



HAL
open science

Codes of conduct for algorithmic news recommendation: the Yandex.News controversy in Russia

Françoise Daucé, Benjamin Loveluck

► To cite this version:

Françoise Daucé, Benjamin Loveluck. Codes of conduct for algorithmic news recommendation: the Yandex.News controversy in Russia. *First Monday*, 2021, 26 (5). hal-03217497

HAL Id: hal-03217497

<https://hal.science/hal-03217497v1>

Submitted on 10 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Codes of conduct for algorithmic news recommendation: The Yandex.News controversy in Russia

by Françoise Daucé and Benjamin Loveluck

Abstract

In Russia, since 2011, the Yandex.News aggregator (Yandex.Novosti) — the Russian equivalent to Google News — has been suspected of political bias in the context of protests against electoral fraud followed by the Ukrainian crisis. This article first outlines the issues associated with automated news recommendation systems, their role as “algorithmic gatekeepers” and the questions they raise in terms of news diversity and possible manipulation. It then analyses the controversies which have developed around Yandex.News, particularly since the authorities have decided to regulate the way it operates through a law adopted in 2016. Finally, it provides an audit of Yandex.News aggregation in 2020, through a quantitative analysis of its database of sources and of the Top 5 results presented on the Yandex homepage. It shows the discrepancy between the diversity of the Russian online mediasphere and the narrowness of the Yandex.News media sample. This research contributes to the sociology of digital platforms and the study of “governance by algorithms”, showing how the Yandex news aggregator is a key asset in the Russian government’s overall disciplining of the country’s media and digital public sphere, in an ongoing effort to assert “digital sovereignty”.

Contents

[Introduction](#)

[1. Algorithmic gatekeeping in digital media ecosystems: Context and issues](#)

[2. Yandex.News as political controversy: Defining the “right” news aggregation](#)

[3. The Yandex rankings as a gateway to the algorithm and its transformations](#)

[Conclusion](#)



This paper is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).

Codes of conduct for algorithmic news recommendation: The Yandex.News controversy in Russia
by Françoise Daucé and Benjamin Loveluck.

First Monday, Volume 26, Number 5 - 3 May 2021

<https://firstmonday.org/ojs/index.php/fm/article/download/11708/10131>

doi: <https://dx.doi.org/10.5210/fm.v26i5.11708>

Introduction

Russia is among the few countries in the world where Google does not massively dominate the online search industry. In 2020, the Russian-language equivalent Yandex held just under half the market share (about 45 percent) [1]. Yandex has long benefited from a certain degree of autonomy, and its founders have even, at different moments, expressed political disagreement with the Kremlin. However, as a national economic champion and as a key player in the organization of information, it has been under tight scrutiny. This is especially the case since the 2011–2012 protests against electoral fraud and the 2014 annexation of Crimea, which have also constituted general turning points for Russia due to the increasing control exerted over the media, Internet and civil society (Oates, 2013; Soldatov and Borogan, 2015; Wijermars and Lehtisaari, 2020).

A case in point is the Yandex.Novosti (“Yandex.News”) aggregator — the Russian equivalent of Google News, launched in 2004 — which is the focus of this article. When they first appeared, search engines and recommendation systems such as aggregators were designed as tools which would make the diversity of content on the Web more manageable. As a vast body of research has shown however, these platforms occupy a strategic place and have become key intermediaries in channeling information to end users *qua* citizens. Thus, they wield a form of power in shaping the perception of social reality, which scholars, policy-makers and civil society alike are still in the process of defining precisely. As a contribution to this effort, we would like to shift the perspective towards the less-studied Russian context. Our research shows how the Russian authorities have attempted, both directly and indirectly, to discipline the news recommendation system provided by Yandex as part of a wider endeavour to assert control over the circulation of information.

We first present the role and responsibilities of news aggregators in digital media ecosystems. Drawing on the more familiar model of Google News, we spell out the issues raised by automated news recommendation, as part of a larger set of questions concerning the role played by search engines and social media platforms in ensuring media diversity (Helberger, 2019) and in shaping the public sphere through “algorithmic gatekeeping” (Napoli, 2014; Nechushtai and Lewis, 2019). Algorithms are often criticized for their opaqueness and unaccountability (Pasquale, 2015; Saurwein, *et al.*, 2015) and despite long-standing claims of neutrality on the part of these actors, it has now become increasingly clear that issues related to news curation are not merely technical or commercial.

We then engage in a critical analysis of the Yandex.News platform and its algorithm (Kitchin, 2017; Seaver, 2019), which relies on two different types of evidence: controversy analysis and algorithm auditing.

In the second part we present the political controversies triggered by Yandex.News during the 2010s. Drawing on social studies of science and technology (STS), we consider controversies associated with socio-technical systems as a privileged path of inquiry (Latour, 2005; Marres, 2007). We look at the attention garnered by the aggregator and the objections raised by policy-makers and end-users, particularly focusing on accidents, disturbances and alleged malfunctions; they represent key entries towards understanding the role of this algorithmic “black box” as a contested producer of meaning (Bucher, 2016). In 2016, laws targeting news aggregators compelled Yandex to restrict the results displayed on its news page to officially registered media (Daucé, 2017). Drawing on interviews with current and former members of the Yandex company, with Search Engine Optimization (SEO) professionals, and with

Russian journalists and editors confronted with the platform, we show there are three rival interpretations of what constitutes the “right” aggregation results. The first one is defended by Yandex itself as an “objective” output of its algorithms. Another is the criticism raised by the authorities, for whom Yandex.News may promote “unpatriotic”, “fake” or otherwise problematic news. The last one is put forward by the journalists, editors and SEO professionals who rely on their sense of what can be considered “newsworthy” to criticize the platform.

In a third part, in order to test the algorithm itself, we compare Yandex.News’s huge database of partners with the news rankings provided by the algorithm for the Top 5 news items which are presented on the Yandex homepage [2]. We show the discrepancy between the diversity of the Russian online mediasphere in the database, and the narrowness of the media sample represented by Yandex.News rankings in 2020, exclusively dominated by a small set of only 14 media outlets (news agencies, state-funded media and private publications which are “loyal” to the government). We then contrast the narrowness of the aggregator’s results with the diversity of content circulating on social networks. This contrast suggests the existence of two online media spheres: the first one overwhelmingly dominated by the “registered” media aggregated by Yandex.News, and a second one relying on social networks where “alternative” media sources can carve out a space.

This research provides a case study of “governance by algorithms” (Musiani, 2013; Just and Latzer, 2017; see also Gillespie, 2018), and contributes to understanding how news recommendation systems can be prone to certain biases (Kulshrestha, *et al.*, 2019; Trielli and Diakopoulos, 2019) or may be susceptible to political filtering in subtle ways (Jiang, 2014). The Yandex case also provides insights into the specificities of Russian Internet policy as an assertion of “digital sovereignty” (Nocetti, 2015; Musiani, *et al.*, 2019). It exposes the new “codes of conduct” — involving both computer code and legal code (Lessig, 1999) — which may be set up in the networked public sphere in contemporary societies. It is revealing of both the strategic nature of search platforms and algorithms today (Pasquale, 2015) and the ways in which the Russian government increasingly seeks to assert its dominance over Internet governance understood as a dimension of “information security” (Maréchal, 2017), illustrating how information infrastructures can be held in check in order to indirectly regulate online speech (Sivetc, 2019).



1. Algorithmic gatekeeping in digital media ecosystems: Context and issues

1.1. Automated news recommender systems, news diversity and the shaping of the public sphere

The Yandex.News aggregator can be described as an *automated news recommender system*. The most well-known example of such a service is the Google News aggregator, which was first launched in 2002 and taken out of beta in 2006 [3]. Initially the service aimed at providing a broad overview of trending news, by presenting the user with “clusters” of related articles. Gradually, other languages and country-based editions were developed, and features such as e-mail alerts, personalisation and recommendation of news stories were added. In

2021, the service indexes tens of thousands of news Web sites across the world and is interwoven with the main Web search service.

Recommender systems exist since the early days of the Web and cover a wide range of different applications, from shopping and e-commerce to music, movies and news (Jannach, *et al.*, 2011; Ricci, *et al.*, 2015). They involve automated filtering, which may rely on various parameters but generally draws on users' actions (backlinks, clicks, searches, choices, preferences etc.) to approximate a selection of relevant information which this same user or other users will likely be interested in. News recommender systems rely on various methods, which can be based primarily on an analysis of the content itself (the nature of the publications, including for instance its "freshness"), on the activity it generates (click rate, engagement metrics on social media such as likes and shares etc.), on forms of "collaborative filtering" and interest patterns in a given community, or on the users' preferences — this latter case involving a personalization of the aggregated news items based on collected behavioural data, which can be self-determined or inferred (Karimi, *et al.*, 2018). Generally, news recommender systems combine these approaches to different degrees, depending also on whether the user can be easily tracked (*e.g.*, is logged in or otherwise identified as a unique user).

Beyond assessing relevance and pushing content for which others have shown an interest or content which is similar to one's previous interests, recommending news items also requires providing a *qualitative* selection. This dimension is more difficult to define, since it involves fetching items which might not be within a users' direct scope or attention. One key measure of quality is based on the *diversity* of the news which an aggregator is able to provide. Such a criterion may be assessed in terms of user satisfaction; however, it also involves a much wider discussion regarding the role and responsibility of the media as a central institution of democracy, expected to adequately inform citizens as well as providing a diverse public forum for debating ideas and opinions (Helberger, 2019). Moreover, the key parameters of this diversity include dimensions such as plurality of topics covered, as well as variety in editorial policies, ideological perspectives, narrative genres etc. (Helberger, *et al.*, 2018). Another related concern is whether automated recommender systems may be *biased* against certain types of content, underrepresenting them in the results (Kulshrestha, *et al.*, 2019). Indeed the overall picture may appear diverse but key issues may be left out (*e.g.*, a major corruption scandal), certain topics may bear comparatively less weight (*e.g.*, politics vs. sports), or their editorial treatment may downplay their importance (*e.g.*, by focusing on less crucial but more entertaining aspects).

Very early on, the "politics of search engines" has been presented as a crucial issue with decisive implications for the shaping of the public sphere (Introna and Nissenbaum, 2000). News aggregators which provide visibility to "automatically" selected news items, have also come under scrutiny for their increasingly central role and the growing power which they have harnessed within media ecosystems. There are now many different channels through which news is distributed, which include search engines and news aggregators, as well as social media. As a result, publishers have become increasingly dependent on digital platforms (Nielsen and Ganter, 2018). All such digital intermediaries have, in effect, themselves become gatekeepers alongside journalists (Napoli, 2014), often leveraging user behaviour to shape an overall picture of "curated flows" of information (Thorson and Wells, 2016, 2015).

Google, but also Yandex, have generally presented their services as "neutral", but such claims to objectivity have been criticized for different reasons. For the past decade, because of their

increasingly powerful personalisation features, some of the main Web services and particularly Google's search engines have been suspected of entrapping users in "filter bubbles" and "echo chambers" (Pariser, 2011; Bozdag, 2013). By making users oblivious to certain types of information or to alternative perspectives, and sometimes reinforcing existing prejudices or biases, these services would be in the process of undermining the public sphere. Search algorithms and automated recommender systems have also been criticised for promoting outrage and conspiracy theories, with for instance the YouTube recommendation algorithm being presented as "the great radicalizer" [4]. However, the reality of these phenomena is difficult to assess precisely (Flaxman, *et al.*, 2016; Bruns, 2019), particularly for search engines (which have even been shown to increase information diversity, see Fletcher and Nielsen, 2018). In the case of Google News, even personalisation features don't seem to reduce news diversity (Haim, *et al.*, 2018). However, although *individual* filter bubbles may be difficult to assess empirically, "algorithmic news curation still represents a concern for source diversity since it can concentrate societal attention on a narrow range of privileged outlets." [5].

1.2. Disciplining the algorithm to control the news?

Before looking more closely at Yandex.News, it is worth noting that the relationship of news outlets with Google News has been complicated. News aggregators provide visibility for news content in exchange for access to publications, and although Google News is a purveyor of traffic (Calzada and Gil, 2020), it can also be perceived as an outlet in its own right benefiting from the content (titles and "snippets" of text) provided by the news media [6]. Beyond these copyright and business model issues, news aggregators affect the publication of the news itself: content must comply with criteria valued by their algorithms in terms of relevance, "freshness", frequency of updates, metadata, backlinks, mobile friendliness, etc. Editors deploy search engine optimization (SEO) strategies to ensure their content is efficiently referenced and promoted, constantly monitoring audience analytics to understand what "works" and what does not, which stories gain traction while others do not. Depending on how heavily a news Web site relies on it for traffic (and thus for advertising revenues), it will need to tailor its content and adapt its publication strategies in order to be "picked up" by the platforms — which directly affects the work of journalists and conventional understandings of "news value" or "noteworthiness" (Boyer, 2013; Belair-Gagnon and Holton, 2018; Diakopoulos, 2019a).

The algorithms deployed by these platforms can therefore be perceived as an "invisible hand" deciding which topics will be singled out as relevant and which news outlets will be pushed on the forefront according to sometimes unfathomable criteria — profoundly affecting the nature of journalism itself in the process, as professionals adjust the form and nature of their published content in order to satisfy these constraints (Brake, 2017; Christin, 2020). The role played by algorithms in sorting news items, directing visibility and attention, framing issues and setting agendas, is increasingly questioned — particularly considering that

"Trending algorithms may reflect back what is popular while raising awareness among an even broader set of people, in effect helping to conjure an interested public around an issue. On the other hand, questions may reasonably be raised when a newsfeed fails to notify its users of important civil unrest while continuing to amuse and divert attention to popular events." [7]

The impact of search engines on user choices and preferences can be far reaching, and in a series of controlled experiments it has even been shown to sway undecided voters (Epstein and Robertson, 2015).

Given its central role in the distribution of news today, Google is regularly suspected of providing visibility to illegitimate sources (*e.g.*, the controversial image board 4chan after the Las Vegas shooting in 2017) [8] or of being politically biased in favour of certain types of (“left-leaning”) news sources (Diakopoulos, 2019b). It has also been claimed that traffic referred by search engines seems to benefit mainly a small number of already highly visible national news providers — thus entrenching already-existing media hierarchies and undermining claims to greater pluralism (Hindman, 2018, 2008; Hong and Kim, 2018). This may also be the case with news aggregation: it was recently shown that for Google News, only five news organizations account for nearly half of all recommended news items and legacy media dominate the recommendations (Nechushtai and Lewis, 2019; see also Bui, 2010).

In the Russian political context, the issue raised by the Yandex.News aggregator is even more acute: could it be manipulated for political reasons, either through direct interference with the results or by fooling its algorithm? Or could it be disciplined through regulatory constraints to achieve similar results? Addressing these questions involves assessing not only whether the available body of information is sufficiently diverse, but whether some news considered important for the general public may, in certain circumstances, be intentionally prevented from reaching wide visibility through the aggregator. In practice, this goal could be reached if the automatically “recommended” news only stemmed from controlled sources.



2. Yandex.News as political controversy: Defining the “right” news aggregation

Yandex.News was specifically targeted by different policies and legal initiatives since 2014. We thus provide an overview of the new regulations and the discussion they generated in the press. Moreover, suspicion towards the service was compounded by observations made by journalists and users, who during specific events found a discrepancy between their shared sense of what counted as news, and the automated selection provided by Yandex. We therefore also recount instances of what was considered suspicious activity, based both on published journalistic investigations and on semi-directed interviews we made with key actors, such as journalists, editors, SEO professionals and Yandex former employees. The controversy leads some of them to develop strategies to circumvent the platform.

2.1. Yandex.News in the Russian context

Yandex (a contraction of Yet Another iNDEX) is the name given by Arkady Volozh and Ilya Segalovich to full-text search technologies supporting the Russian language which they developed at Comptek International. The Yandex.ru search engine was launched in 1997 — at about the same time as Google, but in a very different economic context — and contextual advertising was added the next year. However, Russia is one of the very few countries where search is not monopolised by Google, with Yandex owning a share of over 45 percent — thus making it a prized national champion of the Russian digital economy. In 2000, Yandex became an independent entity, and it is now a private globalized company incorporated in the Netherlands. It has been listed on the NASDAQ since 2011. It is Russia’s largest tech company, and its revenues have more than tripled in five years, from 60 billion rubles in 2015 to 218 billion rubles in 2020 [9].

According to journalists in 2017: “On the surface, Yandex and the Kremlin do represent two different Russias with little overlap” [10]. The co-founder of Yandex, Ilya Segalovich, and some of his colleagues actively participated in protests against the results of the elections in late 2011 and 2012. Since then however, the Russian government has increased its pressure on the company (Oates, 2013). Yandex is considered a key national asset, ensuring a degree of independence from foreign (especially American) Web companies. Its activities are constrained by political, legal, technical and economic means (Vendil Pallin, 2017). In 2009 Yandex’s owners sold a “golden share” (priority share) to Sberbank, the state-controlled Russian savings bank — giving the government *de facto* veto powers over strategic issues. However, diverging strategies recently led to the transfer of this “golden share” to a newly created Public Interest Foundation — a restructuring of its governance meant to reassure both its investors and the Kremlin: the move ensured that Yandex would not fall under foreign control, while still allowing it to operate on global markets [11].

Although it has been criticized, a certain degree of loyalty towards the state has also brought the company certain advantages. For instance, Yandex won an antitrust conflict with Google and is expected to benefit from a new Russian law, that makes pre-installation of Russian software on smartphones mandatory [12]. According to this law, Google will not be allowed to pre-install its apps on Android phones, giving Yandex room to develop its products.

Beyond its Web search engine, Yandex has developed a host of other services such as an Internet portal or an e-mail service, as well as more specific ventures such as Yandex.Taxi (ride-hailing, which merged with Uber in 2017), Yandex.Karty (maps and geolocalisation), Yandex.Music (music streaming) or Yandex.Eda (food delivery). Among them is Yandex.Novosti or Yandex.News, the news aggregator which is the focus of our study.

Yandex.News presents a selection of topics and articles which purports to reflect the themes most widely covered by the media at a given moment. To do so, it processes the information published by a range of (mainly Russian) online media. Yandex.News was launched in 2004 and was initially a pilot project led by a team of computer scientists and linguists, who had been hired to develop named-entity recognition and extraction in the news [13]. The Yandex.News team claims that the algorithm works in the absence of human intervention. News from the partners is gathered into topics through the algorithms clustering process. The robot analyzes keywords and facts and groups them by topics, using three main criteria: citation rate, recency and informativity [14].

Lev Gershenzon, the former head of the service, graduated from the Chair of Theoretical and Computational Linguistics at the Moscow University of Social Sciences (RGGU). He joined the Yandex.News Development Team in 2004, when Yandex as a whole only employed about 200 people. At that time Google News already existed, but, according to Gershenzon, “At Yandex, we were stronger and more attractive for users. That is why we placed the Top 5 on the main page and Google never tried to do that” [15]. Indeed, one key dimension of the discussion concerning Yandex.News is that a Top 5 of its aggregation results is constantly presented on the Russian version of the Yandex homepage, just above the search box — ensuring this small selection of news a massive daily viewership, and driving considerable traffic towards the featured publications (see [Figure 1](#)).

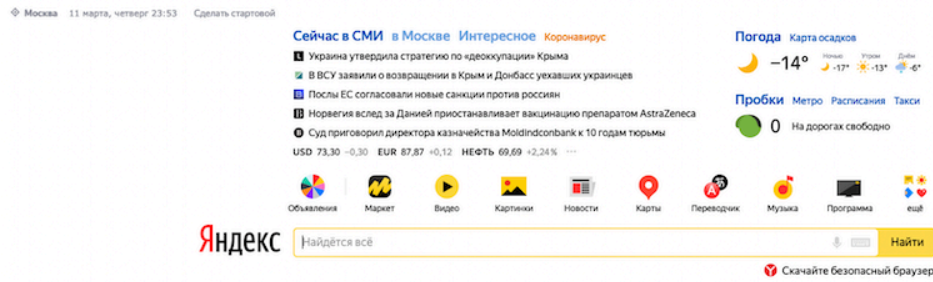


Figure 1: Top of the Yandex homepage, screenshot, 11 March 2021.

Though Google, for instance, also provides a “Top Stories” selection within its search engine, these are relevant to specific searches and only appear along with the rest of the results once a query has been typed. The Top 5 is therefore of particular relevance in understanding how Yandex as a platform functions as a news recommender system. In 2017, according to Grigori Bakunov, Yandex Technical Director, “The daily audience of the five news items that appear on the Yandex homepage is the same as the homepage — approximately 20 million people, depending on the day. Six million visit the Yandex.News page daily” [16].

2.2. *Legal constraints: Domesticating news curation*

The controversies which arose after 2012 put an end to the belief in the objectivity of the aggregator. 2012 was a decisive year for freedom of expression in Russia, and a “watershed moment” in Internet regulation (Lonkila, *et al.*, 2020). For a decade, the public sphere had seemed to be thriving and had remained relatively free — especially online. This culminated in a bold outburst of contestation following the 2011 general elections, which had seen fraud on a massive scale. The wave of protests — which also followed the Arab Spring — triggered a harsh response from the authorities, who launched a gradual tightening of the rules governing public expression (Denisova, 2017).

One of the most significant measures was the establishment by law in 2012 of a blacklist of censored Web sites, which ISPs were now required to prevent access to [17]. It originally targeted child pornography, information related to drug production and distribution, and information encouraging suicide. However, the notion of “prohibited information” was extended to “incitements to illegal action” or “promoting extremism”, and in 2014 was used to block opposition Web sites such as Grani.ru, Kasparov.ru or the Livejournal blog of opposition leader Alexey Navalny. The list is managed and regularly updated by Roskomnadzor, the communication and media watchdog which also grants licenses for mass media in Russia. Moreover, these Web sites were no longer allowed to be referenced by search engines. Yandex therefore had to remove them from its search results.

Control over the public sphere stepped up again in 2014, during the conflict with Ukraine and the occupation of Crimea. Dissenting views were scrutinized in a context of strong patriotic rhetoric. Media campaigns were launched by the authorities, involving both state media and a now infamous “troll army”, in order to fuel support for pro-Russian opinions and undermine the credibility of any pro-Ukrainian voices (Fedor, 2015; Mejias and Vokuev, 2017). At that time — in a global context which included the Snowden revelations of dragnet surveillance of

the Internet by U.S. intelligence agencies — Putin declared that Yandex had begun as a project under Western influence, and that the Internet in general was a “special CIA project”.

Yandex.News, in particular, was at the heart of a political controversy, after being accused of partiality by the authorities for providing visibility to information which didn't align with the official narrative. The site Pravda.ru wondered if “Yandex lights a ‘Maidan’ in Russia?” (referring to the protests in Kiev which led to the regime change in Ukraine) [18]. The newspaper was outraged by the headlines chosen by the news aggregator and claimed that it was necessary to legally regulate its activity. In May 2014, Putin's press secretary, Dmitry Peskov, declared that Yandex.News should be registered as a mass media, which would put it under the control of Roskomnadzor.

This led to the adoption in 2016 of a law on news aggregators, designed to extend control to such intermediaries and specifically targeting Yandex.News [19]. News aggregators receiving over one million daily visitors became legally responsible for any content published in their results (and at risk of heavy fines in case of violations), unless the selected media are officially registered with Roskomnadzor. Formal agreements were set up between Yandex.News and the media, with 6700 new “partnerships” established in 2016. The law went into effect on 1 January 2017, and as a consequence all non-registered media (including dissenting voices such as Mediazona) as well as all foreign media (such as the BBC in Russian, as well as exiled media such as Meduza) disappeared from the Top 5 results presented on the Yandex homepage as well as from Yandex.News.

After the law was adopted, Tatyana Isaeva, head of Yandex.News since 2012, announced she was leaving the company. She argued in an interview with *Meduza* that her work was made less interesting by the law and that the very aims of the aggregator — highlighting important news and making different points of view available to the user — were undermined: “The aggregator is really meant to cover the [news] picture with one glance. If this picture is the same in all its parts, it is absolutely unclear why an aggregator is needed” [20]. Isaeva also mentioned that the situation in Crimea was a critical moment for the aggregator, which sourced news coming from both Russia and Ukraine, and presented radically different perspectives on the conflict:

On one page headlines were collected that directly contradicted each other. And it was absolutely unclear who we were talking about at all, when in one and the same story the same people in one headline were called separatists, in another headline militia, in the third defenders, and in the fourth ... some bad words. And it happened so because technically the stories of the Russian issue could have been covered by the Ukrainian media on absolutely the same grounds as the headlines of the Russian media. And this, in general, was quite justified until now, because journalists cover very similar topics in Russia and Ukraine. And when we saw this outrage, we understood that we distort people's perspective, because the service was not ready for such a turn — for information warfare, there were no precedents. [21]

As a consequence, the service was made “impossible to use”, and Yandex decided to divide the media between “domestic” and “foreign” for each of the two countries — and then make sure that the “domestic” media would appear higher in the results: registered Russian news outlets in Russia, and Ukrainian sources in Ukraine.

Despite the adoption of the law, the Yandex aggregator is still criticized by the authorities. In August 2019, Yandex was accused by Russian Duma deputies of spreading “fake news” after an inaccurate story by the daily newspaper *Kommersant* made it to the top of the

Yandex.News selection. The article stated that the Duma was considering a ban on the use of old vehicles, sparking widespread outrage and leading deputies to issue clarifications, according to which this was a recommendation which concerned only professional vehicles and not personal ones. However, the news stayed on top of the selection even after it had been strongly denied, leading some deputies to take aim at the news aggregator. Andrey Isyaev accused Yandex of “deliberately exacerbating the social and political situation” and of “foreign interference”, while Adalbi Shkhangoshev said he had called Yandex CEO Yelena Bunina and asked her to revise the news aggregator’s algorithm — in response to which Yandex threatened to close its aggregation service [22]. This new scandal arose in a difficult political context for the government, with opposition movements demonstrating during the summer to denounce the refusal to let them register at the local Duma elections.

2.3. Gaming the algorithm? Criticism and suspicion towards Yandex.News

For several years now, journalists and political activists have also been claiming that the Yandex.News rankings are biased — but for different reasons. According to them, the biases may come from the aggregator itself, which undermines information about the opposition, or from official actors who have learned how to mislead the algorithm.

According to one of our interviewees, a journalist from *Kommersant*, “the algorithm is used by power-dependent newspapers that publish news which are pushed up by the algorithm” [23]. These techniques involve pushing the boundaries of search engine optimisation (SEO), and gaming the Yandex.News algorithms by artificially creating multiple sources of information. They can be understood as exploiting vulnerabilities of the platform in order to engineer increased visibility. In a similar way, “junk news” Web sites have been shown to increase their “discoverability” in Google Search (through keyword optimisation this time) for disinformation purposes (Bradshaw, 2019). Such attempts at shaping search results are in fact as old as search itself, and algorithms are normally updated at regular intervals to counter these manipulations — which keep cropping up however.

During the Moscow City Duma elections in 2014, for example, according to an investigation by RBK, dozens of unknown media Web sites wrote the same news about the alleged successes of Sergey Sobyenin, Moscow mayor since 2010 and candidate for re-election — and the news made its way to the top results of Yandex.News. The Moscow mayor’s office had learned how to influence the service [24]. First, several hundred district newspapers, city Web sites and government agency Web sites (many of them recently created) had been registered with Yandex’s “Database of media and official sources”. Then, any positive news in favour of the Moscow mayor was sent by the Moscow Information Technologies OJSC (owned by the Moscow authorities) to publishing companies. These would rewrite the news to ensure the articles wouldn’t be identified as duplicates by the Yandex algorithm, before being published on these local Web sites.

Independent journalists and political activists denounced the progressive hold on the aggregator by actors defending a “patriotic” stance (legislators, regional administrations, official media ...). Leading defenders of online freedoms who initially supported the Yandex.News aggregator became sceptical that it could remain free from political interventions. They have been collecting evidence of its biases and denounce its partiality.

The “law on aggregators” in 2016 entrenched this partiality by severely restricting the news sources Yandex could promote. For example, in March 2017, major demonstrations took

place in Moscow and across Russia. The initial protests flared up after Alexey Navalny's Anti-Corruption Foundation (FBK) released an investigative video showing multiple examples of alleged embezzlement by ex-Prime Minister Dmitry Medvedev. In Moscow, a thousand people were arrested — but surprisingly, according to journalist Alexey Kovalev, the protests were hardly mentioned on the main Yandex.News page and even on the local news feed for Moscow [25]. Yandex argued that its algorithms hadn't been tampered with, but that this was a consequence of the law on aggregators which had considerably reduced the number of available sources — thus illustrating how such a legal initiative could function as an indirect control mechanism (Wijermars, 2021).

Political suspicions against the news aggregator resurfaced in April 2020. *TJournal* (a Russian Web site devoted to technology) showed that the Yandex search engine and the Yandex.News services only returned negative content when searching for Alexey Navalny [26]. When questioned by *TJournal*, Yandex said that the priority given to negative publications about Navalny was an “experiment”. In 2021, after Navalny returned from Germany and was sentenced to prison, demonstrations broke out in Moscow. The activists we spoke to underlined the discretion of Yandex.News concerning these events. An activist close to Navalny already considered in 2018 that: “Yandex.News transmits state propaganda. The aggregator has been destroyed (*sloman*)” [27]. Some of these observers point out that Yandex avoids conflicts with state authorities so as not to jeopardize its multiple activities (Yandex.Taxi, Yandex.Eda etc.) [28].

All these suspicions lead to a de-legitimation of the algorithm in the eyes of journalists, as well as Web professionals and programmers who seek ways to bypass it. Some of them now consider the service to be useless. As Lev Gershenzon remarked in 2016: “Aggregators make sense (...) only when there is something to aggregate. If all independent, interesting, professional publications on a federal scale can be counted on the fingers of one hand, rocket technology for their aggregation and processing is not needed — you can simply add them to your bookmarks” [29]. The idea of closing the service seems to have been considered by Yandex executives themselves. According to well-known journalist A. Plyushchev, from *Ekho Moskvy*: “Well, you know, I once talked to A. Volozh, the head of Yandex, and that was before the law on aggregators was passed. And he told me that if the law was adopted, he would close the service. (...) Well, the law was softened a bit, and the service, as you can see, did not close. I still doubt if that was the right decision. Because, well, I think, unfortunately, the state did everything possible to manipulate both the media and extraction in search engines” [30]. In May 2018, in an open letter to Yandex CEO Yelena Bunina, A. Plyushchev advised her to shut down the Yandex.News service or to rename it Yandex.Propaganda [31].

Another scenario concerns the development of an alternative aggregator. In 2019, from abroad, Telegram founder Pavel Durov announced his intention of developing a news aggregator on his platform: “We have a chance to create the first effective and free news aggregator in the history of the Internet,” said Durov. “We can start recommending articles from the Recommended Articles block after reading each article in Telegram, gradually bringing it into service with an hourly selection and a global search on all the news in the world” [32]. P. Durov announced his aggregator will be beyond the control of the Russian security services and political censorship, unlike local operators [33]. He invited Yandex.News developers to participate in the creation of his service [34] and announced a competition to develop an algorithm from the Data Clustering Contest [35].

To summarize, the aggregator claims to be neutral and objective but, on the one hand, the authorities denounce its propensity to relay discontent and destabilise the political situation while on the other hand, journalists, Web professionals and activists underline its institutional framing to promote a “loyal” agenda. Neither the law nor Yandex’s rules allows one to understand how the algorithm works. This uncertainty about the relative weight of each news in the algorithm opens a space of controversies for the actors who interact with the algorithm.

3. The Yandex rankings as a gateway to the algorithm and its transformations

In this section, we present an audit of the aggregation algorithm (Kitchin, 2017). In line with similar approaches where direct access to a “black-boxed” algorithm or the data processing itself is impossible (Jiang, 2014; Robertson, *et al.*, 2018; Nechushtai and Lewis, 2019; Trielli and Diakopoulos, 2019), it is based on an analysis of the declared input on the one hand, and the observable output on the other. We thus relied both on the news sources allegedly used by Yandex.News and on the results provided by its Top 5 feature on the Yandex homepage, collected over several periods of time as detailed below. We then used this information to assess the diversity of sources in the featured publications as a key dimension of news pluralism, using diversity of sources in this case (and due to the strong prescriptive effect of the Yandex Top 5) as a proxy for exposure diversity — *i.e.*, the attention collected by these sources (Helberger, 2012; Helberger, *et al.*, 2018).

3.1. From the Yandex.News database of sources to its rankings

Though it is difficult to investigate the algorithm itself, one can look at the input that the aggregator feeds on. The database of “partner” media which Yandex.News officially draws upon is publicly available online [36] and was first published in March 2004. At that time, it included 460 references (news agencies, online media, print titles, radio and television Web sites) [37]. In December 2019, it included 7,107 Web sites that constitute the media panorama in which Yandex.News operates. By way of comparison, Google News lists 4,500 English-language sources. The Yandex database does not only include media officially registered with Roskomnadzor but brings together registered or unregistered media, pro-state or independent sources, Russian and foreign content, public and private Web sites. The complete database is very heterogeneous and representative of the diversity of the Russian Internet (Table 1).

Media	Daily newspapers	220	3%	56%
	Monthly newspapers	212	3%	
	Weekly newspapers	661	9%	
	Weekly magazines	78	1%	
	News agencies	481	7%	
	Radio	65	1%	
	Online media	1,957	28%	

	TV channels	264	4%	
Other online sources	Official sources	520	7%	44%
	Thematic Web sites (politics, society, sports, culture ...)	2,631	37%	
	Other	18	0%	
	Total	7,107	100%	

Concerning the media sources, the database includes officially registered media such as daily newspapers *Rossijskaia Gazeta*, *Izvestia*, *Moskovskij Komsomolets*, state news agencies (RIA Novosti, TASS ...), Web sites of TV channels and online media as well as administrative sources and thematic commercial Web sites. Moreover, some newspapers registered all their local newsrooms, potentially giving them more weight in the database. *Komsomolskaia Pravda* is registered 55 times with all its local branches. The same phenomenon is obvious concerning *Sputnik* (15), *Moskovskij Komsomolets* (19 references), RIA (28) or the GTRK official TV channel (46). In Moscow, local online media are registered in every district. Of the media considered more liberal, only *Kommersant* (19) and RBK (18) have a substantial network of editorial offices in the regions. The other media usually have only one editorial office in Moscow. The importance of the regional media registered in the database allows Yandex.News to offer regional news rankings. It can also possibly raise particular news in the rankings through the relay effects of regional newsrooms, as discussed earlier in the case of the Moscow City Duma elections of 2014.

Independent media also remain in the database, and this even includes opposition media that are not accredited with Roskomnadzor such as *Meduza*, *Mediazona*, *Dozhd'* TV (tvrain.ru), human rights NGO *OVD-Info*, the press service of exiled former oligarch and political opponent Mikhail Khodorkovsky, or even Kasparov.ru and Grani.ru Web sites which were banned by Roskomnadzor in 2014. However, as the editor in chief of Grani.ru Yulia Berezovskaya explains: “It is true that Yandex.News chose not to kick us out despite the ban (unlike Rambler News) but they have ‘diminished’ us so much (...) that our presence in this aggregator is only symbolic” [38]. The public database is therefore a remnant of digital freedoms from the 2000s. Since 2016, “On the main page of the service and in the top on Yandex.News, you can only show publications who have media registration. Those who do not continue to appear in Yandex.News search results — there are about seven thousand such sources” explains Grigorij Bakinov from Yandex [39].

The database reflects the diversity of the Russian digital space, from a geographical, thematic, and even political point of view. It probably allows the operation of various services offered by Yandex, including regional news ranking and information services customized by the Yandex.Dzen application, a personal recommendation service that creates a feed of content that automatically adjusts to the interests of its users since 2016.

In 2019, the database provided little information on the relative political weight of the media which can appear in the Top 5 of Yandex.News. To the question “Where can we know the weight (of a partner)?” Yandex answers “nowhere”. Because we are unable to access the algorithm itself, which remains secret, the analysis of its ranking choices allows a better understanding of the outputs produced by the robot.

During the month of June 2020, we conducted a quantitative analysis of the news selected by Yandex.News and presented as the Top 5 news on the Yandex homepage (Table 2). We carried out a systematic scraping of news: between 1 June and 30 June 2020, we automatically collected the Yandex.News rankings every two hours and listed a total of 3,011 references [40]. The data was collected on a server based in France, but we controlled for possible personalisation and localisation features by checking at different moments the news items with results appearing for a user based in Russia and found no difference. It appeared that, during this period, only a small group of 14 media outlets were cited in the Top 5 — an extremely narrow sample considering the over 7,000 sources listed in the Yandex.News database. We then extended the scraping to the period June–December 2020 and obtained the same results, with the same 14 media appearing in the Top 5 over this period. Finally, we categorized these 14 media according to their type, the nature of their ownership and their general editorial positioning, as detailed below and colour-coded in the table.

Table 2: Media cited in the Yandex.News rankings (1–30 June 2020 and June–December 2020).				
Table colour code				
State media	Private “loyal” patriotic media	Private “loyal” liberal media	Independent registered media	Independent non- registered media
Media outlet	Type	Ownership	Number of citations 1–30 June 2020	Number of citations June–December 2020
RIA Novosti	Press agency	State	494	3,816
Gazeta.ru	Online newspaper	Private (Rambler Media Group/V. Potanin)	386	3,282
Izvestia	Newspaper	Private (National Media Group)	352	2,666
RBK	Newspaper	Private (Grigori Beryozkin)	296	2,211
Lenta.ru	Online newspaper	Private (Rambler Media Group/V. Potanin)	267	2,010
RT in Russian	Online TV channel	State	255	2,222
Kommersant	Newspaper	Private (Alisher Usmanov)	204	1,593
Regnum	Press agency	Private (Boris Sorkin)	179	1,569
Rossijskaia Gazeta	Official government newspaper	State	151	1,353
TASS	Press agency	State	115	1,271

Vesti.ru	News service on television, on radio and online	State	96	767
Vedomosti	Newspaper	Private (Ivan Eremin)	81	1,005
BFM.ru	Radio	Private (Yegor Altman)	67	622
Interfax	Press agency	Private	56	751

The data strikingly shows the concentration of information on Yandex.News among a few large media players: public press agencies, state-funded media, leading newspapers and mainstream online publications. An over-representation of specific news publishers has also been shown to exist in the case of Google News (Schroeder and Kralemann, 2005; Haim, *et al.*, 2018), but not to such an extent. This is a much narrower range than the results observed by Nechushtai and Lewis (2019) in the case of Google News in the U.S. for instance where, although a small selection of 14 outlets also dominated the aggregator, a long tail of other publications also figured in the results. Trielli and Diakopoulos (2019), looking at the Google Top Stories box in the U.S., found that only 20 news sources accounted for over half of the featured articles and that a “left-leaning ideological skew” could be observed in the selection of sources; however, again, a considerable long tail of over 650 other news sources appeared at least once in the remaining half of the total 6,302 links collected over a one-month period.

Moreover, although nuances can be detected between these 14 major media, what appears clearly is that in 2020, “officially sanctioned” media reached Yandex’s heights more easily. Indeed, most publications within the 14 selected outlets are related to the Kremlin: they are either directly funded by the state, or are privately owned by “loyalist” figures or entities and thus indirectly “managed” by the authorities. Since 2014, the Russian media panorama is usually divided between “state-owned” and “independent private” media [41]. The boundaries of these different categories are blurred and debatable. Indeed, if state-owned media are clearly identified (RIA Novosti, RT, *Rossijskaia Gazeta*, TASS, Interfax), “independent private” media are more difficult to classify. To facilitate the reading of [Table 2](#), we propose here two categories to qualify them. “Private loyal ‘patriotic’ media” refers to general news media which have been transformed from within by the departure and replacement of their editorial teams between 2012 and 2014. They officially remain as “facade” but have undergone hostile takeovers, with their teams replaced by journalists who are loyal to the authorities (Chupin and Daucé, 2017; Daucé, 2020; Kovalev, 2020). This is mostly the case with Lenta.ru or Gazeta.ru. “Private loyal ‘liberal’ media” refers to general news and business media which used to be considered “liberal” but whose political staffs were reorganized between 2018 and 2020 (this mainly concerns RBK, *Kommersant*, *Vedomosti*). Yandex.News works in a context of economic and political reshuffling of the Russian media space, whereby different types of constraints have led the media spectrum to be narrowed down (Wijermars and Lehtisaari, 2020). As a consequence, the main sources used by Yandex.News, which the algorithm builds on to paint a picture of the daily news on the Web, have been profoundly altered. Conversely, independent media outlets are increasingly sidelined as they do not benefit from the traffic referred by Yandex.News and thus also the advertising revenue, reducing exposure diversity but also making them less viable commercially (Kovalev, 2020; Wijermars, 2021).

3.2. Yandex.News rankings facing Russian social networkss

In Western countries, “social media are becoming central to the way people experience news” (Hermida, *et al.*, 2012). In Russia, in the beginning of the 2010s, a convergence took place between online news and social media (Pancenکو, 2011). Since 2016, however, we observe a growing gap between how people experience news on Yandex.News and on social networks. To show this differentiation of media spaces, we relied on different sources: the audience analysis of the company Medialogia (which measures both the media citation rates and their circulation on social networks) [42], surveys of news consumption by the Levada Sociological Center [43], and metrics produced by media themselves.

Since January 2017, Medialogia offers two media rankings. The first one, the Citation Index ranking, is based on Medialogia’s media database, which includes about 43,400 sources: TV, radio, newspapers, magazines, news agencies, online media and blogs. News aggregators are not taken into account when calculating this ranking. The second is based on hyperlinks shared and commented on social networks (Twitter, Facebook, VKontakte etc.) [44]. As a representative of Medialogia explains:

The second ranking “emerged in response to a request from media outlets to compare their performances on social networks. The Citation Index ranking shows the quality of content and credibility of the media in a professional environment, while the social media data reflects users’ interest in and trust in the media’s content.” [45]

In June 2020, the Citation Index ranking was fairly consistent with the rankings of Yandex.News, whereas the social media data differed significantly (see [Table 3](#)).

Table 3: Fourteen most cited media in the Yandex Top 5, in the Medialogia Citation index and on social networks in June 2020.			
June 2020	Yandex rankings (as scraped by the authors)	Medialogia: Citation index in mass media	Medialogia: Hyperlinks shared on social networks
1	RIA Novosti	TASS	RIA Novosti
2	Gazeta.ru	RIA Novosti	Openmedia.io
3	Izvestia	<i>Interfax</i>	Meduza.io
4	RBC	RBC	Russian.rt.com
5	Lenta.ru	<i>Izvestia</i>	MBKh Media
6	RT in Russian	<i>Kommersant</i>	RBC.ru
7	<i>Kommersant</i>	Russian.rt.com	TASS
8	<i>Regnum</i>	<i>Rossijskaia Gazeta</i>	<i>Ekho Moskv</i>
9	<i>Rossijskaia Gazeta</i>	<i>Vedomosti</i>	Znak.com
10	TASS	<i>Forbes</i>	Zona.media
11	Vesti.ru	360tv.ru	<i>Radio Svoboda</i>
12	<i>Vedomosti</i>	<i>Komsomolskaia Pravda</i>	Tsargrad.tv

13	BFM.ru	Gazeta.ru	Lenta.ru
14	Interfax	Moskovskij Komsomolets	Anews.com

Table colour code				
State media	Private “loyal” patriotic media	Private “loyal” liberal media	Independent registered media	Independent non-registered media

In June 2020, some Internet sources cited on Russian social networks such as Meduza.io, OpenMedia.io, MBKh Media or Mediazona never appear in the Yandex Top 5 rankings because they were not registered with Roskomnadzor [46]. Most of them are considered to be critical of state policies. The discrepancy between the narrow selection of media on Yandex.News and the greater pluralism on social networks shows the divergence between two different media spaces in Russia: registered media Web sites aggregated by Yandex.News on one side and news contents from alternative non-registered media circulating on social networks on the other. According to available data, these two media spaces attract different audiences. A media consumption study carried out by the Levada Sociological Center in Russia in February 2020 showed that people over 40 years old get their information mainly from official Web sites or from television while younger people (18–39 years old) secure it mostly from social networks [47].

The top-ranked independent media on social networks reflect the preferences of this audience. The ranking also reflects the dissemination strategies of the media themselves. Excluded from Yandex’s rankings since 2016, they carry out dissemination actions on social networks. The example of *Meduza* is very enlightening here. *Meduza* is a Riga-based online newspaper created by Galina Timchenko after she was fired from the news Web site Lenta.ru in 2014. According to its own metrics, in 2020 69 percent of its audience was younger than 45 [48]; moreover, traffic came mostly from direct connections and social networks, which is presented as a badge of honour with its traffic being “certified organic” (Figure 2). *Meduza* does not obtain any traffic from Yandex.News (compared with Lenta.ru, *Kommersant*, RBK which are more dependent on the aggregator). It has an active presence on social networks, as shown by the data presented in its media kit for advertisers.

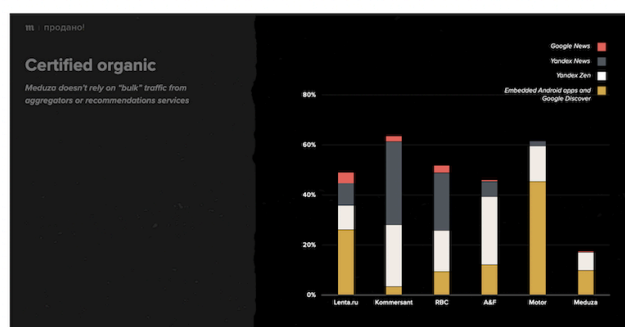


Figure 2: *Meduza* media kit, at <https://meduza.io/static/ads/mediakit-eng.pdf>, accessed 20 December 2020.

Alternative, dissenting or independent media have found it increasingly hard to operate in the officially regulated Russian media space, aggregated by Yandex.News. Some of them have been banned outright, while others have seen a sharp fall in traffic which has entailed dwindling revenues from advertising. Most of them have resorted to other business models (paywalls, subscriptions, fundraising etc.), leading to a fragmentation of the alternative media space and new market inequalities in access to independent news. These media have also migrated and relocated to other spaces and new types of distribution, disseminating their content on social media such as Facebook, Twitter, Instagram or Telegram.




Conclusion

The recent history of Yandex.News in Russia highlights how platform regulation can be leveraged to set up a form of “governance by algorithms” of the media and the public sphere. Initially presented as a technical means to “objectively” account for the diversity of online content, the aggregator sparked techno-political controversy in the 2010s: it was criticized by the authorities for promoting “unpatriotic” or “fake” news, while conversely journalists, Web professionals and end users increasingly suspected that *inconvenient* truths would find it difficult to reach its top rankings. The adoption in 2016 of a law on news aggregators, allowing only officially “registered” sources to be displayed by the service, clearly showed the intention to domesticate the platform in order to limit the visibility of protests and discontent in the public sphere. This regulation took place in a complex digital ecosystem which articulates different levels of gatekeeping, including Yandex.News and other platforms, the telecommunications watchdog Roskomnadzor, as well as media outlets and journalists.

The different types of evidence presented in this research — regulatory policies, public controversies, and a summary audit of the algorithm — indicate that Yandex as a news recommender system abides by both legal and technical “codes of conduct” ensuring that the information it promotes and amplifies remains in check. As a result, in 2020, the service gave visibility to only 14 media outlets which are themselves closely supervised by Russian authorities through Roskomnadzor. The tight control of the algorithm’s rankings is obvious compared to Google News which, although it gives pride of place to a small selection of major outlets, also accounts for a long tail of diverse publications and in any case, does not display a default selection of news on its search engine homepage. Yandex.News therefore only represents a facade of information pluralism. Moreover, it no longer reflects the diversity of content that still circulates in the Russian digital space. Although no outright censorship can yet be demonstrated at the level of Yandex.News, the aggregator appears to be an important cog in the machine of tightening control exerted by the authorities over the overall Russian media ecosystem.

However, governance by algorithms remains imperfect and takes place in a complex technical, political, legal and economic context where national and international platforms coexist and compete. Journalists and publishers seek alternative channels to distribute information, relying on social media such as Telegram or Twitter. Moreover, despite the new regulatory constraints, controversy over Yandex.News resurfaces periodically in times of political tension. The Russian authorities justify their efforts to control the media agenda and to reassert their sovereignty over the public sphere by denouncing information framed as

“unpatriotic”, “fake” or otherwise problematic. Paradoxically, this research also highlights how media players and news professionals, along with the new hurdles they face, are gradually developing critical views of the role and functioning of platforms and their algorithms — uncovering the political stakes of these key infrastructures. 

About the authors

Françoise Daucé is Professor at the School for Advanced Studies in the Social Sciences (EHESS) and director of the Center for Russian, Caucasian and East-European Studies (CERCEC).

E-mail: dauce [at] ehess [dot] fr

Benjamin Loveluck is Associate Professor at i3-SES, Telecom Paris, France.

E-mail: benjamin [dot] loveluck [at] telecom-paris [dot] fr

Notes

1. Statcounter, at <https://gs.statcounter.com/search-engine-market-share/all/russian-federation>, accessed 20 March 2021.

2. We thank Fabrice Demarthon (CNRS/CERCEC) for his help in extracting this database and scraping results from the Yandex homepage.

3. K. Bharat, “And now, News,” *Google Official Blog* (23 January 2006), at <https://googleblog.blogspot.com/2006/01/and-now-news.html>.

4. Z. Tufekci, “YouTube, the great radicalizer,” *New York Times* (10 March 2018), at <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>; see also O’Callaghan, *et al.*, 2015.

5. Trielli and Diakopoulos, 2019, p. 3.

6. This has led to heated discussions in most countries, which have usually found a settlement except in Spain, where the service is closed due to the introduction of a “link tax” requiring Google to pay a fee to display the text snippets.

7. Diakopoulos, 2019a, p. 183.

8. A. Robertson, “After its 4chan slip-up, is it time for Google to drop Top stories?” *The Verge* (3 October 2017), at <https://www.theverge.com/2017/10/3/16413082/google-4chan-las-vegas-shooting-top-stories-algorithm-mistake>.

9. Yandex 2020 financial results, at <https://ir.yandex/financial-releases?year=2020&report=q4>.

10. E. Osetinskaia, “Yandex, a Russian success story and Putin’s high-tech tiger,” Op-ed for the *Moscow Times* (27 September 2017), at <https://themoscowtimes.com/articles/yandex-a-russian-success-story-and-putins-high-tech-tiger-59029>.

11. M. Seddon, “Yandex agrees restructuring with Kremlin,” *Financial Times* (18 November 2019), at <https://www.ft.com/content/999e3ca6-09db-11ea-bb52-34c8d9dc6d84>.

12. On 16 March 2021, the Duma adopted in its third (final) reading a bill on penalties for selling smartphones, tablets and personal computers without pre-installed Russian software. Bill № 757430-7. Amendment of Article 14.8 of the Code of Administrative Offences of the Russian Federation, <https://sozd.duma.gov.ru/bill/757430-7>.

13. L. Gershenson, interview with the authors, January 2020.

14. These principles are available at <https://yandex.ru/promo/news/index>.

15. *Ibid.*

16. “Protesty ne v tope,” *Radio Svoboda — Krym Realii* (29 March 2017), at <https://ru.krymr.com/a/28397904.html>.

17. Federal Law № FZ-139, 28 July 2012.

18. “Yandex ‘razzhigaet’ Majdan v Rossii?” *Pravda*, at <http://www.pravda.ru/topic/yandex-617/>, accessed 25 August 2016.

19. Federal Law № FZ-208, 23 June 2016.

20. Interview with Tatyana Isaeva, *Meduza* (24 October 2016), at <https://meduza.io/feature/2016/10/24/oschuscheniya-chto-ot-mediasredy-otstali-net>.

21. *Ibid.*

22. “Deputaty popali v Yandex.Novosti,” *Kommersant* (16 August 2019), at https://lenta.ru/news/2019/08/16/ya_novosti/.

23. Interview with D., Moscow, September 2019.

24. Zh. Ulyanova and D. Luganskaya, “Rassledovanie RBK: kak chinovniki perekhitrili Yandex,” (22 October 2014), at http://top.rbc.ru/technology_and_media/22/10/2014/5447a659cbb20fd5d33b94d.

25. A. Kovalev, “Hear no evil, see no evil, report no evil,” *Moscow Times* (27 March 2017), at <https://www.themoscowtimes.com/2017/03/27/hear-no-evil-see-no-evil-report-no-evil-a57550>.

26. “Yandex vseimi svoimi servisami risuet obraz Naval’nogo,” *TJournal* (26 April 2020), at <https://tjournal.ru/news/162614-yandeks-vseimi-svoimi-servisami-risuet-obraz-navalnogo>.

27. Interview with A., activist of the Anti-Corruption Foundation, Moscow, September 2019.

28. Interview with D., Moscow, September 2019.

29. L. Gershenson, “Nothing to aggregate,” *Republic* (20 April 2016), at <https://republic.ru/posts/66965>.

30. Interview with Aleksandr Plyushchev, Moscow, March 2019.

31. <https://t.me/PlushevChannel/2533>.

32. E. Eremenko and E. Bryzgalova, “Durov sozdast agregator novostej,” *Vedomosti* (7 June 2019), at <https://www.vedomosti.ru/technology/articles/2019/06/07/803708-durov-sozdast>.

33. See the article by Ksenia Ermoshina and Francesca Musiani in this issue.

34. E. Eremenko and E. Bryzgalova, *art. cit.*

35. The first stage of the contest took place in November 2019 and the second stage in May 2020.

36. See <https://yandex.ru/news/smi>.

37. https://yandex.ru/company/press_releases/2004/0311.

38. Mail exchange with Yulia Berezovskaya, 27 May 2020.

39. “Protesty ne v tope,” *Radio Svoboda — Krym Realii* (29 March 2017), at <https://ru.krymr.com/a/28397904.html>.

40. We analysed the code of the Yandex homepage and found that 10 news references were presented at any given time. We therefore set up a Node.js script to collect these 10 references every two hours: four references occupy the places 1 to 4 of the Top 5, while the fifth place is likely occupied by the six other references on a rotating basis. The script uses two main Node libraries, Puppeteer for scrapping and Mongoose for database registration. After manually analyzing the html code of the homepage and several other pages of the Web site, we wrote the javascript code to scrape the content of the 10 top news (title, date, source name, source url, rank on the homepage). The data was then registered in a MongoDB database using the Mongoose library.

41. See for example “Media compass: Russia’s changing media landscape” (2 April 2014). Published by the *Calvert Journal* (supported by a non-profit U.K. registered charity created in 2009) at <https://www.calvertjournal.com/features/show/2234/russian-media-independent-compass>.

42. Medialogia is Russia’s first automated real-time media monitoring and analysis system, and was established in 2003. It was acquired by VTB Bank in 2019. Before the deal, the main owner was the IBS group, owned by

Anatoly Karachinsky and Sergey Matsotsky. According to its new owners, Medialogia will help improve the marketing strategies of its clients (“VTB Priobrel kontrol’nij paket v ‘Medialogii’,” *RBC* (17 January 2019), at <https://www.rbc.ru/rbcfreenews/5c4040da9a7947c79d040b64>).

43. The Levada Center is a Russian nongovernmental polling and sociological organization (named after its founder professor Yuri Levada). It is generally regarded as one of the few independent sociological research centres in Russia.

44. See Medialogia Web site: <https://www.mlg.ru/ratings/media/federal/5766/>.

45. Mail exchange with N., Medialogia employee, 15 December 2020.

46. <https://www.mlg.ru/ratings/media/socmedia/7492/>.

47. Levada Center, “Istochniki novostej i doverie SMI (Sources of information and trust in the media)” (27 February 2020), at https://www.levada.ru/cp/wp-content/uploads/2020/02/SMI_tablitsa.pdf. These results are coherent with the research on media consumption in Russia done by the audit company Deloitte in 2020, “Mediapotreblenie v Rossii — 2020 (Media consumption in Russia — 2020),” at <https://www2.deloitte.com/content/dam/Deloitte/ru/Documents/technology-media-telecommunications/russian/media-consumption-russia-2020.pdf>.

48. <https://meduza.io/static/ads/mediakit-eng.pdf>.

References

Valerie Belair-Gagnon and Avery E. Holton, 2018. “Boundary work, interloper media, and analytics in newsrooms: An analysis of the roles of Web analytics companies in news production,” *Digital Journalism*, volume 6, number 4, pp. 492–508.
doi: <https://doi.org/10.1080/21670811.2018.1445001>, accessed 22 April 2021.

Dominic Boyer, 2013. *The life informatic: Newsmaking in the digital era*. Ithaca, N.Y.: Cornell University Press.
doi: <https://doi.org/10.7591/cornell/9780801451881.001.0001>, accessed 22 April 2021.

Engin Bozdag, 2013. “Bias in algorithmic filtering and personalization,” *Ethics and Information Technology*, volume 15, number 3, pp. 209–227.
doi: <https://doi.org/10.1007/s10676-013-9321-6>, accessed 22 April 2021.

Samantha Bradshaw, 2019. “Disinformation optimised: Gaming search engine algorithms to amplify junk news,” *Internet Policy Review*, volume 8, number 4.
doi: <https://doi.org/10.14763/2019.4.1442>, accessed 22 April 2021.

David R. Brake, 2017. “The invisible hand of the unaccountable algorithm: How Google, Facebook and other tech companies are changing journalism,” In: Jingrong Tong and Shih-Hung Lo (editors). *Digital technology and journalism: An international comparative perspective*. Cham, Switzerland: Palgrave Macmillan, pp. 25–46.
doi: https://doi.org/10.1007/978-3-319-55026-8_2, accessed 22 April 2021.

Axel Bruns, 2019. *Are filter bubbles real?* Cambridge: Polity Press.

Taina Bucher, 2016. “Neither black nor box: Ways of knowing algorithms,” In: Sebastian Kubitschko and Anne Kaun (editors). *Innovative methods in media and communication research*. Cham, Switzerland: Palgrave Macmillan, pp. 81–98.
doi: https://doi.org/10.1007/978-3-319-40700-5_5, accessed 22 April 2021.

CamLy Bui, 2010. “How online gatekeepers guard our view — News portals’ inclusion and ranking of media and events,” *Global Media Journal*, volume 9, number 16, pp. 1–41, and at <https://www.globalmediajournal.com/open-access/how-online-gatekeepers-guard-our-view-news-portals-inclusion-and-ranking-of-media-and-events.pdf>, accessed 22 April 2021.

Joan Calzada and Ricard Gil, 2020. "What do news aggregators do? Evidence from Google News in Spain and Germany," *Marketing Science*, volume 39, number 1, pp. 134–167.
doi: <https://doi.org/10.1287/mksc.2019.1150>, accessed 22 April 2021.

Angèle Christin, 2020. *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton, N.J.: Princeton University Press.

Ivan Chupin and Françoise Daucé, 2017. "Termination of journalists' employment in Russia: Political conflicts and ordinary negotiation procedures in newsrooms," *Laboratorium*, volume 9, number 2, pp. 39–58.
doi: <https://doi.org/10.25285/2078-1938-2017-9-2-39-58>, accessed 22 April 2021.

Françoise Daucé, 2020. "Disguising the Internet? Website design and control in Russia," *Digital Icons*, number 20, at <https://www.digitalicons.org/issue20/disguising-the-internet-website-design-and-control-in-russia>, accessed 22 April 2021.

Françoise Daucé, 2017. "Political conflicts around the Internet in Russia: The case of Yandex.Novosti," *Laboratorium*, volume 9, number 2, pp. 112–132.
doi: <https://doi.org/10.25285/2078-1938-2017-9-2-112-132>, accessed 22 April 2021.

Anastasia Denisova, 2017. "Democracy, protest and public sphere in Russia after the 2011-2012 anti-government protests: Digital media at stake," *Media, Culture & Society*, volume 39, number 7, pp. 976–994.
doi: <https://doi.org/10.1177/0163443716682075>, accessed 22 April 2021.

Nicholas Diakopoulos, 2019a. *Automating the news: How algorithms are rewriting the media*. Cambridge, Mass.: Harvard University Press.

Nicholas Diakopoulos, 2019b. "Audit suggests Google favors a small number of major outlets," *Columbia Journalism Review* (10 May), at https://www.cjr.org/tow_center/google-news-algorithm.php, accessed 22 April 2021.

Robert Epstein and Ronald E. Robertson, 2015. "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections," *Proceedings of the National Academy of Sciences of the United States*, volume 112, number 33 (18 August), pp. E4512–E4521.
doi: <https://doi.org/10.1073/pnas.1419828112>, accessed 22 April 2021.

Julie Fedor, 2015. "Introduction: Russian media and the war in Ukraine," *Journal of Soviet and Post-Soviet Politics and Society*, volume 1, number 1, pp. 1–2, at <https://spps-jspps.autorenbetreuung.de/files/00-fedor.pdf>, accessed 22 April 2021.

Seth Flaxman, Sharad Goel and Justin M. Rao, 2016. "Filter bubbles, echo chambers, and online news consumption," *Public Opinion Quarterly*, volume 80, number S1, pp. 298–320.
doi: <https://doi.org/10.1093/poq/nfw006>, accessed 22 April 2021.

Richard Fletcher and Rasmus Kleis Nielsen, 2018. "Automated serendipity: The effect of using search engines on news repertoire balance and diversity," *Digital Journalism*, volume 6, number 8, pp. 976–989.
doi: <https://doi.org/10.1080/21670811.2018.1502045>, accessed 22 April 2021.

Tarleton Gillespie, 2018. "Regulation of and by platforms," In: Jean Burgess, Alice Marwick and Thomas Poell (editors). *Sage handbook of social media*. London: Sage, pp. 254–278.
doi: <http://dx.doi.org/10.4135/9781473984066.n15>, accessed 22 April 2021.

Mario Haim, Andreas Graefe and Hans-Bernd Brosius, 2018. "Burst of the filter bubble? Effects of personalization on the diversity of Google News," *Digital Journalism*, volume 6, number 3, pp. 330–343.
doi: <https://doi.org/10.1080/21670811.2017.1338145>, accessed 22 April 2021.

Natali Helberger, 2019. "On the democratic role of news recommenders," *Digital Journalism*, volume 7, number 8, pp. 993–1,012.
doi: <https://doi.org/10.1080/21670811.2019.1623700>, accessed 22 April 2021.

Natali Helberger, 2012. "Exposure diversity as a policy goal," *Journal of Media Law*, volume 4, number 1, pp. 65–92.
doi: <https://doi.org/10.5235/175776312802483880>, accessed 22 April 2021.

Natali Helberger, Kari Karppinen and Lucia D'Acunto, 2018. "Exposure diversity as a design principle for recommender systems," *Information, Communication & Society*, volume 21, number 2, pp. 191–207.
doi: <https://doi.org/10.1080/1369118X.2016.1271900>, accessed 22 April 2021.

Matthew Hindman, 2018. *The Internet trap: How the digital economy builds monopolies and undermines democracy*. Princeton, N.J.: Princeton University Press.

Matthew Hindman, 2008. *The myth of digital democracy*. Princeton, N.J.: Princeton University Press.

Sounman Hong and Nayeong Kim, 2018. "Will the Internet promote democracy? Search engines, concentration of online news readership, and e-democracy," *Journal of Information Technology & Politics*, volume 15, number 4, pp. 388–399.
doi: <https://doi.org/10.1080/19331681.2018.1534703>, accessed 22 April 2021.

Lucas D. Introna and Helen Nissenbaum, 2000. "Shaping the Web: Why the politics of search engines matters," *Information Society*, volume 16, number 3, pp. 169–185.
doi: <https://doi.org/10.1080/01972240050133634>, accessed 22 April 2021.

Dietmar Jannach, Markus Zanker, Alexander Felfernig and Gerhard Friedrich, 2011. *Recommender systems: An introduction*. Cambridge: Cambridge University Press.

Min Jiang, 2014. "The business and politics of search engines: A comparative study of Baidu and Google's search results of Internet events in China," *New Media & Society*, volume 16, number 2, pp. 212–233.
doi: <https://doi.org/10.1177/1461444813481196>, accessed 22 April 2021.

Natascha Just and Michael Latzer, 2017. "Governance by algorithms: Reality construction by algorithmic selection on the Internet," *Media, Culture & Society*, volume 39, number 2, pp. 238–258.
doi: <https://doi.org/10.1177/0163443716643157>, accessed 22 April 2021.

Mozhgan Karimi, Dietmar Jannach and Michael Jugovac, 2018. "News recommender systems — Survey and roads ahead," *Information Processing & Management*, volume 54, number 6, pp. 1,203–1,227.
doi: <https://doi.org/10.1016/j.ipm.2018.04.008>, accessed 22 April 2021.

Rob Kitchin, 2017. "Thinking critically about and researching algorithms," *Information, Communication & Society*, volume 20, number 1, pp. 14–29.
doi: <https://doi.org/10.1080/1369118X.2016.1154087>, accessed 22 April 2021.

Alexey Kovalev, 2020. "The political economics of news making in Russian media: Ownership, clickbait and censorship," *Journalism* (14 August).
doi: <https://doi.org/10.1177/1464884920941964>, accessed 22 April 2021.

Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi and Karrie Karahalios, 2019. "Search bias quantification: Investigating political bias in social media and Web search," *Information Retrieval Journal*, volume 22, numbers 1–2, pp. 188–227.
doi: <https://doi.org/10.1007/s10791-018-9341-2>, accessed 22 April 2021.

Bruno Latour, 2005. *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.

- Lawrence Lessig, 1999. *Code and other laws of cyberspace*. New York: Basic Books.
- Markku Lonkila, Larisa Shpakovskaya and Philip Torchinsky, 2020. “The occupation of Runet? The tightening state regulation of the Russian-language section of the Internet,” In: Mariëlle Wijermars and Katja Lehtisaari (editors). *Freedom of expression in Russia’s new mediasphere*. Abingdon: Routledge, pp. 17–38.
- Nathalie Maréchal, 2017. “Networked authoritarianism and the geopolitics of information: Understanding Russian Internet policy,” *Media and Communication*, volume 5, number 1, pp. 29–41.
doi: <http://dx.doi.org/10.17645/mac.v5i1.808>, accessed 22 April 2021.
- Noortje Marres, 2007. “The issues deserve more credit: Pragmatist contributions to the study of public involvement in controversy,” *Social Studies of Science*, volume 37, number 5, pp. 759–780.
doi: <https://doi.org/10.1177/0306312706077367>, accessed 22 April 2021.
- Ulises A. Mejias and Nikolai E. Vokuev, 2017. “Disinformation and the media: The case of Russia and Ukraine,” *Media, Culture & Society*, volume 39, number 7, pp. 1,027–1,042.
doi: <https://doi.org/10.1177/0163443716686672>, accessed 22 April 2021.
- Francesca Musiani, 2013. “Governance by algorithms,” *Internet Policy Review*, volume 2, number 3.
doi: <https://doi.org/10.14763/2013.3.188>, accessed 22 April 2021.
- Francesca Musiani, Benjamin Loveluck, Françoise Daucé and Ksenia Ermoshina, 2019. “‘Digital sovereignty’: Can Russia cut off its Internet from the rest of the world?” *The Conversation* (28 October), at <https://theconversation.com/digital-sovereignty-can-russia-cut-off-its-internet-from-the-rest-of-the-world-125952>, accessed 13 May 2020.
- Philip M. Napoli, 2014. “Automated media: An institutional theory perspective on algorithmic media production and consumption,” *Communication Theory*, volume 24, number 3, pp. 340–360.
doi: <https://doi.org/10.1111/comt.12039>, accessed 22 April 2021.
- Efrat Nechushtai and Seth C. Lewis, 2019. “What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations,” *Computers in Human Behavior*, volume 90, pp. 298–307.
doi: <https://doi.org/10.1016/j.chb.2018.07.043>, accessed 22 April 2021.
- Rasmus Kleis Nielsen and Sarah Anne Ganter, 2018. “Dealing with digital intermediaries: A case study of the relations between publishers and platforms,” *New Media & Society*, volume 20, number 4, pp. 1,600–1,617.
doi: <https://doi.org/10.1177/1461444817701318>, accessed 22 April 2021.
- Julien Nocetti, 2015. “Russia’s ‘dictatorship-of-the-law’ approach to Internet policy,” *Internet Policy Review*, volume 4, number 4.
doi: <https://doi.org/10.14763/2015.4.380>, accessed 22 April 2021.
- Sarah Oates, 2013. *Revolution stalled: The political limits of the Internet in the post-Soviet sphere*. Oxford: Oxford University Press.
doi: <https://doi.org/10.1093/acprof:oso/9780199735952.001.0001>, accessed 22 April 2021.
- Derek O’Callaghan, Derek Greene, Maura Conway, Joe Carthy and Pádraig Cunningham, 2015. “Down the (white) rabbit hole: The extreme right and online recommender systems,” *Social Science Computer Review*, volume 33, number 4, pp. 459–478.
doi: <https://doi.org/10.1177/0894439314555329>, accessed 22 April 2021.
- Eli Pariser, 2011. *The filter bubble: What the Internet is hiding from you*. New York: Penguin Press.
- Frank Pasquale, 2015. *The black box society: The secret algorithms that control money and information*. Cambridge, Mass.: Harvard University Press.

Francesco Ricci, Lior Rokach and Bracha Shapira (editors), 2015. *Recommender systems handbook*. New York: Springer.

doi: <https://doi.org/10.1007/978-1-4899-7637-6>, accessed 22 April 2021.

Ronald E. Robertson, David Lazer and Christo Wilson, 2018. “Auditing the personalization and composition of politically-related search engine results pages,” *WWW '18: Proceedings of the 2018 World Wide Web Conference*, pp. 955–965.

doi: <https://doi.org/10.1145/3178876.3186143>, accessed 22 April 2021.

Florian Saurwein, Natascha Just and Michael Latzer, 2015. “Governance of algorithms: Options and limitations,” *info*, volume 17, number 6, pp. 35–49.

doi: <https://doi.org/10.1108/info-05-2015-0025>, accessed 22 April 2021.

Roland Schroeder and Moritz Kralemann, 2005. “Journalism ex machina — Google News Germany and its news selection processes,” *Journalism Studies*, volume 6, number 2, pp. 245c247.

doi: <https://doi.org/10.1080/14616700500057486>, accessed 22 April 2021.

Nick Seaver, 2019. “Knowing algorithms,” In: Janet Vertesi and David Ribes (editors). *digitalSTS: A field guide for science & technology studies*. Princeton, N.J.: Princeton University Press, pp. 412–422.

Liudmila Sivetc, 2019. “State regulation of online speech in Russia: The role of Internet infrastructure owners,” *International Journal of Law and Information Technology*, volume 27, number 1, pp. 28–49.

doi: <https://doi.org/10.1093/ijlit/eay016>, accessed 22 April 2021.

Andrei Soldatov and Irina Borogan, 2015. *The red Web: The Kremlin's wars on the Internet*. New York: PublicAffairs.

Kjerstin Thorson and Chris Wells, 2016. “Curated flows: A framework for mapping media exposure in the digital age,” *Communication Theory*, volume 26, number 3, pp. 309–328.

doi: <https://doi.org/10.1111/comt.12087>, accessed 22 April 2021.

Kjerstin Thorson and Chris Wells, 2015. “How gatekeeping still matters: Understanding media effects in an era of curated flows,” In: Timothy Vos and François Heinderyckx (editors). *Gatekeeping in transition*. New York: Routledge, pp. 25–44.

doi: <https://doi.org/10.4324/9781315849652>, accessed 22 April 2021.

Carolina Vendil Pallin, 2017. “Internet control through ownership: The case of Russia,” *Post-Soviet Affairs*, volume 33, number 1, pp. 16–33.

doi: <https://doi.org/10.1080/1060586X.2015.1121712>, accessed 22 April 2021.

Mariëlle Wijermars, 2021. “Russia’s law ‘On news aggregators’: Control the news feed, control the news,” *Journalism* (15 January).

doi: <https://doi.org/10.1177/1464884921990917>, accessed 22 April 2021.

Mariëlle Wijermars and Katja Lehtisaari (editors), 2020. *Freedom of expression in Russia's new mediasphere*. Abingdon: Routledge.