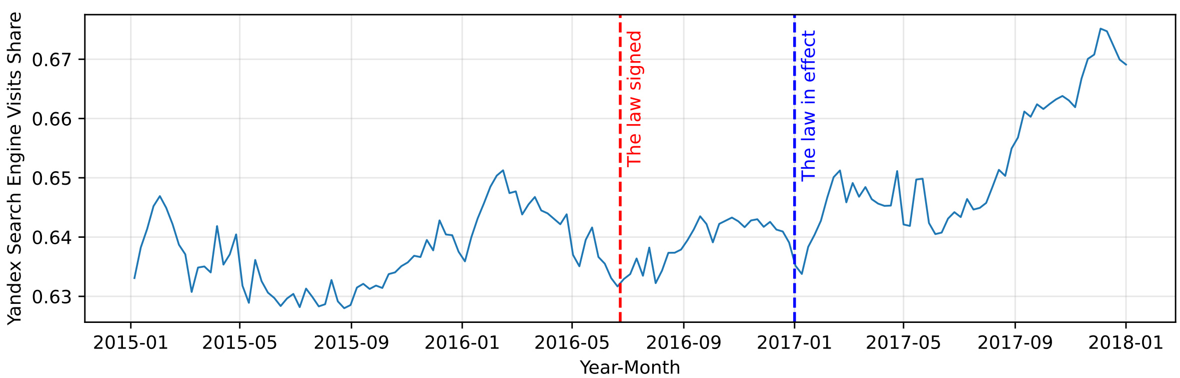
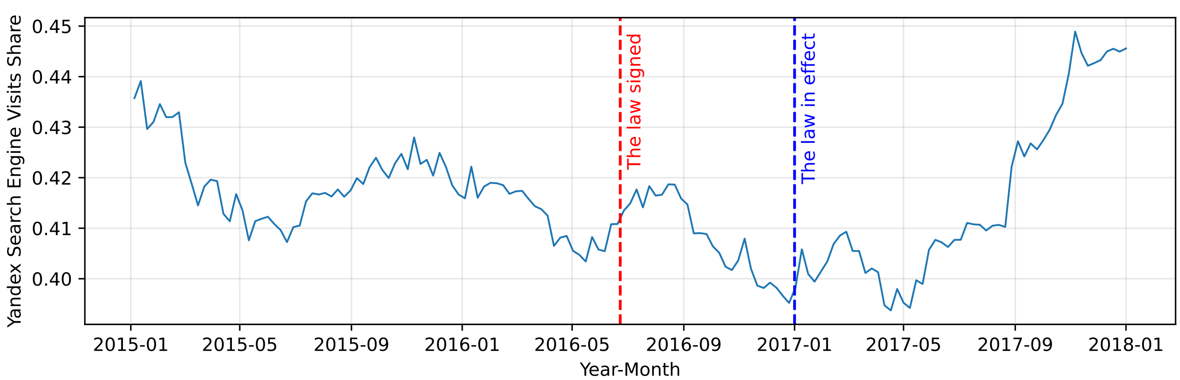


Figure A3: Yandex Search Engine Market Share – Mobile Devices



A4 Foundation of Public Opinion Survey Data

To describe individuals who can be more effectively directed by Yandex Top-5, we use data from the 2011 survey conducted by Foundation of Public Opinion[25], one of the leading independent sociological organizations in Russia.

The dataset includes demographic characteristics of 56,900 individuals, along with their responses to questions related to political beliefs. We aggregate these responses by demographic groups – specifically age, gender, and region – to align with our Yandex's Top-5 Effectiveness estimates.

For the regression analysis, we construct the X variables by calculating the share of respondents in each demographic group who chose particular answer options for each survey question. Table A6 presents the variable names, the full survey questions (translated into English), and the corresponding answer options whose shares were extracted.

Only respondents who report frequent internet use are used to compute the shares.

Table A6: Survey questions and answer options used to construct regression variables.

Variable Name	Full Question	Answer option
Gov. party supporters	If in December 2011 you participate in the State Duma elections, which party are you most likely to vote for?	"United Russia (Dmitry Medvedev)"
Low income	Which statement best describes the material situation of your family?	"We do not have enough money even for food" OR "We have enough money for food, but cannot buy clothes or shoes"
Likely protesters	Do you personally allow or exclude the possibility of participating in any protest actions?	"I allow"
See corruption as an issue	In your opinion, which of the tasks listed on the card should the president and the new government focus on the most? Please select no more than five tasks.	"Fight against corruption"
Support democracy	In your opinion, which of the tasks listed on the card should the president and the new government focus on the most? Please select no more than five tasks.	"Protection of human rights and civil liberties" OR "Development of civil society and local self-government" OR "Development of democratic institutions and political competition"

[25] https://fom.ru/

A5 News Events Detection Algorithm

We use the Topic Detection and Tracking (TDT) algorithm developed by Allan et al. (1998) and used in the economic literature (e.g. see Cagé et al., 2020) to cluster articles into news events.

The approach consists of four major steps:

- **Preprocessing** – we remove HTML tags, perform stemming, and tokenize the text.
- **Vectorization** – we combine article titles and body texts and transform them into numeric vectors. Since there are multiple methods for text vectorization, we test several approaches to select the most suitable one and check for the robustness of our results. We focus on several widely used methods: TF-IDF, Word2Vec and Word2Vec weighted with TF-IDF.
- **Clusterization** – in TDT, the clustering process takes into account both the vector representations and the publication time of articles. We use a three-day sliding window, and within each such period we cluster articles based on their cosine similarity. If there is a news article that belongs to two different clusters across the overlapping time periods, we merge these clusters into a single news event. If no new similar news articles are published the next day, then the news event is finalized.
  There are several hyperparameters that must be selected at this stage. First, we compare different clustering algorithms suitable for our task, including a density-based approach and agglomerative clustering.
  Second, for each of the algorithms, we test different radius values that define clusters size by setting a threshold for cosine similarity.
  We keep the sliding window size (3 days) and the overlap period (1 day) unchanged.
- **Postprocessing** – we remove duplicate articles from the results. An article is considered a duplicate if two or more observations published by one outlet clustered into the same news event with a very small radius[26].

  We also exclude all articles published by Ukrainian outlets (which were initially included to our dataset to help find the best clustering settings). After these adjustments, we say that all clusters with only one article are not news events and change their status to "unclustered".

[26] We apply the DBSCAN algorithm with a radius parameter of 0.001 to cluster TF-IDF vectors of news articles. To validate the robustness of this approach, we compare the results with Agglomerative Clustering applied to the same TF-IDF vectors. The two methods show a high degree of agreement: 99.92% of the articles identified as potential duplicates by DBSCAN are also flagged by Agglomerative Clustering, while 100% of the duplicates detected by the Agglomerative approach are also found by DBSCAN. We rely on TF–IDF vectors instead of word embeddings, since our goal is to identify near-duplicates by exact word overlap rather than by semantic similarity.

Additionally, we "uncluster" 1,063 events which are longer than half a month (15 days), because such events usually correspond to recurring reports rather than real events. For example, weather reports, daily war reports, or horoscopes (Table A7).

Table A7: Example of Removed Longest Events

Event ID	Length (days)	Sample Titles
2885230	1311	Гидрометцентр: В Москве 5 августа возможны дожди с грозами
		Hydrometcenter: In Moscow on August 5, rain with thunderstorms is possible
		В Украине сегодня без осадков, местами жара до +34° (карта)
		In Ukraine today no precipitation, in places heat up to +34° (map)
		В аэропортах Москвы отменено около 80 авиарейсов и около 60 задержано
		About 80 flights canceled and 60 delayed at Moscow airports
		МЧС сделало метеопредупреждение на 6 октября
		EMERCOM issued a weather warning for October 6
		В Липецк ненадолго вернутся ночные заморозки
		Night frosts will briefly return to Lipetsk
4523211	236	К перемирию в Сирии присоединились уже 972 населенных пункта
		972 settlements have already joined the ceasefire in Syria
		В Дамаске за сутки было зафиксировано пять нарушений перемирия
		Five ceasefire violations were recorded in Damascus in 24 hours
		Россия поддерживает продление режима перемирия в Сирии на 48 часов
		Russia supports extending the ceasefire in Syria by 48 hours
		В Сирии за сутки зафиксировано три нарушения режима прекращения огня
		Three violations of the ceasefire regime recorded in Syria in 24 hours
		В Сирии 207 населенных пунктов присоединились к перемирию
		207 settlements in Syria joined the ceasefire
7291694	232	Гороскоп на 22 октября для всех знаков зодиака
		Horoscope for October 22 for all zodiac signs
		Гороскоп на 17 июля: Рыбы, не делайте поспешных выводов
		Horoscope for July 17: Pisces, do not rush to conclusions
		Гороскоп на 10.05.2018 для знаков зодиака
		Horoscope for May 10, 2018, for zodiac signs
		Гороскоп на 1 мая для всех знаков зодиака
		Horoscope for May 1 for all zodiac signs
		Гороскоп на сегодня для всех знаков зодиака - Рыбы, Овен, Телец, Весы, Водолей
		Horoscope for today for all zodiac signs – Pisces, Aries, Taurus, Libra, Aquarius
1735093	172	Теплая погода сохранится в Москве и Подмосковье
		Warm weather will continue in Moscow and the Moscow region
		Погода на завтра: На юге Украины дожди с грозами, температура до +28
		Weather for tomorrow: In southern Ukraine, rain with thunderstorms, temperature up to +28
		Главное за день: Богатырева в розыске, Путин рассказал, как помог Януковичу
		Main news of the day: Bogatyreva wanted, Putin explained how he helped Yanukovych
		До начала декабря в Петербурге не будет снега
		No snow in St. Petersburg until the beginning of December
		На предстоящей неделе в Красноярске будет дождливо и прохладно
		Next week in Krasnoyarsk will be rainy and cool
2535399	168	В выходные липчанам прогнозируют дождь и град
		Rain and hail expected in Lipetsk over the weekend
		Воскресенье во Владивостоке будет таким же теплым, как и суббота
		Sunday in Vladivostok will be as warm as Saturday
		Погода на завтра, 12 февраля
		Weather for tomorrow, February 12
		В течение недели в Украине дожди и заморозки будут чередоваться с потеплением
		During the week in Ukraine, rain and frosts will alternate with warming
		Апрельское солнце подогреет столицу
		April sun will warm the capital

Thus, implementing TDT requires selecting several hyperparameters: the vectorization method, the clustering algorithm, and the cluster radius. To find the most suitable settings and verify the robustness, we test several commonly used options. For vector representation, we try TF-IDF, Word2Vec (calculating an average vector for each article) and Word2Vec weighted with TF-IDF (using a weighted average). For clustering we compare two approaches: DBSCAN and Agglomerative clustering with complete linkage. We sort through different radius values for each algorithm.

TDT is an unsupervised algorithm, so there is no single clear criterion for selecting the optimal settings. We use two measures to guide this choice. First, our initial dataset includes not only the 50 Russian outlets of interest but also 19 Ukrainian outlets that publish articles in Russian.

We consider the clustering to be more accurate if it effectively separates Russian and Ukrainian news events. Second, we try to avoid excessive fragmentation in the final clusters.

To evaluate how well the clustering separates Russian and Ukrainian news, we calculate the isolation index used as a measure of segregation by Gentzkow and Shapiro (2011).

Let $c \in C$ represent a single news event, i.e., a cluster of articles, and let C indicate the full clustering – that is, the partitioning of all articles in the dataset (we consider all unclustered articles as a separate cluster labeled "unclustered"). Let ru and ukr denote the total number of articles published by Russian and Ukrainian outlets in the dataset. Define $articles = ru + ukr$ as the total number of articles. Similarly, for each cluster c, let ru_c and ukr_c indicate the number of articles by Russian and Ukrainian outlets in that cluster, so $articles_c = ru_c + ukr_c$.

Then, the measure of segregation is calculated as follows:

$$S = \frac{\sum_{c \in C} \left(\frac{ukr_c}{ukr} * \frac{ukr_c}{articles_c} \right) - \left(\frac{ukr}{articles} \right)}{1 - \left(\frac{ukr}{articles} \right)} \tag{A2}$$

Table A8 presents the results for each clustering.

We choose Agglomerative clustering with a radius of 0.6 on TF-IDF vectors as our main specification for the event detection model, because it achieves the highest segregation score.

For robustness checks, we also use two alternative specifications: Agglomerative clustering with a radius of 0.04 on Word2Vec vectors, which is the second-best option according to our metrics; and DBSCAN with a radius of 0.35 on TF-IDF, which has the best performance among the density-based clustering algorithms.

Table A8: Clustering Quality by Vectorization, Algorithm, and Radius

Vectorization	Clustering	Radius	Segregation	Share Clustered	# Clusters
TF-IDF	Agglomerative	0.300	0.341	0.406	3249416
		0.350	0.385	0.482	3539821
		0.400	0.418	0.549	3669509
		0.450	0.440	0.608	3661717
		0.500	0.455	0.659	3554150
		0.550	0.463	0.703	3380271
		0.600	0.466	0.743	3170092
		0.650	0.462	0.780	2948266
		0.700	0.452	0.817	2718771
	DBSCAN	0.150	0.169	0.185	1558906
		0.200	0.230	0.268	1933052
		0.250	0.280	0.362	2133530
		0.300	0.313	0.454	2151632
		0.350	0.329	0.536	2050910
		0.400	0.318	0.605	1904851
		0.450	0.295	0.661	1754202
		0.500	0.260	0.709	1605426
		0.600	0.169	0.786	1246448
		0.800	0.074	0.935	86537
Word2Vec	Agglomerative	0.010	0.137	0.148	1423882
		0.030	0.375	0.512	3497956
		0.040	0.412	0.643	3510413
		0.050	0.405	0.738	3182186
		0.060	0.370	0.805	2705827
		0.070	0.322	0.852	2180347
		0.100	0.196	0.920	943748
		0.150	0.119	0.948	205896
	DBSCAN	0.010	0.141	0.155	1349820
		0.020	0.237	0.372	1843707
		0.030	0.206	0.577	1494763
		0.040	0.125	0.716	966070
		0.050	0.072	0.811	526740
		0.060	0.054	0.874	248623
		0.070	0.059	0.910	108567
		0.100	0.090	0.948	10960
		0.150	0.105	0.956	1234
		0.200	0.108	0.957	241
Word2Vec + TF-IDF	Agglomerative	0.010	0.115	0.126	1223426
		0.030	0.313	0.403	3080790
		0.040	0.375	0.543	3438385
		0.050	0.392	0.652	3349252
		0.060	0.374	0.735	3029465
		0.070	0.337	0.797	2606181
		0.100	0.209	0.898	1317055
		0.150	0.118	0.942	325846
		0.200	0.106	0.952	92688
	DBSCAN	0.010	0.118	0.130	1193450
		0.020	0.210	0.268	1813370
		0.030	0.218	0.455	1725257
		0.040	0.157	0.615	1247854
		0.050	0.093	0.732	780750
		0.100	0.073	0.938	24265
		0.200	0.108	0.957	636

A6 Detecting Big and Sensitive Events

We classify news events obtained after implementing the TDT algorithm into two categories: "big" and "sensitive".

We define "big" events as clusters with articles from 10 or more different outlets. In total, 135,282 events (the top 5% by the number of outlets in the cluster) are labeled as "big" in our dataset.

Among "big" news events, we identify those that are under-reported by government-owned outlets compared to independent ones. We call such events "sensitive".

To detect such clusters, we implement an algorithm developed by Simonov and Rao (2022), but apply it at the event level rather than the word level, as in the original paper. For each event and outlet, we compute the share of articles written about this event. Then, for each event, we order the outlets in descending order by this share and compute the difference between the average rank of government-owned outlets and the average rank of independent ones. This yields a value that indicates how sensitive the event is: a higher value means that the event was more underrepresented by government-owned outlets. Table A9 reports the most "sensitive" events.

Table A9: The Five Most Sensitive News Events (Largest Under-Coverage by Government-Controlled Outlets Relative to Independent Ones)

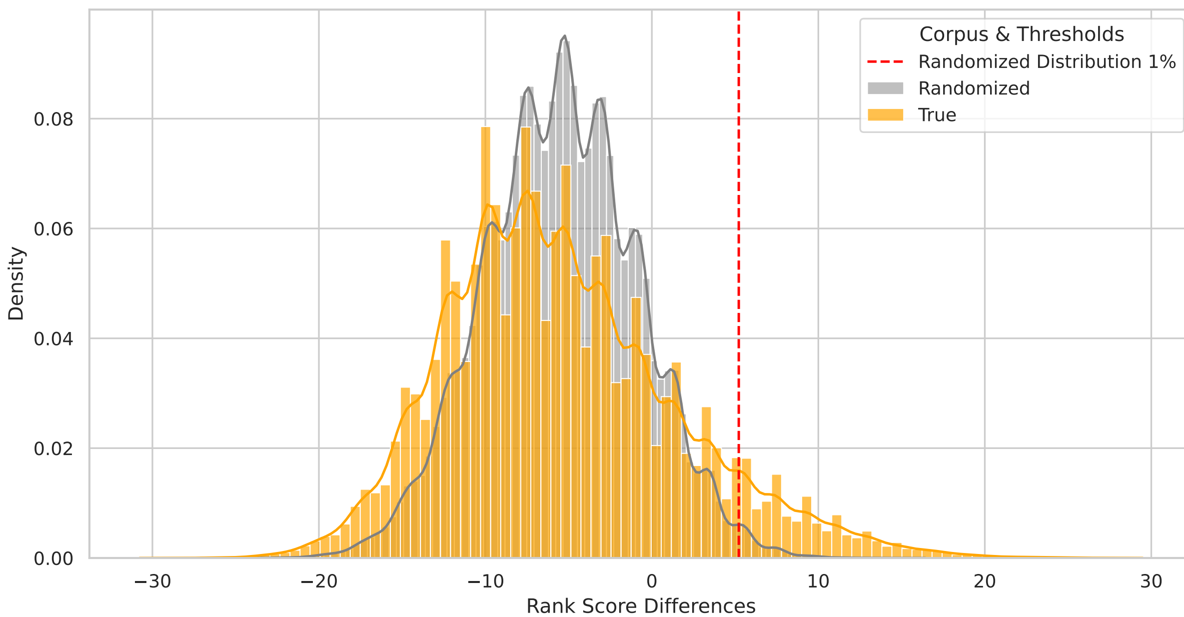
Index	Examples of Titles (Original and Translated) of News Articles in this Event
1	Навальный назвал стоимость расследования о семье генпрокурора
	Navalny revealed the cost of the investigation into the prosecutor general's family
2	Роскомнадзор стал третьей стороной в иске Дерипаски о его фотографиях в соцсетях
	Roskomnadzor joined as a third party in Deripaska's lawsuit over his photos on social media
3	СМИ: Арестованные сотрудники ФСБ сливали информацию ЦРУ
	Media: arrested FSB officers were leaking information to the CIA
4	Навальный удалил с сайта материалы про Дерипаску
	Navalny removed materials about Deripaska from his website
5	Навальный проиграл Усманову
	Navalny lost to Usmanov

We exclude events which were reported by fewer than 10 outlets.

To choose a threshold for sensitive events, we adopt the strategy described in Simonov and Rao (2022).We shuffle the shares within each outlet, recompute the rankings for each event, and thereby obtain the distribution of possible differences between the average ranks of government-owned and independent outlets.

As in Simonov and Rao (2022), the true distribution has heavier tails than the simulated distribution. We label 8,734 big events as sensitive, as their rank difference exceeds the top 1% of the randomized distribution. Figure A4 shows the true and simulated distributions of event ranks, along with the selected threshold.

Figure A4: True and Randomized Distributions of Ranks Differences for Events



A7 Additional Tables and Figures

Figure A5: Weekly Average Number of Snapshots per Day for Yandex Top-5 on Webarchive

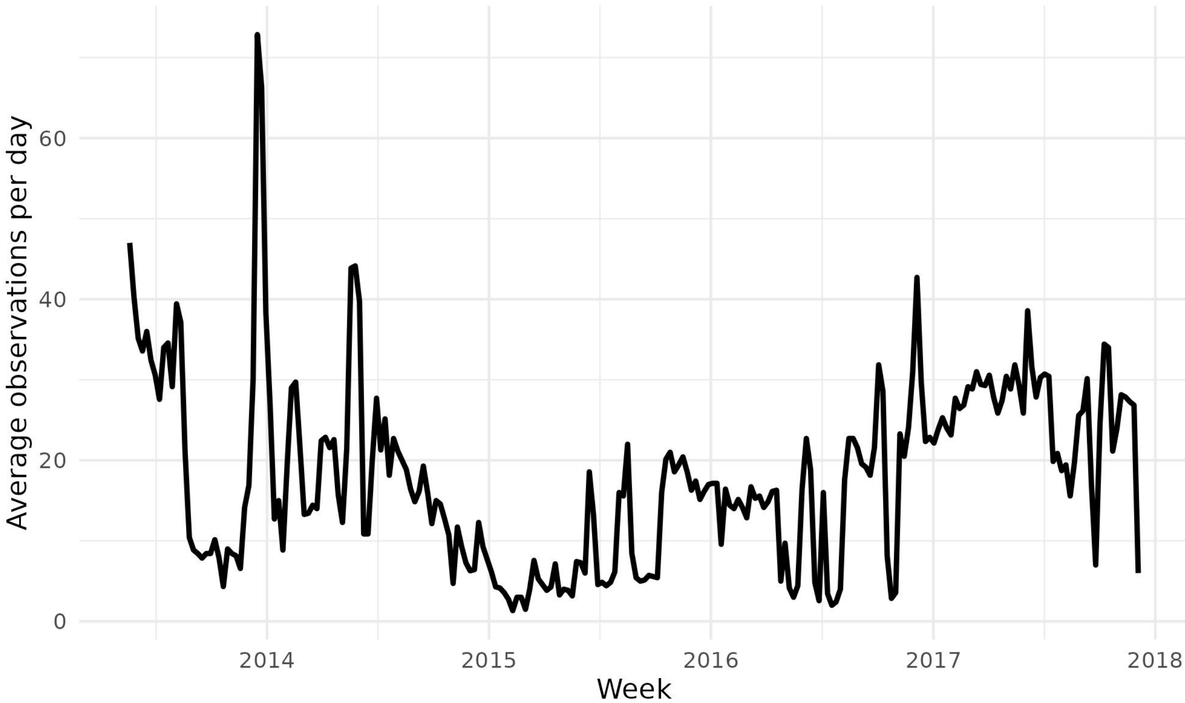


Table A10: Effect of Outlet References in Yandex's Top-5: By Position

	Dependent variable:			
	y_{jd}			
	(1)	(2)	(3)	(4)
$X(j \text{ in position } 1)_d$	92.00***	95.05***	97.47***	97.26***
	(15.33)	(18.90)	(18.77)	(18.75)
$X(j \text{ in position } 2)_d$	83.92***	95.34***	96.86***	96.82***
	(15.61)	(18.85)	(18.75)	(18.79)
$X(j \text{ in position } 3)_d$	58.33***	67.20***	68.60***	68.61***
	(13.94)	(15.67)	(15.61)	(15.54)
$X(j \text{ in position } 4)_d$	55.28***	59.29***	61.29***	61.65***
	(13.55)	(15.72)	(15.68)	(15.70)
$X(j \text{ in position } 5)_d$	51.50***	57.09***	58.63***	58.44***
	(13.53)	(15.70)	(15.69)	(15.72)
$I(j \text{ in top-6})_d$		10.57***		
		(3.43)		
$\sum_{p=6}^{10} X(j \text{ in position } p)_d$			17.69*	
			(9.33)	
$X(j \text{ in position } 6)_d$				28.76
				(24.28)
$X(j \text{ in position } 7)_d$				9.55
				(22.70)
$X(j \text{ in position } 8)_d$				18.05
				(26.39)
$X(j \text{ in position } 9)_d$				4.98
				(25.46)
$X(j \text{ in position } 10)_d$				28.10
				(25.85)
Fixed Effects:				
- Week x Outlet	Y	Y	Y	Y
- Week Day x Outlet	Y	Y	Y	Y
- Date	Y	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks	Weeks
Observations	22,971	20,084	20,084	20,084
Adjusted R^2	0.95	0.95	0.95	0.95

Notes: The dependent variable y_{jd} is the number of daily visitors to outlet j, in thousands; coefficients are in thousands of visitors for full-day occupancy of the given position. Positions 1–5 appear on Yandex's front page; positions 6–10 do not. Standard errors, clustered by week, in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A11: Effect of Outlet References in Yandex's Top-5: By Type of Visitors

	Dependent variable:				
	y_{jd}	y_{jd}^{New}	y_{jd}^{Regular}	$y_{jd}^{\text{Pageviews}}$	$y_{jd}^{\text{Av. Pageviews}}$
	(1)	(2)	(3)	(4)	(5)
$X(j \text{ in top-5})_d$	374.03***	157.06***	3.41	681.30***	-0.76***
	(38.13)	(15.42)	(2.79)	(88.10)	(0.09)
$I(j \text{ in top-6})_d$	10.64***	7.40***	-0.21	15.51	-0.06***
	(3.43)	(1.73)	(0.23)	(10.24)	(0.01)
Fixed Effects:					
- Week x Outlet	Y	Y	Y	Y	Y
- Week Day x Outlet	Y	Y	Y	Y	Y
- Date	Y	Y	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks	Weeks	Weeks
Observations	20,084	18,655	18,655	18,655	18,655
Adjusted R^2	0.95	0.88	1.00	0.95	0.96

Notes: Dependent variables in columns 1–4 are in thousands (visitors or pageviews); column 5 is pageviews per visitor. New visitors are defined as visitors who came to this site for the first time in 91 days. Regular visitors are defined as visitors who visited this site for at least 7 days in the last 28 days. Standard errors, clustered by week, in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A12: Effect of Outlet References in Yandex's Top-5: Overall Traffic (Weighted by Snapshots per Day)

	Dependent variable:					
	$log(y_{jd})$			y_{jd}		
	(1)	(2)	(3)	(4)	(5)	(6)
$X(j \text{ in top-5})_d$	0.54***	0.53***	0.46***	407.49***	436.95***	422.36***
	(0.04)	(0.04)	(0.04)	(46.03)	(51.01)	(52.00)
$I(j \text{ in top-6})_d$		0.04***	0.04***		8.32**	5.83
		(0.01)	(0.01)		(4.09)	(4.03)
$X(\text{coverage of top-5 events by } j)_d$			0.13***			26.27***
			(0.01)			(9.23)
Fixed Effects:						
- Week x Outlet	Y	Y	Y	Y	Y	Y
- Week Day x Outlet	Y	Y	Y	Y	Y	Y
- Date	Y	Y	Y	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks	Weeks	Weeks	Weeks
Observations	22,971	20,084	16,057	22,971	20,084	16,057
Adjusted R^2	0.98	0.98	0.99	0.96	0.96	0.96

Notes: Observations are weighted by the number of available front-page snapshots on that day. The dependent variable y_{jd} is the number of daily visitors to outlet j, in thousands; coefficients in columns 4–6 are therefore in thousands of visitors. Standard errors, clustered by week, in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A13: Effectiveness of Outlet References in Yandex's Top-5: By Event Topic

	Dependent variable:
	full_formula
$X(j \text{ in top-5})_d$	0.036***
	(0.008)
$X(j \text{ in top-5 with topic Society})_d$	0.034*
	(0.020)
$X(j \text{ in top-5 with topic Incidents})_d$	0.002
	(0.010)
$X(j \text{ in top-5 with topic Domestic politics})_d$	0.006
	(0.011)
$X(j \text{ in top-5 with topic Economy})_d$	0.001
	(0.011)
$X(j \text{ in top-5 with topic Crime})_d$	-0.010
	(0.010)
$X(j \text{ in top-5 with topic Military})_d$	0.001
	(0.013)
$X(j \text{ in top-5 with topic Foreign politics})_d$	-0.017*
	(0.009)
$X(j \text{ in top-5 with topic Culture})_d$	0.003
	(0.013)
$X(j \text{ in top-5 with topic Sports})_d$	-0.018**
	(0.008)
$X(j \text{ in top-5 with topic Local})_d$	0.015
	(0.017)
$X(j \text{ in top-5 with topic Health})_d$	-0.011
	(0.011)
$X(j \text{ in top-5 with topic Environment})_d$	0.040
	(0.030)
$X(j \text{ in top-5 with topic Science})_d$	-0.010
	(0.012)
$X(j \text{ in top-5 with topic Technology})_d$	0.017
	(0.029)
$X(j \text{ in top-5 with topic Education})_d$	-0.022
	(0.029)
$X(j \text{ in top-5 with topic Weather})_d$	0.004
	(0.017)
$X(j \text{ in top-5 with big event})_d$	-0.007
	(0.004)
$I(j \text{ in top-6})_d$	0.001***
	(0.0002)
Fixed Effects:	
- Week x Outlet	Y
- Week Day x Outlet	Y
- Date	Y
Clustered S.E.	Weeks
Observations	20,084
Adjusted R^2	0.957
Notes:	*p<0.1; **p<0.05; ***p<0.01

Table A14: Effectiveness of Outlet References in Yandex's Top-5: By Event Type

	Dependent variable:
	$y_{jd}/y_{jd}^{\text{Yandex}}$
$X(j \text{ in top-5})_d$	0.03^{***}
	(0.004)
$X(j \text{ in top-5 with big IND event})_d$	-0.001
	(0.01)
$X(j \text{ in top-5 with big event})_d$	-0.01
	(0.004)
$I(j \text{ in top-6})_d$	0.001^{***}
	(0.0002)
Fixed Effects:	
- Week x Outlet	Y
- Week Day x Outlet	Y
- Date	Y
Clustered S.E.	Weeks
Observations	20,084
Adjusted R^2	0.96
Notes:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A15: Effectiveness of Outlet References in Yandex's Top-5: By Event Yandex Ranking

	Dependent variable:
	$y_{jd}/y_{jd}^{\text{Yandex}}$
$X(j \text{ in top-5})_d$	0.024***
	(0.003)
$X(j \text{ in top-5, bottom quartile news value})_d$	0.004
	(0.004)
$X(j \text{ in top-5, top quartile news value})_d$	0.003
	(0.005)
$I(j \text{ in top-6})_d$	0.001***
	(0.0002)
Fixed Effects:	
- Week x Outlet	*Y*
- Week Day x Outlet	*Y*
- Date	*Y*
Clustered S.E.	*Weeks*
Observations	20,084
Adjusted R^2	0.957
Notes:	*p<0.1; **p<0.05; ***p<0.01

Table A16: Effectiveness of Outlet References in Yandex's Top-5: By Regional News

	Dependent variable:	
	$y_{jd}/y_{jd}^{\text{Yandex}}$	
	(1)	(2)
$X(j \text{ in top-5})_d$	0.027***	0.027***
	(0.003)	(0.003)
$X(j \text{ in top-5 with news about this region})_d$	0.144***	0.148***
	(0.016)	(0.017)
$I(j \text{ wrote about this region's news})_d$		0.001***
		(0.0001)
$I(j \text{ in top-6})_d$	0.001**	0.0004
	(0.0003)	(0.0003)
Fixed Effects:		
- Week x Outlet x Region	Y	Y
- Week Day x Outlet x Region	Y	Y
- Date x Region	Y	Y
Clustered S.E.	$Weeks$	$Weeks$
Observations	597,888	492,192
Adjusted R^2	0.951	0.953
Notes:	*p<0.1; **p<0.05; ***p<0.01	

Table A17: Effectiveness of Outlet References in Yandex's Top-5: By Operation System

	Dependent variable:	
	$y_{jd}/y_{jd}^{\text{Yandex}}$	
	Detailed	Grouped
$X(j \text{ in top-5})_d$ x Windows	0.029***	
	(0.005)	
$X(j \text{ in top-5})_d$ x Android	0.042***	
	(0.009)	
$X(j \text{ in top-5})_d$ x iOS	0.037***	
	(0.007)	
$X(j \text{ in top-5})_d$ x MacOS	0.044***	
	(0.009)	
$X(j \text{ in top-5})_d$ x Desktop		0.029***
		(0.005)
$X(j \text{ in top-5})_d$ x Mobile		0.040***
		(0.008)
$X(j \text{ in top-5})_d$ x Other	0.029***	0.027***
	(0.005)	(0.005)
Controls:		
- $I(j \text{ in top-6})_d$ x OS	Y	Y
Fixed Effects:		
- Week x Outlet x OS	Y	Y
- Week Day x Outlet x OS	Y	Y
- Date x OS	Y	Y
Clustered S.E.	Weeks	Weeks
Observations	39,405	23,643
Adjusted R^2	0.966	0.929

Notes: The estimates correspond to the a estimates from regression (1) with y_{id} normalized by the volume of traffic arriving on `yandex.ru` on day d. *p<0.1; **p<0.05; ***p<0.01.

Table A18: Effectiveness of Outlet References in Yandex's Top-5: By Age

	Dependent variable:								
	top5_efficiency								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$X(j \text{ in top-5})_d$	0.018***	0.011***	0.014***	0.020***	0.023***	0.025***	0.028***	0.030***	0.034***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
$I(j \text{ in top-6})_d$	0.0002	0.0002	0.0003*	0.0003	0.0003	0.0003	0.0004	0.001	0.001*
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Age (years)	<12	12-18	19-24	25-30	31-35	36-40	41-45	46-50	¿50
Fixed Effects:									
- Week x Outlet	Y	Y	Y	Y	Y	Y	Y	Y	Y
- Week Day x Outlet	Y	Y	Y	Y	Y	Y	Y	Y	Y
- Date	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks	Weeks	Weeks	Weeks	Weeks	Weeks	Weeks
Observations	6,920	6,920	6,920	6,920	6,920	6,920	6,920	6,920	6,920
Adjusted R^2	0.963	0.968	0.964	0.960	0.959	0.962	0.965	0.968	0.973

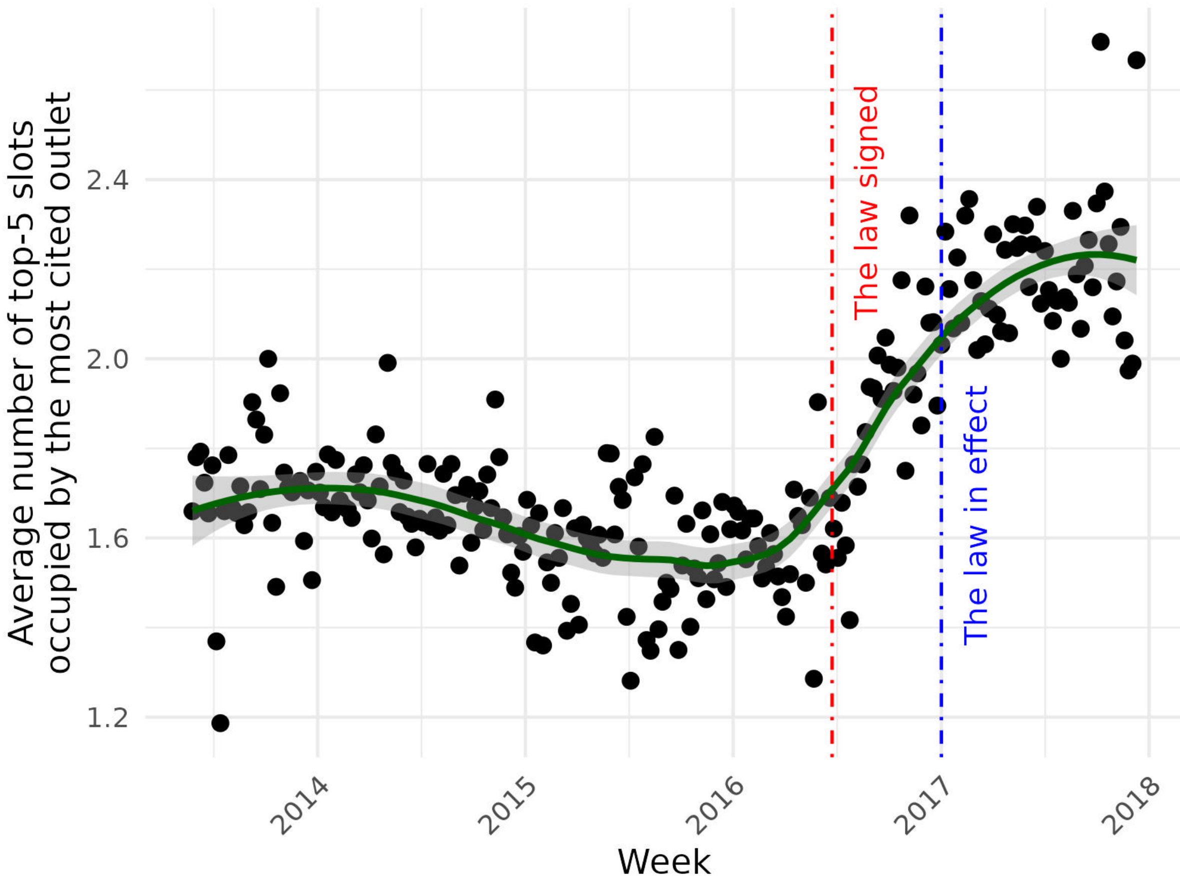
Notes: The estimates correspond to the a estimates from regression (1) with y_{id} normalized by the volume of traffic arriving on `yandex.ru` on day d. *p<0.1; **p<0.05; ***p<0.01.

Table A19: Effectiveness of Outlet References in Yandex's Top-5: By Gender

	Dependent variable:		
	$y_{jd}/y_{jd}^{\text{Yandex}}$		
	(1)	(2)	(3)
$X(j \text{ in top-5})_d$	0.029***	0.018***	0.033***
	(0.004)	(0.003)	(0.005)
$I(j \text{ in top-6})_d$	0.001*	0.0002	0.0004
	(0.0003)	(0.0002)	(0.0003)
Gender	Male	Female	Undetected
Fixed Effects:			
- Week x Outlet	Y	Y	Y
- Week Day x Outlet	Y	Y	Y
- Date	Y	Y	Y
Clustered S.E.	Weeks	Weeks	Weeks
Observations	6,920	6,920	6,920
Adjusted R^2	0.973	0.955	0.974

Notes: The estimates correspond to the a estimates from regression (1) with y_{id} normalized by the volume of traffic arriving on `yandex.ru` on day d. *p<0.1; **p<0.05; ***p<0.01.

Figure A6: The Number of Top-5 Slots Occupied by the Most-Cited Outlet



Notes: Dots represent a weekly average of the number of top-5 slots occupied by the most cited outlet at that moment. For instance, the value of two means that in an average snapshot, two out of five slots are occupied by the same outlet (e.g., ria.ru).

Table A20: Implied Outlet Weights Before and After the Law

		Probability of Citation (in %)				
Outlet	Type	Before the Law	After the Law	Difference	S.E. of the Difference	Tstat
Forbes.ru	ind	6.83	0.18	-6.65	0.25	-26.68
novayagazeta.ru	ind	3.98	0.01	-3.97	0.15	-26.08
vesti.ru	gov	4.34	0.64	-3.70	0.15	-24.14
tvrain.ru	ind	2.72	0.05	-2.68	0.14	-19.18
vz.ru	gov	2.57	0.05	-2.52	0.10	-24.62
Republic.ru	ind	2.27	0.11	-2.16	0.17	-13.06
echo.msk.ru	ind	2.39	0.32	-2.07	0.11	-18.60
iz.ru	inf	7.64	5.75	-1.89	0.45	-4.20
ntv.ru	gov	2.00	0.29	-1.71	0.15	-11.74
1tv.ru	gov	1.60	0.08	-1.52	0.18	-8.38
aif.ru	gov	1.39	0.05	-1.34	0.11	-12.64
rusplt.ru	inf	0.66	0.00	-0.66	0.08	-8.00
NewsRu.com	ind	0.65	0.00	-0.65	0.07	-9.65
ng.ru	inf	0.67	0.02	-0.65	0.07	-9.45
bfm.ru	inf	0.72	0.11	-0.61	0.11	-5.82
mk.ru	inf	0.15	0.00	-0.15	0.03	-5.17
Lenta.Ru	inf	7.24	7.15	-0.09	0.34	-0.27
bbc.com/russian	int	0.03	0.00	-0.03	0.02	-1.51
sobesednik.ru	inf	0.02	0.00	-0.02	0.01	-2.50
trud.ru	inf	0.02	0.00	-0.02	0.01	-1.44
rosbalt.ru	inf	0.01	0.00	-0.01	0.02	-0.52
svoboda.org	int	0.00	0.00	-0.00	0.02	-0.23
dni.ru	gov	0.00	0.00	-0.00	0.01	-0.43
Kp.ru	inf	0.01	0.01	-0.00	0.02	-0.11
russian.rt.com	gov	1.37	1.37	-0.00	0.17	-0.01
tjournal.ru	ind	0.00	0.00	-0.00	0.01	-0.10
newtimes.ru	ind	0.00	0.00	-0.00	0.08	-0.01
ridus.ru	inf	0.00	0.00	-0.00	0.01	-0.07
FB.ru	inf	0.00	0.00	-0.00	0.00	-0.03
zona.media	ind	0.00	0.00	-0.00	0.00	-0.00
golos-ameriki.ru	int	0.00	0.00	-0.00	0.02	-0.00
meduza.io	int	0.00	0.00	0.00	0.00	0.00
Znak.com	inf	0.00	0.00	0.00	0.00	0.06
polit.ru	inf	0.00	0.00	0.00	0.01	0.05
fontanka.ru	inf	0.00	0.00	0.00	0.00	0.10
utro.ru	inf	0.00	0.00	0.00	0.01	0.14
Life.ru	inf	0.00	0.00	0.00	0.01	0.16
dw.com/ru	int	0.00	0.01	0.01	0.00	1.45
svpressa.ru	inf	0.00	0.01	0.01	0.01	0.70
snob.ru	ind	0.01	0.06	0.05	0.03	1.45
the-village.ru	ind	0.00	0.06	0.06	0.04	1.58
vedomosti.ru	ind	2.92	3.28	0.36	0.23	1.53
regnum.ru	inf	2.93	4.21	1.28	0.19	6.56
kommersant.ru	inf	3.69	7.47	3.78	0.31	12.12
rbc.ru	ind	7.10	11.82	4.72	0.34	14.00
rg.ru	gov	10.69	15.42	4.73	0.31	15.40
gazeta.ru	inf	7.08	12.38	5.30	0.34	15.39
tass.ru	gov	3.91	9.36	5.45	0.35	15.73
ria.ru	gov	12.36	19.73	7.37	0.46	16.03

Notes: Implied probabilities of citations based on outlet weights α_j and α'_j from equation (4).