Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)

The Limits of De-politicizing—and Also of Annotation: A Case Study in Russian Media Outlets' Social Media Posts, 2016–2024

Liancheng Gong[1], **Daniel J. Hopkins**[23], **Samuel Wolken**[2]
[1]University of Maryland, College Park, College of Information
[2]University of Pennsylvania, Department of Political Science
[3]University of Pennsylvania, School of Engineering and Applied Science
gonglc@umd.edu, danhop@sas.upenn.edu, sam.wolken@asc.upenn.edu

Abstract

We seek to understand the impact of the February 2022 invasion of Ukraine on political coverage and engagement with it. Accordingly, we introduce new annotated data with more than 2 million social media posts from prominent Russian-language print and online media outlets. After manually annotating 6,661 posts that appeared on V'Kontakte (VK) as well as additional posts on Facebook and Telegram—and supplementing human annotation with Large Language Models (LLMs)—we assess the validity and reliability of our measures before considering substantive questions. Varied analyses point to one conclusion: the ecosystem for political news durably changed with the invasion. Post-invasion, political posts spiked, and social media users became more likely to engage with political posts versus non-political posts. Moreover, analyses using embeddings illustrate that the similarity between independent and government-oriented media posts briefly grew after the invasion. By analyzing both outlets' coverage and engagement with it, this research indicates that autocrats' strategies for managing news coverage are not static. Instead, autocratic regimes may abandon de-politicization in favor of more invasive approaches when events drive heightened engagement with news. This research also provides methodological tools for future research. It evaluates the value and limitations of deploying LLMs to annotate and analyze text in Russian, and also of translating text from Russian to English before annotation or analysis. This research contributes to our understanding of the possibilities and limits of using LLMs to measure more abstract concepts, too.

Introduction

In autocracies and democracies alike, politicians seek to build support by influencing the information that is salient to citizens (Gehlbach, Sonin, and Svolik 2016; Rosenfeld and Wallace 2024). Autocratic regimes can do so directly through tactics such as censorship or by owning or pressuring media outlets (King, Pan, and Roberts 2013, 2014). However, they can also intervene indirectly. One such tactic is to demobilize and de-politicize society (Gerschewski 2023), either by diverting attention to apolitical news (Fredheim 2017) or by framing political news so as to generate apathy and cynicism (Applebaum 2024). Another tactic is to deflect blame (Rozenas and Stukal 2019).

Still, such strategies are not employed in a vacuum. They can be undermined by the widespread availability of online information, and also by events that draw citizens' attention to politics. Events may also change the demand for political news coverage, potentially upending autocrats' existing strategies for managing the news media via de-politicization.

Here, we study the interplay of audience engagement, news coverage, and events by analyzing prominent Russian-language media outlets' social media posts in the period between 2016 and 2024. We focus on 1,623,310 social media blurbs posted by 20 news organizations on V'Kontakte (VK), a major Russian social media site (see also Park et al. 2022). We also provide supplemental analyses of blurbs posted by the same outlets on Facebook and Telegram.

Even after extensive efforts to refine, streamline, and improve our codebook, we find that some of our categories—such as whether a post has political content or mentions Russian President Vladimir Putin—can be much more reliably measured than other categories, both by human annotators and by Large Language Models (LLMs). After characterizing the over-time and cross-outlet distribution of political attention, we investigate related questions including the performance of classifiers on Russian versus English-language text. Our analyses show a spike in political coverage with the February 2022 invasion alongside a related spike in audience interest in political posts. Additionally, independent outlets' production of political content declined in the wake of new legal restrictions in 2022, at which point these independent outlets' coverage also became increasingly similar to other outlets' coverage. These varied empirical observations jointly indicate the limits of de-politicizing strategies for managing media content, as some events generate spikes in interest and engagement with political topics.

This manuscript also offers methodological contributions. The emergence of LLMs has the potential to transform multiple parts of the annotation process. For one thing, LLMs provide researchers working with foreign-language text the option to translate that text into a different language, a move which may potentially increase a research team's capacity to annotate text but may also degrade its accuracy. LLMs can be deployed directly to annotate multi-lingual text, too. We find that for some tasks, classifiers perform slightly better on

translated, English-language text.

The next section provides background on the Russian media environment. This paper then summarizes prior research before detailing its Data and Methods and then its Results.

Background

Here, we briefly provide background about the evolution of the media environment in contemporary Russia to motivate our subsequent analyses.[1]

Given the centrality of television as an information source in Russia, in the early period after Vladimir Putin's election as President in 2000, government efforts focused on reining in independent television stations such as NTV and ORT, with newspapers and radio given a wider berth (Frye 2022; see also Enikolopov, Petrova, and Zhuravskaya 2011). But over time, the government and its allies intervened more actively in print and online journalism as well (Vendil Pallin 2017). Such interventions accelerated after 2011-2013, a period which saw the announcement that Putin would seek to return to the presidency in 2012 as well as extensive protests on that and other issues (Koesel and Bunce 2012). As Appendix Table 15 details, during the same period, the Russian government also asserted greater control over Russian print and online media through ownership by the state or Kremlin allies. The killing of 48 journalists in Russia between 1999 and 2023 is very likely to have influenced coverage on politically sensitive topics as well (Bjorklund and Smith 2023).

Government encroachment on media freedom intensified following February 2022, when Russian forces dramatically expanded their operations in Ukraine, launching a full-on invasion targeting Kyiv. At the outset of the invasion in March 2022, Russia implemented new, strict laws governing media outlets, Russian Federal Laws No. 31-FZ and No. 32-FZ. These laws criminalized "discrediting the armed forces" (Troianovski and Safronova 2022; Thompson Reuters Foundation and the Committee to Protect Journalists 2022), and they were subsequently expanded to include Russian state actors and pro-Russian groups. Their enactment and enforcement had the effect of driving some independent media outlets—including TV Rain and *Novaya Gazeta* (both outlets we study)—out of Russia (Troianovski and Safronova 2022). The Russian government subsequently blocked those outlets' websites and also blocked the social media website Facebook. In short, the 2022 Ukraine War ushered in a new era of overt media censorship in Russia.

Our analyses below focus on VK, a popular Russian social media website whose usage within Russia has increased in the wake of 2022 moves by the Russian government to ban Facebook and Instagram. Even in 2019, VK reported more users in Russia than platforms including Instagram and Facebook (DataReportal 2019; Enikolopov, Makarin, and Petrova 2020). However, in 2021, the Russian government-connected firm Gazprom took control of VK, giving "the Russian state significant involvement in a key part of Russia's online space" (Moscow Times 2021). According to Mediascope (2025), by early 2025, VK had 94 million users in Russia, making it the country's largest social media platform and also outpacing Telegram by a wide margin (DataReportal 2025).

[1]For theoretical models of autocratic politics, see Svolik 2012 and Gehlbach, Sonin, and Svolik 2016. On contemporary Russian autocracy, see Colton 2016, Baturo and Elkink 2021, Rosenfeld 2021, and Frye 2022.

Prior Research

In the period since the Second World War, researchers have developed nuanced conceptions of the various pathways through which media might influence the relationship between citizens and their government (Bennett and Iyengar 2008). Media need not directly change citizens' views to be influential—in some cases, it can serve as a coordinating point for collective action, while in others it can act to demobilize or prevent collective action (King, Pan, and Roberts 2013; Roberts 2018; Frye 2022; Porter et al. 2024). In that vein, King, Pan, and Roberts (2013) find that Chinese censorship activity targets not expressions of negative opinion but instead posts that may generate collective action (but see Miller 2025). Like democracies, autocracies have developed various approaches to managing the media environment both at home and abroad, although autocracies face fewer limits with respect to media censorship, and commonly employ state repression to shape coverage (Gehlbach, Sonin, and Svolik 2016).

There are various tactics that autocratic governments can use to shape the information available to their citizens, some more oppressive than others (Guriev and Treisman 2018). Such tactics are likely to vary over time, across media types/outlets, and across places even within a single country (Litvinenko and Nigmatullina 2020). Some focus on the ownership of media outlets, either directly or indirectly (Gehlbach and Sonin 2014; Gehlbach, Sonin, and Svolik 2016; Simonov and Rao 2022). Beyond government directives about what news to cover (e.g. Waight et al. 2025), autocratic governments can employ other, more subtle tactics to influence the information environment. Field et al. (2018) uses evidence from *Izvestia*, a state-aligned newspaper we include below, to demonstrate that when economic performance declines, the paper calls more attention to the U.S. In a related finding, Rozenas and Stukal (2019) shows that Russia's largest state-owned television station does not censor bad economic news, but instead frames it as due to external influences rather than government failings (see also Lankina and Watanabe 2017; Park et al. 2022; Applebaum 2024). Autocratic governments can also deploy more decentralized forms of propaganda via social media (Lu et al. 2025).

Another tactic is to de-emphasize politics (Gerschewski 2023). For example, Fredheim (2017) studies the online Russian newspapers *Lenta* and *Gazeta*—two of the outlets we study here—and finds that pro-government editorial changes at both papers led to sizable declines in the fraction of news devoted to domestic or international politics. Instead, the papers shifted to human interest stories. Editors who are more sympathetic to the Kremlin can thus shift content to avoid more sensitive political topics (see also Goode 2025). Classical treatments of totalitarianism such as Arendt (1948) emphasize that such governments seek to mobilize the public and collapse the distinction between the

personal and the political. However, autocratic governments may seek to do the reverse by de-politicizing news coverage and encouraging citizens to withdraw from the political arena (Hjermann 2023).

The market for news—and governments' incentive to react—depends on events, too (Simonov and Rao 2022). Some events may be especially politically sensitive, while others may not. Events may also change the demand for political news, and so force autocratic governments to adopt new strategies for managing information. In contrast to a strategy of de-politicization, Nieman and Labzina (2025) illustrate that in the run-up to the invasions of Georgia and Ukraine, state-owned media outlets in Russia increased their coverage of those countries. Events may alter the demand for news, a sometimes-neglected but critical factor, even in autocracies (Shirikov 2024). They may even force autocratic regimes into new strategies for managing news. We turn now to assessing these possibilities empirically.

Data and Methods

Data Compilation

Our goal is to study the content of and online engagement with social media posts by prominent Russian media outlets, primarily newspapers. Given the government's particular efforts to control information on television (Pomerantsev 2014; Frye 2022), newspapers are likely to display greater variation in their political coverage.

To identify our sample of media outlets with varying relationships to the Russian government, we first began with a list of the highest-circulation newspapers in Russia. We then extended the list with prominent outlets studied by prior research (Frye 2022; Gehlbach, Lokot, and Shirikov 2023; Shirikov 2024). Because outlets that were more independent from or critical of the Russian government were less common, especially after March 2022, we then augmented the list by including such outlets even when their circulation was lower or when they did not distribute print newspapers or magazines (e.g. TV Rain). Next, we evaluated each outlet's relationship with the government qualitatively over time, with Appendix Table 15 summarizing the results and our classifications of each. Note that during this period, the trend was towards increased government influence over media outlets.

Next, we confirmed that we could obtain posts on VK from each source between January 1st, 2016 and October 1st, 2024, resulting in a list of 20 prominent media outlets which had VK posts. Those media outlets are: aif.ru, gazeta.ru, iz.ru, kommersant.ru, lenta.ru, life.ru, mediazona.ru, meduza.ru, mk.ru, newizv.ru, newtimes.ru, ng.ru, novayagazeta.ru, pravda.ru, rbc.ru, republic.ru, rg.ru, tvrain.ru, vedomosti.ru, and vesti.ru. These outlets were chosen because they vary in their relationship to the Russian government; because they at one time had large audiences within Russia; and because they posted on social media sites such as VK, Facebook, and Telegram in the period in question. Appendix Figures 9, 10, and 11 provide an example post from aif.ru (*Arguments and Facts*) on each platform. Appendix Figure 12 illustrates the monthly posts by outlet for the VK data, while Appendix Figures 13 and 14 do so for the Facebook and Telegram posts, respectively. Appendix Figure 15 reports the distribution of word counts for the VK posts in our sample.

This project originally focused on Facebook posts, so we began with January 1st, 2016, the first date for which URL shares were available through Facebook's Social Science One portal. However, subsequent events—including the 2022 invasion of Ukraine, the declining use of Facebook after Russia's designation of the company as an extremist/terrorist organization (BBC 2022), and the incomplete data availability evident in Figure 13—led us to shift our emphasis to VK. Extending these analyses back to include the 2014 invasion of Crimea is a valuable direction for future research.

In total, we collect 1,623,310 VK posts for subsequent analyses using the vk.com API. We study the text (but not the images) in these posts, which we refer to as "posts" or "blurbs" interchangeably.[2] To probe whether our results were specific to VK, we also repeat this process with a set of news blurbs posted by many of these same media outlets on Facebook, which we obtain via the Social Science One collaboration. However, the Facebook blurbs are truncated in time, both because Social Science One did not make blurbs after October 2022 available and because of the 2022 ban on Facebook within Russia. In addition, we download 799,949 posts from these outlets on Telegram, another widely used social media site in Russia. Table 1 lists the outlets along with the dates for which posts were available and the total number of posts.

Annotation

To quantify the volume of political news coverage and better understand the posts' content over time, we next develop an annotation scheme which quantifies key elements of their texts. Media outlets cover a wide range of issues, events, and individuals, and we are interested in the extent to which coverage focuses on political issues as well as international issues such as Ukraine. Our codebook thus asks annotators to indicate whether a blurb is political and whether it mentions Russian President Putin explicitly. It also measures more subjective elements of the posts, such as whether it is "unambiguously positive" or "unambiguously negative" towards Putin as well as whether the story is "unambiguously good news" or "unambiguously bad news" for the Russian government.

Appendix B includes the detailed codebook, which was the product of extensive iteration and revision. Following best practice (see esp. Hopkins and King 2010; Grimmer, Roberts, and Stewart 2022; Halterman and Keith 2025), select authors and several undergraduate annotators applied our preliminary codebook to Facebook blurbs from news outlets. We then revised the codebook after estimating inter-

[2]In the Social Science One data set, the unit of analysis is an external URL (in this context, typically a news article), rather than a Facebook post. Facebook makes available a generic "blurb" summary for each URL (rather than the text of individual Facebook posts containing the URL).

Media Outlet	Earliest Date	Latest Date	# Posts
aif.ru	2016-01-01	2024-09-10	61904
gazeta.ru	2016-01-19	2024-09-12	85352
iz.ru	2016-03-08	2024-09-10	94988
kommersant.ru	2017-05-03	2024-09-10	10419
lenta.ru	2016-01-12	2024-09-10	40067
life.ru	2016-01-01	2024-09-12	65874
mediazona.ru	2016-01-07	2024-03-28	22524
meduza.ru	2016-01-21	2024-05-14	86282
mk.ru	2016-01-02	2024-09-12	384335
newizv.ru	2016-01-04	2024-09-11	12068
newtimes.ru	2016-02-04	2024-09-11	44341
ng.ru	2016-05-18	2024-09-11	78738
novayagazeta.ru	2016-01-19	2023-07-28	24306
pravda.ru	2016-01-01	2024-09-11	108215
rbc.ru	2016-01-11	2024-09-30	172559
republic.ru	2016-01-02	2024-09-12	49109
rg.ru	2016-01-01	2017-05-03	6522
tvrain.ru	2016-01-01	2022-03-03	89192
vedomosti.ru	2016-01-11	2024-09-11	104692
vesti.ru	2016-01-08	2024-09-12	81823

Table 1: Summary of VK Posts by Media Outlet

coder reliability, studying cases of coder disagreement and discussing those cases with Russian- and English-speaking annotators. For example, our political category was initially defined as "1 if there is any relevance to politics/government," whereas the finalized codebook is substantially more detailed. It provides clarifications (e.g., "This category should only include explicit references to politics, government institutions, government policy, military actions, or diplomacy"; see Appendix B) as well as guidelines to handle specific cases (e.g., "References to political leaders are not sufficient unless discussing government or politics").

To be sure, the very definition of "political" has been contested for centuries (e.g. Marx and Engles 1845), and we do not pretend to resolve such disputes. Our goal was to produce sufficiently clear instructions such that we developed a measure which could obtain reasonable inter-annotator agreement while providing a plausible operationalization of the underlying construct. We thus opt for a narrow definition of "political" while fully acknowledging the value of analyses with more expansive conceptions (Goode 2025). For instance, a human-interest story about Crimea may not contain explicit political content and so would not be classified as political even if it serves a political purpose by advancing the Russian state's narrative or shifting public attention away from other events (Mattingly and Yao 2022; Shirikov 2024; Goode 2025). The finalized codebook also includes multiple conditional annotations. For example, we only annotate posts as to whether the news is good or bad for the Russian government if they explicitly mention a Russian government actor.

To annotate a subset of the VK posts, we draw stratified random samples of 1,950 posts, with 50 posts before 2022 and 50 posts after 2022 from each of 20 outlets. By stratifying, we can be sure we have sufficient sample sizes from before and after the 2022 invasion, making our analyses more robust to changes in language during this period and enabling us to test key assumptions (Egami et al. 2024).[3] After translating the posts using GPT-4 from OpenAI, we then employed two undergraduates who coded the posts in Russian and another three undergraduates who annotated the same 1,950 VK posts in English translation. For VK posts, we have 6,661 human annotations from five separate human annotators, 3,918 in English and 2,743 in Russian.[4]

Our primary interest is in tracking changes over time, so a key assumption is that our annotators are not differentially accurate at specific points in time. Appendix Figure 17 investigates this possibility, finding that levels of annotator disagreement for the "political" category do not systematically rise or decline over time.

Multi-lingual Annotation

One of the ways in which LLMs may assist researchers is with multi-lingual annotation. The simplest route to label our text corpus would be to provide text samples directly to an LLM in the original language for zero-shot classification. However, LLMs tend to perform worse on tasks with text samples in languages other than English (Nikolich et al. 2024). This performance gap reflects the composition of training data for LLMs (which is skewed towards English), leading to shortcomings in cultural knowledge as well as inefficiencies in the tokenization of non-English words (Nikolich et al. 2024; Tikhomirov and Chernyshev 2023). Thus, this approach may result in suboptimal label quality, compromising downstream analyses. Nonetheless, for a U.S.-based research team like ours, it is easier to find English-speaking research assistants than Russian-speaking research assistants. Accordingly, one question we analyze is whether research assistants who annotate the blurbs in their original Russian outperform those who annotate blurbs that have been translated into English. Specifically, we employ GPT-4 to translate each post into English and assign multiple Russian and English speakers to annotate each post.

Given that GPT-4 is a generative language model, it will not necessarily follow the conditional rules laid out in our codebook, and may instead provide forbidden combinations of annotations, a form of hallucination. In fact, we find rates of such hallucinations as high as 9% (for the combination not mentioning the Russian government but identifying good news for it), although others (such as identifying mentions of democratic activities in non-political posts) are much less common (0.07%). To address these issues, we impose consistency by coercing the dependent categories to prohibit impossible combinations. For example, all posts which have no mention of the Russian government do not

[3] rg.ru provides only 50 posts each due to lack of consistent availability over time.

[4] With the 1,874 Facebook posts, we similarly used GPT-4 to translate the posts into English, and employed three undergraduates to code the posts in Russian and another two who annotated them in English translation. In total, we have 4,842 human annotations of these Facebook posts from 5 separate annotators, 2,084 in English and 2,758 in Russian.

Label	All annotations	All GPT	All human
Panel A: Unconditional α (n=20347; 13686;6661)			
POLITICAL	0.688	0.826	0.519
PUTIN	0.917	0.940	0.887
PUTIN NEG	0.525	0.621	0.365
RUSSIAN GOVT	0.590	0.718	0.338
GOOD NEWS	0.379	0.537	0.208
BAD NEWS	0.330	0.446	0.148
NEGATIVE GOVT	0.371	0.450	0.217
Panel B: Conditional α			
Given `PUTIN`*=1* (n=1559; 1083; 476)			
PUTIN NEG	0.515	0.596	0.382
Given `RUSSIAN GOVT`*=1* (n=6124, 4251, 1873)			
GOOD NEWS	0.437	0.583	0.334
BAD NEWS	0.403	0.559	0.309
NEGATIVE GOVT	0.483	0.596	0.309

Table 2: Unconditional and conditional Krippendorff's Alphas for VK annotations. Sample sizes: total number of annotations, number of GPT annotations, number of human annotations. We use GPT-4o and GPT-o3-mini to annotate both English and Russian posts, resulting in four GPT-based models under "All GPT."

Label	English	Russian	Combined
Panel A: Unconditional α (n=3918; 2743; 6661)			
POLITICAL	0.480	0.594	0.519
PUTIN	0.890	0.886	0.887
PUTIN NEG	0.458	0.665	0.365
RUSSIAN GOVT	0.160	0.424	0.338
GOOD NEWS	0.114	0.252	0.208
BAD NEWS	0.117	0.169	0.148
NEGATIVE GOVT	0.122	0.295	0.217
Panel B: Conditional α			
Given `PUTIN`*=1* (n=273, 203, 476)			
PUTIN NEG	0.439	0.735	0.382
Given `RUSSIAN GOVT`*=1* (n=1144; 729; 1873)			
GOOD NEWS	0.233	0.256	0.334
BAD NEWS	0.233	0.248	0.309
NEGATIVE GOVT	0.226	0.538	0.309

Table 3: Krippendorff's Alpha: Human Annotations by Language. Sample sizes: n=number of human annotations in English; in Russian; all human annotations.

contain good news, bad news, or negative coverage of the Russian government by imposition.

Table 2 reports the inter-coder reliability—measured via Krippendorff's Alpha (Krippendorff 1970)—for the VK sample which we focus on here. Krippendorff (2004) indicates that an Alpha above 0.8 is high, and that an Alpha above 0.667 permits "tentative conclusions." With that in mind, we present the results for key categories in Table 2. The human annotators only agree at high levels as to whether posts explicitly mention Putin, with the coders reading LLM-translated texts performing basically comparably to those reading in the original Russian (Table 3). But in other, more subjective categories, agreement among annotators is lower—and there, the annotators who read posts in Russian always have higher inter-coder agreement than those reading the English translations. Inter-coder reliability typically improves when conditioning on posts for which a category such as "good news" is relevant, but as the bottom panels of Tables 2 and 3 make clear, this is not always the case. As a result, we focus below on the categories with higher inter-coder reliability and also perform additional corrections or aggregation to address this issue.

Automated Annotation and Validation

Because human annotation is costly and time-consuming, researchers have assessed the extent to which LLMs can replace human annotators on text-labeling tasks. Influential research has furnished evidence that LLM performance exceeds other classification approaches and even trained humans on social science classification tasks (Törnberg 2024; Rathje et al. 2024; Gilardi, Alizadeh, and Kubli 2023). However, LLM performance varies across tasks and data sets (Pangakis, Wolken, and Fasching 2023; Egami et al. 2024; Wang et al. 2024; Halterman and Keith 2025). Classification errors by LLMs may reflect an incomplete understanding of the underlying concept being measured (Wallach et al. 2024; Halterman and Keith 2025), which can introduce systematic biases in labels assigned by an LLM. Recognizing these potential pitfalls, we validate LLM performance on a task-by-task basis (Pangakis and Wolken 2025). By doing so, we do not defer the comprehension of our measured topics to the LLM but instead, ensure that the LLM is serving as a measurement tool consistently aligned to our human annotators' judgment.

In Table 4, we use GPT-o3-mini to annotate the texts, defining our benchmark as the majority of labels from all human annotations in Russian and English for the VK and Facebook data sets. (We employ GPT-o3-mini due to its high levels of accuracy and low latency given its cost for large-scale labeling.) The results from the LLM roughly track the results from the human annotators above: the categories for which we saw higher inter-coder reliability are also those where GPT matches human annotations more closely. For example, the GPT annotations for the "Putin" category closely track the human annotations as measured via the F1 statistic, something that is true to a slightly lesser extent for the "political" and the "Russian government" categories. But from there, performance falls off steeply, with GPT not doing very well identifying whether human annotations are negative references to Putin (for Facebook posts), whether posts are good news or bad news for the Russian government, or whether they discuss the Russian government in a negative light. Overall, the F1 scores vary from a low of 0.22 to a high of 0.92, and so closely mirror the findings of the meta-analysis in Egami et al. (2024).

Multi-lingual Prediction

A related question is whether classifiers perform better when the predictive features are generated from the original Russian text or from the translated English text. It is possible that Russian's rich morphology may lead to reduced perfor-

Lbl	Src	T%	P%	[TN,FP;FN,TP]	F1
Pol	VK	47.9%	53.4%	[[619, 162], [80, 639]]	0.84
	FB	58.2%	46.5%	[[624, 143], [358, 712]]	0.74
PT	VK	7.2%	7.5%	[[1381, 11], [6, 102]]	0.92
	FB	6.9%	4.7%	[[1706, 5], [45, 81]]	0.76
PT-	VK	0.8%	0.7%	[[1486, 2], [4, 8]]	0.73
	FB	2.0%	0.8%	[[1797, 3], [26, 11]]	0.43
RG	VK	24.5%	29.0%	[[972, 160], [93, 275]]	0.68
	FB	46.7%	23.4%	[[899, 81], [509, 348]]	0.54
GN	VK	5.1%	5.3%	[[1374, 49], [46, 31]]	0.39
	FB	5.9%	1.6%	[[1714, 15], [93, 15]]	0.22
BN	VK	2.7%	4.7%	[[1402, 58], [27, 13]]	0.23
	FB	18.1%	6.5%	[[1447, 57], [271, 62]]	0.27
G-	VK	2.7%	3.6%	[[1419, 40], [27, 14]]	0.29
	FB	20.5%	6.3%	[[1413, 48], [308, 68]]	0.28

Table 4: Confusion matrices between GPT-o3-mini annotations and gold-standard labels for the VK (n=1,500) and Facebook (or "FB") (n=1,837) data sets. Both datasets are comprised of the intersection between human and GPT-o3-mini annotations. T% refers to the percentage of samples with a given label among the human annotations; P% refers to the percentage of samples in the GPT-o3-mini labels. The matrix format is [TN, FP; FN, TP], where TN = True Negatives, FP = False Positives, FN = False Negatives, and TP = True Positives. The gold standard comes from the human annotations, with majority rule used to decide in cases of annotator disagreement. Labels are renamed for brevity: Pol = Political; PT = Putin; PT- = Putin Negative; RG = Russian Government; GN = Good News; BN = Bad News; G- = Negative Government

mance when standard NLP approaches are applied directly to Russian texts. To answer this question, we employ a variety of approaches to classification separately with the Russian and English text and then compare the results in Tables 5 and 6.

Pre-processing decisions can be influential on the subsequent estimation (Denny and Spirling 2018), and that may be especially true when classifying across multiple languages. To prepare the data for the classifiers—Logistic Regression, Support Vector Classification (SVC), and Random Forests—we use bag-of-words unigram features with language-specific normalization. For English, we lowercase, tokenize with a regex, remove NLTK stopwords, and apply Snowball stemming; for Russian, we lowercase, tokenize with a Cyrillic regex, remove NLTK stopwords, and lemmatize with `pymorphy3`. This results in 4,167 Russian unigram features and 3,473 English unigram features. Full preprocessing and model details are provided in Appendix A.2.

Tables 5 and 6 present the accuracy and F1 scores for the listed classification approaches applied to Russian-language text (left columns) or the translated English-language text (right columns). We assess these different approaches against final "gold standard" labels derived via majority vote from available human annotations. When classifying text as political, standard classifiers such as Logistic Regression, SVC, and Random Forests work slightly better with the translated English-language text than with

	Russian		English		
Model	Acc	F1	Acc	F1	R>E
SVC	0.797	0.806	0.836	0.845	0
Random Forest	0.810	0.823	0.823	0.841	0
Logistic Regression	0.813	0.818	0.846	0.852	0
GPT-o3-mini	0.854	0.871	0.864	0.881	0
GPT-4o	0.820	0.843	0.854	0.870	0

Table 5: Model Performance Results for the `POLITICAL` label across languages and models. Data source: VK posts, sample size 390 (20% test set). For the gold standard, we use majority votes among all human annotators as the true label. In cases of a tie, we assign a label of 1.

	Russian		English		
Model	Acc	F1	Acc	F1	R>E
SVC	0.977	0.866	0.977	0.866	1
Random Forest	0.977	0.862	0.967	0.787	1
Logistic Regression	0.972	0.853	0.979	0.892	0
GPT-o3-mini	0.992	0.960	0.985	0.917	1
GPT-4o	0.992	0.955	0.983	0.906	1

Table 6: Model Performance Results for the `PUTIN` label across languages and models. Data source: VK posts, n=390 (20% test set).

the Russian-language text. However, the differences for the Putin classification are substantively tiny and typically in the opposite direction. While such results are far from definitive, it is possible that off-the-shelf classifiers and standard pre-processing techniques have been optimized for English-language text. The GPT-based classifiers typically outperform other classifiers, with GPT-o3-mini slightly outperforming GPT-4o, and so we employ that LLM in analyses described subsequently to expand our sample sizes.

Over-time Correction

Above, we saw that for many annotation categories, the reliability of the human annotators was quite low. However, computational social scientists can sometimes avoid such issues through aggregation—after all, we are frequently interested not in individual-level analysis but instead in characterizing collectivities (Hopkins and King 2010). Accordingly, our results below include over-time analyses in which we aggregate posts at the weekly, monthly, or outlet level. While doing so, we employ the correction for misclassification outlined by Hopkins and King (2010), and so should recover unbiased over-time estimates so long as the mis-classification isn't correlated with time-varying factors. This technique allows us to move forward with the over-time analyses even as we remain concerned about some of the low levels of inter-coder reliability for individual posts revannotations.

Feature	VK (n=3500)	VK Gold (n=1950)
POLITICAL	55.63	55.38
PUTIN	7.80	8.87
PUTIN POSITIVE	0.71	1.33
PUTIN NEGATIVE	1.29	1.18
RUSSIAN GOVT	28.80	36.87
GOOD NEWS	4.26	2.00
BAD NEWS	6.29	2.51
NEGATIVE GOVT	5.54	3.74
NEAR ABROAD	12.63	15.44
US	5.80	8.72
DEMOCRACY	4.63	4.31
OPPOSITION	3.23	4.77
ETHNIC GROUPS	0.26	0.15
RUSSIANS ABROAD	0.20	1.23

Table 7: Proportion of VK posts in each category, as labeled by GPT-o3-mini (left) and by gold-standard human annotators (right). These proportions reflect all labeled VK samples.

Domain	N	Pol	PT+	PT-	PT	RG	GN	BN
aif.ru	100	.40	.03	.01	.05	.32	.03	.01
gazeta.ru	100	.60	.02	0	.09	.35	.05	0
iz.ru	100	.53	.01	0	.05	.30	.02	0
kommersant.ru	100	.39	0	0	.01	.59	.02	.01
lenta.ru	100	.41	.02	.01	.06	.24	0	.02
life.ru	100	.47	.02	0	.09	.31	.04	.01
mediazona.ru	100	.64	0	.05	.11	.54	0	.08
meduza.ru	100	.69	0	.04	.19	.41	0	.06
mk.ru	100	.48	.01	0	.05	.20	.01	.01
newizv.ru	100	.60	.01	.01	.11	.53	.01	.01
newtimes.ru	100	.33	0	0	0	.04	0	0
ng.ru	100	.58	.02	0	.05	.21	.05	0
novayagazeta.ru	100	.55	.01	.02	.09	.43	.01	.04
pravda.ru	100	.64	.02	0	.06	.26	.03	.03
rbc.ru	100	.81	0	0	.09	.54	.03	.09
republic.ru	100	.47	.02	.03	.10	.30	0	.05
rg.ru	50	.36	0	0	.02	.12	0	0
tvrain.ru	100	.78	.01	.05	.25	.59	.01	.02
vedomosti.ru	100	.65	.01	.01	.06	.45	.05	.02
vesti.ru	100	.60	.05	0	.21	.52	.03	.03

Table 8: VK sample average scores from the gold-standard annotated data for political, sentiment-related features. N= 1,950.

Results

Results: Descriptive Statistics

In Table 7, we report the fraction of VK posts which were annotated into categories labeled via GPT-o3-mini or human annotators. For "gold standard" human annotations, we choose the majority label where annotators disagree. (Results for Facebook are not shown but are similar.)

Roughly half of the posts are political. We find relatively low fractions of posts—around 8%—which mention Putin, but considerably more (between 29% and 37%) which mention the Russian government in some way. The US comes up occasionally in VK posts, and domestic opposition to the current Russian government comes up relatively infrequently, too. Still, given the inter-coder reliability concerns about some of these categories, we should be cautious in drawing broader lessons from them.

But how do such results vary by outlet? In Table 8 we present the results for our stratified random samples separately for each of our news outlets. As the Table makes clear, political content does vary by outlet, but not in ways that align with the outlet's relationship to the government. Several outlets which we label as "independent" even after 2022—mediazona.ru, meduza.ru, newizv.ru, newtimes.ru, novayagazeta.ru, republic.ru, and tvrain.ru—have relatively high levels of political coverage, but so do pro-government outlets including *Izvestia*. When regressing the values in the "political" column on an indicator for independent outlets, we find that those outlets are on average 0.08 lower in their mention of political content, but with a large standard error (0.075) given the paucity of observations.

Coverage of Politics over Time

Our analyses are on stronger footing when looking at trends over time, as we can aggregate using the over-time correction detailed in Hopkins and King (2010). In Figure 1, we present the over-time trends in several key categories using that correction for misclassification.

With respect to political content, we see a spike early in 2018 which may correspond to that year's Russian presidential election, and perhaps a decrease in political posts in 2020 as the COVID-19 pandemic set in. There is, though, a highly visible spike in early 2022 in political posts and also those mentioning the Russian government, which reflects the invasion of Ukraine that February. Notice that at that time, we see increases in news that appears to be negative for the Russian government as well as posts that contain negative news about the government.

Figure 2 allows us to focus exclusively on the "political" category, and to see the trends when we estimate that time-series using various approaches, from LLMs such as GPT-o3-mini or GPT-4 to some of the classifiers detailed in Tables 5 and 6 above. Here, the core finding is that the basic trends—and especially the spike in political content in early 2022 when Russia launched a full-scale invasion of Ukraine—show up using a variety of estimation techniques. For example, in the 3 months before the invasion—November 2021 to January 2022—the average fraction of the 85 hand-coded blurbs which were political was 0.47. In the subsequent 3 months, however, it leapt to 0.76 (n=114).[5]

Certainly, whether coverage is political or not is a single, coarse indicator. Appendix Figures 18 and 19 help us understand these trends in more detail by plotting the occurrence or co-occurrence of various keywords. For example, Figure 18 illustrates when media posts on VK which men-

[5] Appendix Figures 21 and 22 show the over-time trends in the English- and Russian-language annotations.

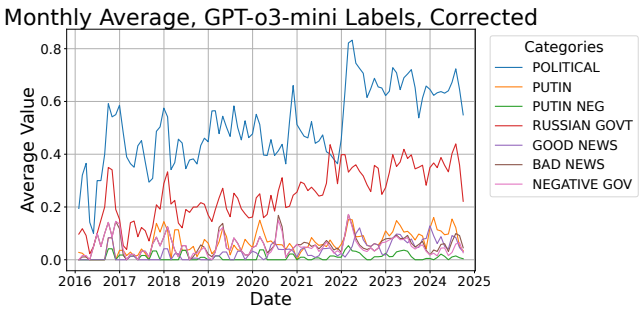

Figure 1: VK GPT-o3-mini annotations (n=7,000; data set includes annotations from Russian texts and texts translated into English).

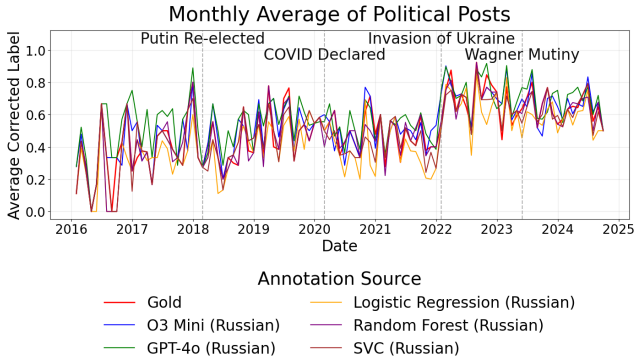

Figure 2: VK monthly percentage of annotations that are POLITICAL with key events overlaid ($n = 11{,}577$ posts from all sources; n=1,950 for each individual source except GPT-4o which is n=1,827).

tion "Ukraine" also mention key nouns such as "NATO" or "Casualty." While NATO is discussed extensively alongside Ukraine after the 2022 invasion—with a particular spike in 2023—hardly any media outlets' coverage of Ukraine employs "Nazi" or "Fascist." That said, "Casualties" received particular attention soon after the invasion, but continued to receive meaningful attention as the war continued.

In Appendix Figure 19, we advance this analysis by plotting the frequency of mentions of key political/governmental figures in our VK data set, including those of Putin alongside Dmitry Medvedev (former President and Deputy Chairman of the Security Council), Yevgeny Prigozhin (former head of the Wagner Group), Alexei Navalny (former opposition politician), and Sergei Shoigu (former Minister of Defense and Secretary of the Security Council). The figure makes clear that Putin receives far more attention than the other figures at almost every moment, with the exception of a spike in coverage of Navalny in late 2020 when he was poisoned and then in January 2021 when he returned to Russia and was immediately arrested. Mentions of Prigozhin also spike somewhat in mid-2023 after his unit's abortive mutiny and his subsequent death. We saw earlier that political coverage is far greater in volume than coverage of Putin specifically, but attention to Putin still dwarfs attention to select other political figures.

Engagement with Political Coverage

News on social media provides media outlets—and also government actors—with real-time information about social media users' demand for news content (Hopkins, Lelkes, and Wolken 2025). Such signals can be especially important in autocratic settings, where concerns about social desirability in surveys are pronounced and so where revealed-preference measures of interest are key (Frye 2022; Shirikov 2024). As a follow-up, we next probe whether VK users' engagement with political content changed after the February 2022 invasion of Ukraine. Specifically, we estimate Pearson's correlations between whether a VK blurb has political content—as estimated by GPT—and the logged number of comments, likes, and reposts of that blurb on the platform.

As Figure 3 illustrates, VK users are at times before 2022 more likely to comment on media posts with political content, but often *less* likely to like or repost such posts. However, there is a clear change with the February 2022 invasion, when political stories begin to generate meaningfully more likes and comments. The spike does tail off as the war goes on, especially for reposting.

Figure 4 shows that posts mentioning Putin typically draw more engagement throughout this period, too. However, the month-to-month relationship is variable and there is little clear acceleration after the February 2022 invasion. Even as consumers of Russian-language news were increasingly drawn to political content, they were not disproportionately drawn to Putin-related content.

Is the heightened engagement we observe actually a product of the war with Ukraine? Figure 5 shows the association between mentions of Ukraine and engagement on VK, which very clearly spikes with the invasion in early 2022. Certainly, the COVID-19 pandemic took place over this period, and so has the potential to confound any over-time inferences. However, in the early period after the invasion, there appears to be a decline in the connection between logged engagement and posts mentioning COVID. Moreover, Appendix Figure 20 tracks the association between engagement and COVID-related keywords, and generally shows more over-time stability alongside differing ebbs and flows. For these reasons, it seems reasonable to attribute the increased engagement with political content chiefly to the invasion of Ukraine.

Distinctiveness of Independent Outlets

Another valuable question is about the distinctiveness of independent outlets' posts, and whether that distinctiveness declined in the wake of the 2022 Ukraine invasion and the onerous new legal restrictions on media outlets.

To explore that question, we randomly sample 100,000 posts from the 1,623,310 posts in the VK data set. We then used logistic regression, fit to the Russian-derived features as described above, to annotate those posts as "political" or not, and restrict our sample to the 53,360 which were labeled "political."[6] Then, we convert posts into embeddings us-

[6]Note that GPT-o3-mini was not available at the time when

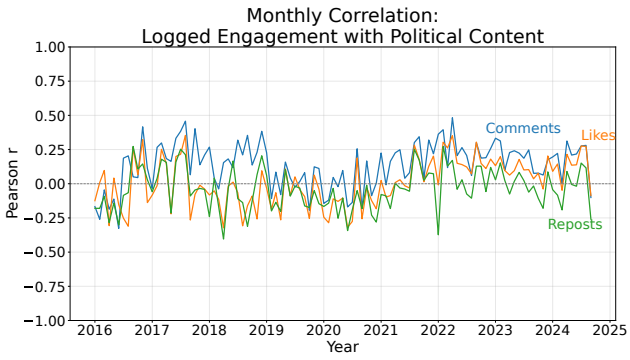

Figure 3: Time series of correlations between the gold standard "Political" category and logged engagement on VK. Data set is all unique posts labeled by majority rule ($n = 9,651$ posts).

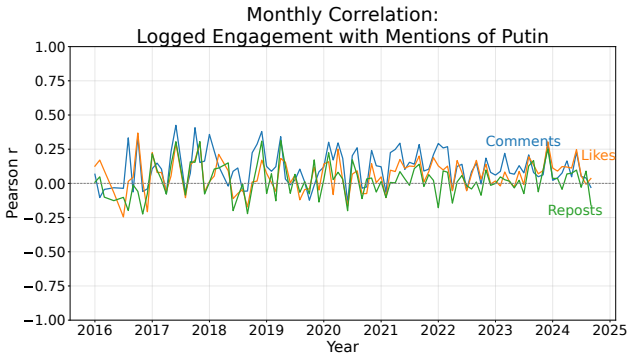

Figure 4: Time series of correlations between mentions of Putin from gold standard labels and logged engagement. Data set is all unique posts labeled by majority rule ($n = 9,651$ posts).

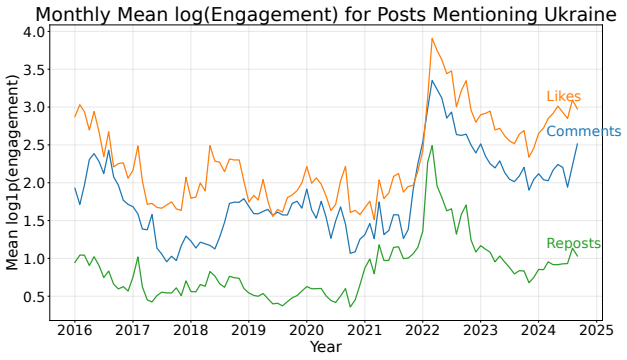

Figure 5: Correlations between mentions of "Ukraine" from keyword searches and logged engagement over time. ($n = 172,218$ posts of $1,623,310$ mentioning Ukraine).

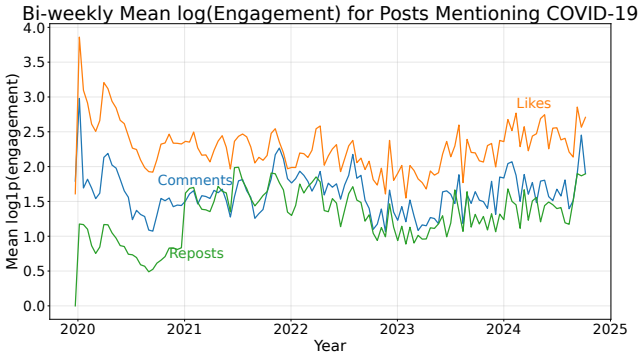

Figure 6: Correlations between mentions of "COVID-19" and "pandemic" from keyword searches and logged engagement on VK since 2020 ($n = 58,169$ posts of $1,623,310$ mentioning COVID).

ing text-embedding-3-small. After separating the posts into those from the independent outlets and those from outlets that are pro-government, we estimate the cosine similarity between all pairs of posts in the two groups. In using cosine similarity to measure the distance between texts, we follow work including Carlson (2019) and Hager and Hilbig (2020).

Figure 7 illustrates a spike in similarity in early 2022, right as the invasion began. During that initial period, the language between independent outlets and other, more pro-government outlets did converge. But that spike in convergence was short-lived, and it echoes a similar spike at the onset of COVID-19 pandemic. In fact, the mean cosine similarity for the period through January 2022 was 0.199, while for the period from February 2022 onward is was very slightly but significantly higher (0.204).[7] Thus, it may well reflect a more common vocabulary to describe new events (such as COVID-19 or the invasion) as well as the effects of the new laws criminalizing certain coverage of the war and military.

Regression Models

An alternative approach to determine the distinctiveness of independent media is to estimate models of the post-level data predicting political coverage. We do so separately in Table 9 for posts with human labels. Employing two-way fixed-effects models with fixed effects for the media outlet and month, we estimate interactions between independent ownership and the period after January 1, 2022. In Figure 8, we supplement these models by reporting the interactions between independent ownership and month.

There is a demonstrable decline in independent outlets' political coverage beginning in 2022 in the human annotations (evident in GPT annotations as well). For the human annotations in Figure 8, there is a discernible drop in independent outlets' relative coverage of politics after 2022.

Conclusion

A central goal of research on autocracies has been to identify variants of autocratic politics and the mechanisms through

conducted these analyses, that classification via logistic regression scales much more affordably than via private LLM, and that Logistic Regression typically performs well in the classification exercises in Tables 5 and 6 above.

[7] $p<0.001$ from a two-sample t-test of the difference in means.

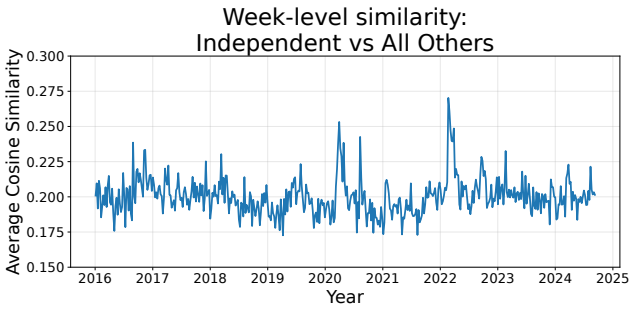

Figure 7: VK Monthly Embedding distance independent vs all others

Interaction Term	Coefficient	Std. Error
Independent Media × Post Dec. 2021	-0.302	0.074

Table 9: Interaction Coefficient (Independent Outlet × Month) from human annotations (disagreements decided by majority rule). Outcome is political content. N=1,950. Fixed effects for outlet and month included but not shown. Standard errors clustered by media outlet.

which they sustain power (Svolik 2012; Gehlbach, Sonin, and Svolik 2016; Baturo and Elkink 2021). One important element of that variation is in how they manage and regulate the information that reaches citizens (Gehlbach and Sonin 2014), especially as the information environment fragments.

This paper uses various Natural Language Processing tools to document over-time variation in a particularly impactful country (Russia) and sector (news content conveyed via social media). Specifically, we seek to contribute to understanding media ecosystems in autocratic regimes by documenting the changes in political news coverage on social media before and after Russia's 2022 invasion of Ukraine. To do so, we employ LLMs and extensive human annotation to generate and analyze a novel data set of prominent newspapers and independent media outlets' social media posts in the period between 2016 and 2024. For more subjective categories, teams of trained Russian- and English-speaking annotators were unable to produce reliable individual-level annotations, a cautionary note about recovering complex, context-dependent constructs.

However, our pipeline was able to reliably measure several constructs, especially after employing a correction for individual-level misclassification. The constructs we focus on here include whether a given social media post was political or mentioned President Putin. There, we find that February 2022 corresponded with critical changes in the Russian media ecosystem and the Russian government's role within it. Outlets provided more political coverage, and VK users were more likely to engage with the political coverage they did provide. Independent outlets produced political coverage that was more similar to pro-government outlets, but also saw their reach on social media increasingly curtailed. It seems, then, that February 2022 reflected a shift away from a strategy of de-politicization, and inaugurated a period of

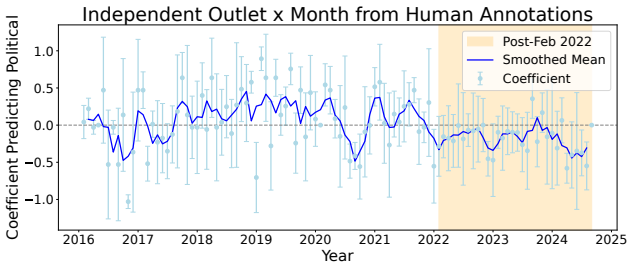

Figure 8: Coefficient on interactions between independent outlet and month indicators when predicting political content. Dataset: human annotations over time.

more overt restrictions on media outlets that were not pro-government. Prior to 2022, the Russian state was already undertaking extensive efforts to shape news coverage, and these efforts greatly intensified after the invasion. As the sum total of our evidence makes clear, 2022 represented a watershed movement that inaugurated a new, more heavy-handed approach to censorship and control of information.

Limitations and Future Research

Our analysis has various limitations, several of which represent areas for improvement in future research. Our analyses focus primarily on VK blurbs rather than other social media platforms, so platform-specific policies and content distribution techniques may influence exposure and survivorship. As a consequence, this research is likely to provide a partial accounting of social media activity in Russia. Such analyses could productively be expanded to include non-textual information (such as images) and a wider range of media outlets (including television channels).. As detailed above, several labels exhibit only moderate inter-coder agreement, and translation/LLM choices may add noise; improving measurement and identifying which higher-order constructs are not amenable to consistent annotation is a key avenue for future work.

The engagement estimates detailed here are correlational and may be confounded by unobserved, time-varying shocks, limiting our capacity to make causal inferences. For example, we do not know to what extent the new regime of increased censorship and repression of independent media outlets was planned before the invasion versus being a response to initial failures of some parts of the Russian invasion and the corresponding spike in media interest. That said, our varied analyses of news coverage and engagement with that coverage have documented clearly that the Russian news ecosystem changed in a meaningful and durable way right around the February 2022 invasion of Ukraine.

Acknowledgments

The authors thank research assistants including Aruzhan Aussat, Griffin Bond, Kaynath Chowdhury, Amina Ford, Edwin Klanke, Kseniya Shalanskaya, Yuliya Solyanyk, and Dina Zhanybekova and research coordinators Nilou Davis and Gall Sigler. The authors also thank members of the Uni-

versity of Pennsylvania American Politics Working Group (July 2025), attendees at the 2025 Annual Meeting of the Society for Political Methodology (July 2025; Emory University) and Adeline Lo, Jane Esberg, and Anton Shirikov for incisive comments and/or feedback. The authors thank the University of Pennsylvania School of Arts and Sciences for funding and have no competing interests to disclose.

References

Applebaum, A. 2024. Democracy is Losing the Propaganda War. *The Atlantic*, June: 30–40.

Arendt, H. 1948. *The Origins of Totalitarianism*. Harcourt Brace.

Baturo, A.; and Elkink, J. A. 2021. *The new Kremlinology: understanding regime personalization in Russia*. Oxford University Press.

BBC. 2022. Russia confirms Meta's designation as extremist.

Bennett, W. L.; and Iyengar, S. 2008. A new era of minimal effects? The changing foundations of political communication. *Journal of communication*, 58(4): 707–731.

Bjorklund, K.; and Smith, S. J. 2023. Ukraine war: reports suggest that Russia has been deliberately targeting journalists—which is a war crime.

Carlson, T. N. 2019. Through the grapevine: Informational consequences of interpersonal political communication. *American Political Science Review*, 113(2): 325–339.

Colton, T. J. 2016. *Russia: What Everyone Needs to KnowR*. Oxford University Press.

DataReportal. 2019. Digital 2019: The Russian Federation.

DataReportal. 2025. Digital 2025: The Russian Federation.

Denny, M. J.; and Spirling, A. 2018. Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political analysis*, 26(2): 168–189.

Egami, N.; Hinck, M.; Stewart, B. M.; and Wei, H. 2024. Using Large Language Model Annotations for the Social Sciences: A General Framework of Using Predicted Variables in Downstream Analyses.

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

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

Field, A.; Kliger, D.; Wintner, S.; Pan, J.; Jurafsky, D.; and Tsvetkov, Y. 2018. Framing and agenda-setting in Russian news: a computational analysis of intricate political strategies. *arXiv preprint arXiv:1808.09386*, 1–10.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Fredheim, R. 2017. The loyal editor effect: Russian online journalism after independence. *Post-Soviet Affairs*, 33(1): 34–48.

Frye, T. 2022. *Weak Strongman: The Limits of Power in Putin's Russia*. Princeton University Press.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Gehlbach, S.; Lokot, T.; and Shirikov, A. 2023. The Russian Media. In Wengle, S., ed., *Russian Politics Today: Stability and Fragility*, 390–407.

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

Gehlbach, S.; Sonin, K.; and Svolik, M. W. 2016. Formal models of nondemocratic politics. *Annual Review of Political Science*, 19(1): 565–584.

Gerschewski, J. 2023. *The two logics of autocratic rule*. Cambridge University Press.

Gilardi, F.; Alizadeh, M.; and Kubli, M. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30): e2305016120.

Goode, J. P. 2025. Russian Propaganda from V to Z: Projecting Banal and Everyday Nationalism in Unsettled Times. *Nationalities Papers*, 1–21.

Grimmer, J.; Roberts, M. E.; and Stewart, B. M. 2022. *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Guriev, S.; and Treisman, D. 2018. Informational Autocracy: Theory and Empirics of Modern Authoritarianism. *Available at SSRN 2571905*, 1–61.

Hager, A.; and Hilbig, H. 2020. Does public opinion affect political speech? *American Journal of Political Science*, 64(4): 921–937.

Halterman, A.; and Keith, K. A. 2025. Codebook LLMs: Evaluating LLMs as Measurement Tools for Political Science Concepts.

Hjermann, A. R. 2023. Depoliticising democracy through discourse: Reading Russia's descent into autocracy and war with Jacques Rancière's political theory. *New Perspectives. Interdisciplinary Journal of Central & East European Politics and International Relations*, 31(2): 49–76.

Hopkins, D. J.; and King, G. 2010. A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54(1): 229–247.

Hopkins, D. J.; Lelkes, Y.; and Wolken, S. 2025. The rise of and demand for identity-oriented media coverage. *American Journal of Political Science*, 69(2): 483–500.

King, G.; Pan, J.; and Roberts, M. E. 2013. How censorship in China allows government criticism but silences collective expression. *American political science Review*, 107(2): 326–343.

King, G.; Pan, J.; and Roberts, M. E. 2014. Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science*, 345(6199): 1251722.

Koesel, K. J.; and Bunce, V. J. 2012. Putin, popular protests, and political trajectories in Russia: A comparative perspective. *Post-soviet affairs*, 28(4): 403–423.

Krippendorff, K. 1970. Estimating the Reliability, Systematic Error and Random Error of Interval Data. *Educational and Psychological Measurement*, 30: 61–70.

Krippendorff, K. 2004. Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human Communication Research*, 30(3): 411–433.

Lankina, T.; and Watanabe, K. 2017. 'Russian spring'or 'spring betrayal'? The media as a mirror of Putin's evolving strategy in Ukraine. *Europe-Asia Studies*, 69(10): 1526–1556.

Litvinenko, A.; and Nigmatullina, K. 2020. Local dimensions of media freedom: A comparative analysis of news media landscapes in 33 Russian regions. *Demokratizatsiya: The Journal of Post-Soviet Democratization*, 28(3): 393–418.

Lu, Y.; Pan, J.; Xu, X.; and Xu, Y. 2025. Decentralized propaganda in the era of digital media: The massive presence of the Chinese state on Douyin. *American Journal of Political Science*, 1–17.

Marx, K.; and Engles, F. 1845. On the German Ideology. In Tucker, R., ed., *The Marx-Engles Reader*, 146–203.

Mattingly, D. C.; and Yao, E. 2022. How soft propaganda persuades. *Comparative Political Studies*, 55(9): 1569–1594.

Mediascope. 2025. Ratings.

Miller, B. 2025. Fragmented Censorship: How Bureaucratic and Market Forces Constrain China's Information Control Agenda.

Moscow Times. 2021. Gazprom Gains Control of Russia's Top Social Network.

Nieman, M. D.; and Labzina, E. 2025. State-controlled Media and Foreign Policy: Analyzing Russian-language News.

Nikolich, A.; Korolev, K.; Bratchikov, S.; Kiselev, I.; and Shelmanov, A. 2024. Vikhr: Constructing a state-of-the-art bilingual open-source instruction-following large language model for Russian. In *Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024)*, 189–199.

Pangakis, N.; and Wolken, S. 2025. Keeping humans in the loop: Human-centered automated annotation with generative ai. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, 1471–1492.

Pangakis, N.; Wolken, S.; and Fasching, N. 2023. Automated annotation with generative ai requires validation. *arXiv preprint arXiv:2306.00176*, 1–19.

Park, C. Y.; Mendelsohn, J.; Field, A.; and Tsvetkov, Y. 2022. Challenges and Opportunities in Information Manipulation Detection: An Examination of Wartime Russian Media. *EMNLP*, 5209–5235.

Pomerantsev, P. 2014. *Nothing is true and everything is possible: The surreal heart of the new Russia*. Public Affairs.

Porter, E.; Scott, R. B.; Wood, T. J.; and Zhandayeva, R. 2024. Correcting misinformation about the Russia-Ukraine War reduces false beliefs but does not change views about the War. *Plos one*, 19(9): e0307090.

Rathje, S.; Mirea, D.-M.; Sucholutsky, I.; Marjieh, R.; Robertson, C. E.; and Van Bavel, J. J. 2024. GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences*, 121(34): e2308950121.

Roberts, M. 2018. *Censored: distraction and diversion inside China's Great Firewall*. Princeton University Press.

Rosenfeld, B. 2021. *The autocratic middle class: how state dependency reduces the demand for democracy*. Princeton University Press.

Rosenfeld, B.; and Wallace, J. 2024. Information politics and propaganda in authoritarian societies. *Annual Review of Political Science*, 27.

Rozenas, A.; and Stukal, D. 2019. How autocrats manipulate economic news: Evidence from Russia's state-controlled television. *The Journal of Politics*, 81(3): 982–996.

Shirikov, A. 2024. Rethinking propaganda: how state media build trust through belief affirmation. *The Journal of Politics*, 86(4): 1319–1332.

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

Svolik, M. W. 2012. *The politics of authoritarian rule*. Cambridge University Press.

Thompson Reuters Foundation; and the Committee to Protect Journalists. 2022. Understanding the laws relating to "fake news" in Russia.

Tikhomirov, M.; and Chernyshev, D. 2023. Impact of tokenization on llama russian adaptation. In *2023 Ivannikov Ispras Open Conference (ISPRAS)*, 163–168. IEEE.

Törnberg, P. 2024. Large language models outperform expert coders and supervised classifiers at annotating political social media messages. *Social Science Computer Review*, 08944393241286471.

Troianovski, A.; and Safronova, V. 2022. Russia Takes Censorship to New Extremes, Stifling War Coverage.

Vendil Pallin, C. 2017. Internet control through ownership: The case of Russia. *Post-Soviet Affairs*, 33(1): 16–33.

Waight, H.; Yuan, Y.; Roberts, M. E.; and Stewart, B. M. 2025. The decade-long growth of government-authored news media in China under Xi Jinping. *Proceedings of the National Academy of Sciences*, 122(11): e2408260122.

Wallach, H.; Desai, M.; Pangakis, N.; Cooper, A. F.; Wang, A.; Barocas, S.; Chouldechova, A.; Atalla, C.; Blodgett, S. L.; Corvi, E.; et al. 2024. Evaluating Generative AI Systems is a Social Science Measurement Challenge. *arXiv preprint arXiv:2411.10939*.

Wang, X.; Kim, H.; Rahman, S.; Mitra, K.; and Miao, Z. 2024. Human-llm collaborative annotation through effective verification of llm labels. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–21.

Paper Checklist

1. For most authors...

(a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes. The study analyzes public posts from media outlets and discusses reliability and misclassification and translation limits. No privacy violations are evident.

(b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes. The abstract and intro emphasize post-invasion shifts, engagement changes, outlet similarity, and methodological validation—matching the body.

(c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes. We justify human and LLM annotation, inter-coder reliability, confusion matrices, and over-time misclassification correction.

(d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes. We discuss language/translation effects, outlet mix, VK vs. Facebook coverage windows, and low reliability on subjective labels.

(e) Did you describe the limitations of your work? Yes. We have limitation section.

(f) Did you discuss any potential negative societal impacts of your work? No, because we envision positive implications of our findings discussed in Results

(g) Did you discuss any potential misuse of your work? NA

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes. In Appendix A.

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.

2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? NA

(b) Have you provided justifications for all theoretical results? NA

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA

(e) Did you address potential biases or limitations in your theoretical framework? NA

(f) Have you related your theoretical results to the existing literature in social science? NA

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes. All releasable data, code, prompts and instructions are provided in the supplement materials.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No.

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes.

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? Yes. We cite all models/tools used.

(b) Did you mention the license of the assets? NA

(c) Did you include any new assets in the supplemental material or as a URL? Yes. We have data and code in the supplemental material.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes. Discussed in Appendix A how we remove identifiable information.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? Yes.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? Yes. In Appendix A.1.

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? Yes. Human annotation instructions in Appendix B and GPT prompts in Appendix C.

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? No. Analyses were of anonymized, secondary data; no IRB approval required.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Yes

(d) Did you discuss how data is stored, shared, and deidentified? Yes.

Appendix A: Reproducibility Details

A.1 Data

Overview and release. We provide a de-identified corpus of organizational social media blurbs and labels. Two CSVs are included in the released github folder[8]: `vk_posts_cleaned_122825.csv` and `vk_labels_cleaned_122825.csv`. This distribution follows platform Terms of Service. The content is only from public, organizational accounts.

How the data were collected. We scraped VK posts via the VK API; Meta/Facebook blurbs were obtained via approved access through the Social Science One portal. We retain outlet/domain, timestamps, engagement counts, and text fields where available.

Data samples.

Appendix Figures 9, 10, and 11 show sample posts from AIF (*Arguments and Facts*) across the three platforms. These examples come directly from the content we scraped via the platforms' APIs. The GPT-4o translation of the post shown in Figure 9 reads as follows: "The Russian Armed Forces struck a military unit of the Ukrainian Armed Forces in Odessa and warehouses with rocket fuel in the Sumy region, said underground operative Sergey Lebedev. 'It hit the outskirts of Nerubayskoye. There is a military unit located there,' the statement about the strike on Odessa said in the Telegram channel of the coordinator of the Mykolaiv underground. According to him, warehouses with ammunition detonated there, and there are casualties among Ukrainian servicemen. Russian forces also struck warehouses of the Ukrainian Armed Forces with rocket fuel and missile components in the city of Shostka in the Sumy region of Ukraine.

[8] https://github.com/krystalgong/RussianMediaOutletsDatasets

Figure 9: Screenshot of AIF post on VK.

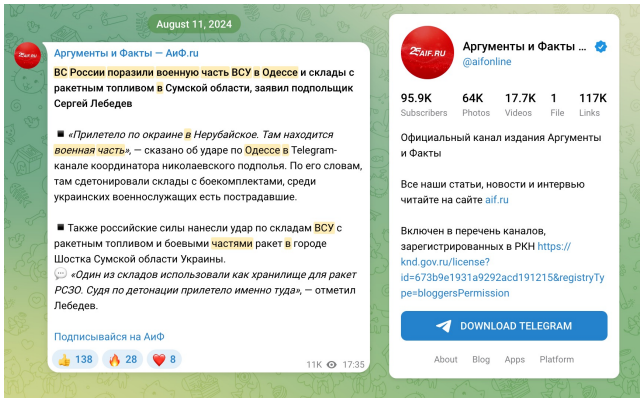

Figure 10: Screenshot of AIF post on Telegram.

Figure 11: Screenshot of AIF post on Facebook.

'One of the warehouses was used as a storage facility for MLRS rockets. Judging by the detonation, it hit exactly there,' Lebedev noted."

How we de-identify and process posts. Records use synthetic `id` keys starting from 1; no private-user identifiers are included. Outlets appear as domains (e.g., `aif.ru`). English translations, where present, are stored in `post_en`. Labels derive from multiple sources (human Russian/English, LLMs, classical ML, and keyword heuristics) and are tracked via `label_source`. We apply minimal normalization to facilitate reproducibility.

VK Tables and schema. Shown in Table 10 and Table 11. Label_source details are shown in Table 12

FAIR commitments. Following FORCE11 FAIR principles, we make releases *Findable* (versioned filenames); *Accessible* (publicly released data with accompanying documentation); *Interoperable* (UTF-8 CSVs, ISO 8601 timestamps, tidy schemas, and documented data dictionaries); and *Reusable* (explicit licenses for released datasets; this datasheet and stated limitations clarify the scope and appropriate reuses).

How the data are stored and shared. Source and working copies reside on secure institutional storage with routine backups and checksums; public artifacts are exported as CSV files. All data sharing complies with the platforms' Terms of Service.

Field	Type	Description
media	string	Outlet/domain (e.g., aif.ru).
post	string (RU)	Blurb text in Russian.
date	datetime (ISO)	Post timestamp.
comments, likes, reposts	int	Platform engagement counts.
POLITICAL_prob	float	Probability score for POLITICAL from Logistic Regression.
id	string/int	Synthetic post identifier; join key to labels.
post_en	string (EN)	Machine translation (may be NaN).

Table 10: VK Posts table (join to labels via id).

Field	Type	Description
POLITICAL, PUTIN, PUTINPOS, PUTINNEG, RUSSIANGOVT, GOODNEWS, BADNEWS, NEGATIVEGOV, NEARABROAD, US, DEMOCRACY, OPPOSITION, ETHGROUP, RUSSIANSABROAD	{0,1}	Binary labels per codebook.
id	string/int	Synthetic post identifier; join key to posts.
label_source	string	Label provenance (human/LLM/ML/heuristic).

Table 11: VK Labels table (join to posts via id).

We do not release code; instead, all table schemas are fully documented in this Appendix in lieu of a separate README.

A.2 Models

For Russian-language posts, we convert text to lowercase, tokenize it using Cyrillic-aware regular expressions, remove Russian stopwords from NLTK, and lemmatize each token using the pymorphy3 morphological analyzer. For English-language posts, we transform text to lowercase, tokenize it using a Latin-character regular expression, remove English stopwords from NLTK, and apply Snowball stemming. In both languages, preprocessing returns a sequence of normalized tokens rather than raw strings.

We represent texts using bag-of-words unigrams constructed with CountVectorizer, supplying language-specific tokenizers and disabling internal lowercasing. We impose no document-frequency thresholds or vocabulary size limits.

label_source	Description
o3_mini_ru, o3_mini_en, gpt_4o_ru, gpt_4o_en	GPT labels
RandomForest_ru, RandomForest_en, LogisticRegression_ru, LogisticRegression_en, SVC_ru, SVC_en	Machine Learning labels
coder_ru_1, coder_ru_2, coder_en_1, coder_en_2, coder_en_3, coder_en_4	Human coder annotations
gold_1123, human_en, human_ru	Majority Vote from human coder annotations

Table 12: Label_source in VK labels

For each language (post for Russian, post_en for English) and each target label (POLITICAL, PUTIN, PUTINPOS, PUTINNEG, RUSSIANGOVT, GOODNEWS, BADNEWS), we train three classifiers implemented in scikit-learn: (1) Logistic Regression (max_iter=1000), (2) a linear Support Vector Classifier with probability estimates (SVC with kernel="linear", probability=True), and (3) a Random Forest classifier with 100 trees (RandomForestClassifier with n_estimators=100). Specifically, we employ L2-regularized logistic regression to downweight uninformative features.

For each label, we use a fixed 80/20 train-test split stratified on the relevant label. Models are trained on the training set and evaluated on the held-out test set. In addition to standard binary predictions, we compute calibrated probabilities using a post-hoc adjustment based on sensitivity and specificity estimated from the test-set confusion matrix, which is then applied to model scores over the full data set.

Our evaluation focuses on accuracy and F1 scores compared with the majority labels from human annotations. For LLM workflows, we use GPT-4 for machine translation (Russian→ English) and GPT-4 and GPT-o3-mini for direct annotation in both Russian and English. Appendix C provides prompts and settings. For the embedding-based similarity analyses, we compute sentence embeddings with text-embedding-3-small and measure cosine similarity as reported in the Results section.

Appendix B: Human Annotation Protocol

We provide the full codebook and annotator instructions in this section. Student research assistants labeled public, organizational posts in two streams: original Russian-language text and English machine translations done by GPT-4. Annotators worked independently, with overlapping assignments to enable reliability checks; we compute Krippendorff's α and report agreement statistics in the paper. We derived the final "gold" labels via majority vote from available human

annotations (language-specific and harmonized snapshots). No personal information about annotators was collected. Annotators were compensated at $15/hour (total compensation: $1,875). All data and labels are stored on secure institutional servers with access control. Public releases are de-identified and comply with platforms' Terms of Service.

POLITICAL
Type: 0/1
Definition: 1 if there is discussion of politics or government. This category includes explicit references to politics, government institutions, policies, military actions, or diplomacy. Broad claims about government authorities (e.g., "corruption among government authorities") are included. Mentions of institutions controlled by the government (e.g., schools, local police, state-owned enterprises) are not sufficient unless in a political context. References to political leaders are not sufficient unless discussing government or politics.

DOMESTIC
Type: 0/1
Definition: If `POLITICAL = 1`, this is 1 if the text primarily discusses domestic (internal Russian) issues.

UKRAINE
Type: 0/1
Definition: If `POLITICAL = 1`, this is 1 if the text primarily discusses Ukraine.

NEARABROAD
Type: 0/1
Definition: If `POLITICAL = 1`, this is 1 if the text primarily discusses any near-abroad country (e.g., Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, including Ukraine).

WHICHNEARABROAD
Type: Text
Definition: If `NEARABROAD = 1`, specify which countries are discussed.

RUSSIANSABROAD
Type: 0/1
Definition: If `POLITICAL = 1`, this is 1 if the article primarily describes the experiences of Russians abroad or perceptions of Russia abroad.

DEMOCRACY
Type: 0/1
Definition: If `POLITICAL = 1`, this is 1 if the article discusses democratic activities (e.g., elections, protests).

OPPOSITION
Type: 0/1
Definition: If `POLITICAL = 1`, this is 1 if the article discusses political opposition to the Russian government.

FOREIGN
Type: 0/1
Definition: If `POLITICAL = 1`, this is 1 if the article primarily discusses a foreign country (any country other than Russia).

WHICHFOREIGN
Type: Text
Definition: If `FOREIGN = 1`, specify which country is discussed.

PUTIN
Type: 0/1
Definition: 1 if Vladimir Putin is explicitly mentioned (either by name or reference such as "President of Russia").

PUTINPOS
Type: 0/1
Definition: If `PUTIN = 1`, this is 1 if the writer's tone toward Putin is unambiguously positive.

PUTINNEG
Type: 0/1
Definition: If `PUTIN = 1`, this is 1 if the writer's tone toward Putin is unambiguously negative.

RUSSIANGOVT
Type: 0/1
Definition: 1 if there is an explicit mention of a specific Russian government agency or political figure by name

GOODNEWS
Type: 0/1
Definition: If `RUSSIANGOVT = 1`, 1 if the article is unambiguously good news for the Russian government.

BADNEWS
Type: 0/1
Definition: If `RUSSIANGOVT = 1`, 1 if the article is unambiguously bad news for the Russian government.

NEGATIVEGOV
Type: 0/1
Definition: If `RUSSIANGOVT = 1`, this is 1 if the tone is explicitly negative toward the Russian government or its policies.

POSITIVEGOV
Type: 0/1
Definition: If `RUSSIANGOVT = 1`, this is 1 if the tone is explicitly positive toward the Russian government or its policies.

US
Type: 0/1
Definition: 1 if the article primarily discusses the United States.

HISTORY
Type: 0/1
Definition: 1 if the article discusses Russian or Soviet history before 1991.

ETHGROUP
Type: 0/1
Definition: 1 if the article primarily discusses ethnic or racial groups (e.g., Ukrainians, Tatars, Estonians).

WHICHETHGROUP
Type: Text
Definition: If `ETHGROUP = 1`, specify which ethnic or racial groups are discussed.

Appendix C: GPT Prompts of Translation and Annotation

Prompt to translate Russian posts into English version

```
Translate the following Russian text to English:\n\nText: \"{text}\"
```

Prompt to annotate posts into structured outputs:

```
Annotate the following text and format the output in JSON, including all annotations with either 0 or 1. \n
   Following is the codebook: \n
POLITICAL -- 0/1, 1 if there is discussion of politics/government. This category should only include explicit references to politics, government institutions, government policy, military actions, or diplomacy. Broad claims about government/authorities (such as "corruption among government authorities") should be coded as 1. References to institutions that are controlled by the government (such as schools, local police, state-owned businesses) are not sufficient for this category unless they are mentioned in a political context. References to political leaders are not sufficient unless the context of the mention is about government and politics.
  IF POLITICAL=1
      DOMESTIC -- 0/1, 1 if this primarily discusses domestic issues
      UKRAINE -- 0/1, if this primarily discusses Ukraine
      NEARABROAD -- 0/1, 1 if the article primarily discusses any near-abroad country (e.g. former members of the USSR: Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, and Lithuania; including Ukraine)
       IF NEARABROAD=1
          WHICHNEARABROAD -- text, enter which countries are discussed
      RUSSIANSABROAD -- 0/1 if this primarily describes the experiences of Russians abroad, or the perceptions of Russia abroad.
      DEMOCRACY -- 0/1, if this primarily discusses democratic activities, such as elections and protests
      OPPOSITION -- 0/1 if this primarily discusses political opposition to the Russian government
      FOREIGN -- 0/1 if this primarily discusses a foreign country, meaning any country other than Russia
       IF FOREIGN =1
          WHICHFOREIGN -- text, which foreign country does it discuss
PUTIN -- 0/1, 1 if there is an explicit mention of Vladimir Putin, whether by name or by referencing or "president of Russia"
   IF PUTIN=1
      PUTINPOS -- 0/1, 1 if the writer's tone towards Putin is unambiguously positive
      PUTINNEG -- 0/1, 1 if the writer's tone towards Putin is unambiguously negative
RUSSIANGOVT -- 0/1, 1 if there is an explicit mention of a specific Russian government agency or Russian political figure by name
   IF RUSSIANGOVT=1
      GOODNEWS -- 0/1, 1 if this is unambiguously good news for the Russian government
      BADNEWS -- 0/1, 1 if this is unambiguously bad news for the Russian government
      NEGATIVEGOV -- 0/1, 1 if the tone of the text is explicitly negative towards the Russian government and its policies
      POSITIVEGOV -- 0/1, 1 if the tone of the text is explicitly positive towards the Russian government and its policies
US -- 0/1, if this primarily discusses the United States
HISTORY -- 0/1, if this discusses Russian/Soviet history from before 1991
ETHGROUP -- 0/1 if this primarily discusses ethnic/racial groups (e.g. Ukrainians, Tatars, Estonians, etc.)
   IF ETHGROUP =1
      WHICHETHGROUP -- text, which ethnic/racial groups does
```

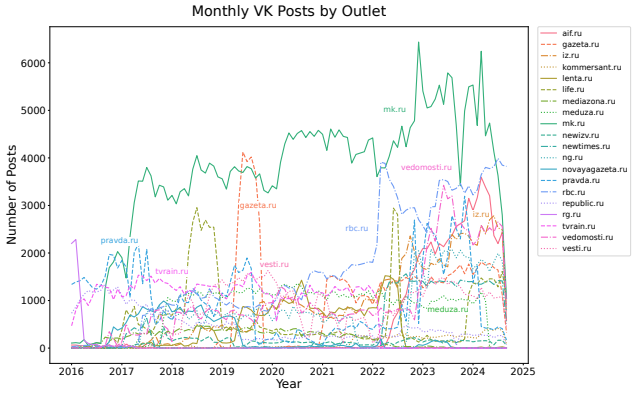

Figure 12: Monthly VK posts by outlet.

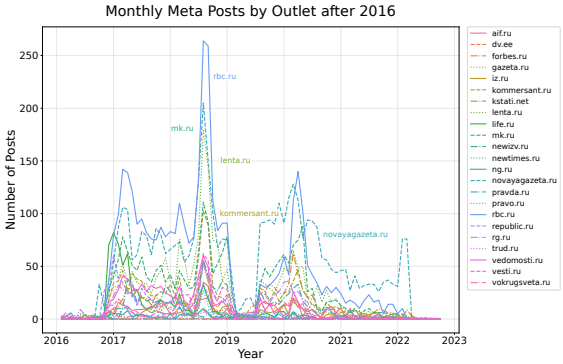

Figure 13: Monthly Facebook posts by outlet.

```
            it discuss?
    Note: only return the JSON format
        but nothing else. Only return the
         features in the following output
         example. \n\n\
    Output example: {"POLITICAL": 1, "
        DOMESTIC": 0, "UKRAINE": 0, "
        NEARABROAD": 0, "RUSSIANSABROAD":
         0, "DEMOCRACY": 1, "OPPOSITION":
         1, "FOREIGN": 1, "PUTIN": 1, "
        PUTINPOS": 0, "PUTINNEG": 1, "
        RUSSIANGOVT": 1, "GOODNEWS": 0, "
        BADNEWS": 1, "NEGATIVEGOV": 1, "
        POSITIVEGOV": 0, "US": 0, "
        HISTORY": 0, "ETHGROUP": 0}
```

Appendix D: Additional VK, Facebook, and Telegram Results

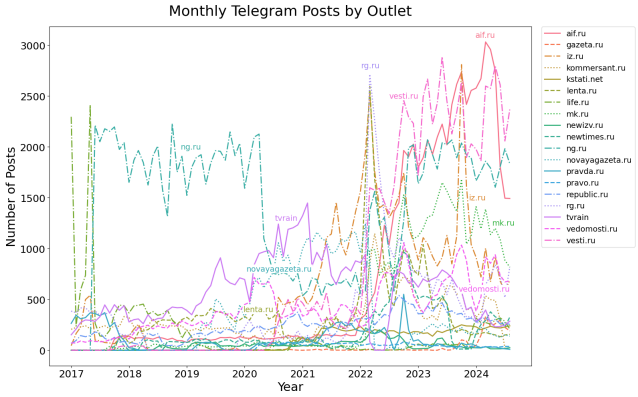

Figure 14: Monthly Telegram posts by outlet.

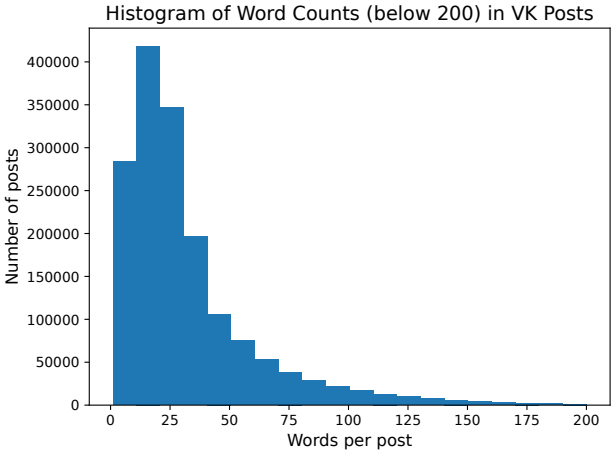

Figure 15: Word Count for VK posts.

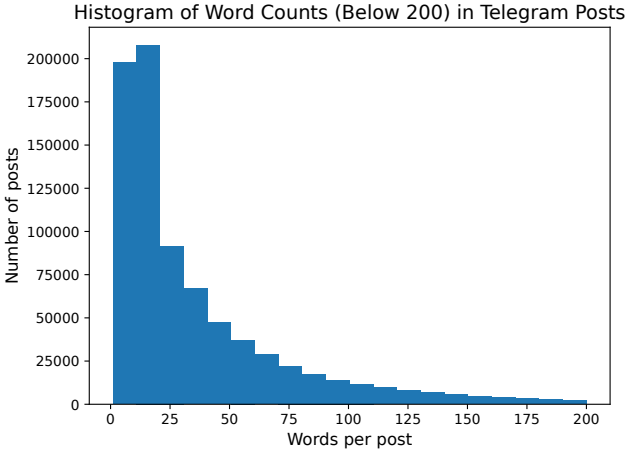

Figure 16: Word Count for Telegram posts.

parent_domain	earliest_date	latest_date	total_posts
aif.ru	2011-11-22	2022-01-29	532
dv.ee	2020-04-01	2020-12-24	3
forbes.ru	2012-01-21	2022-10-10	1062
gazeta.ru	2016-05-31	2021-12-29	726
iz.ru	2017-06-07	2022-04-10	375
kommersant.ru	2013-11-03	2022-07-18	1346
lenta.ru	2011-11-19	2022-03-07	2130
life.ru	2016-02-15	2021-12-11	726
mk.ru	2013-10-14	2022-07-13	2212
newizv.ru	2016-04-20	2022-09-16	911
newtimes.ru	2016-12-22	2022-02-06	182
ng.ru	2013-09-20	2022-06-06	113
novayagazeta.ru	2014-04-26	2022-04-12	4389
pravda.ru	2016-01-09	2022-03-16	149
rbc.ru	2011-01-06	2022-05-27	3891
republic.ru	2016-11-05	2022-07-01	582
rg.ru	2016-06-10	2022-06-23	891
trud.ru	2016-12-28	2020-06-19	13
vedomosti.ru	2015-04-19	2021-11-30	866
vesti.ru	2012-09-04	2021-05-18	676

Table 13: Summary of Facebook/Meta Posts by Media Outlet.

feature	All	All GPT	All human	Human RU	Human EN
POLITICAL	0.559	0.870	0.476	0.497	0.520
PUTIN	0.826	0.957	0.781	0.798	1.000
PUTIN POSITIVE	0.319	0.570	0.252	0.466	0.496
PUTIN NEGATIVE	0.436	0.590	0.386	0.327	0.663
RUSSIAN GOVT	0.500	0.858	0.389	0.418	0.247
GOOD NEWS	0.222	0.624	0.169	0.163	0.446
BAD NEWS	0.210	0.714	0.180	0.223	0.032
NEGATIVE GOVT	0.232	0.728	0.177	0.243	0.092

Table 14: Krippendorff's α for Facebook annotations (n=8,642). "All GPT" aggregates GPT-4o and GPT-o3-mini in English and Russian.

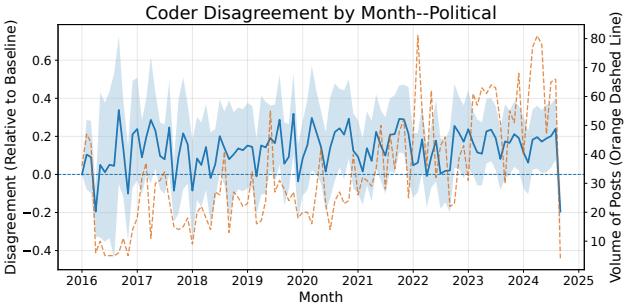

Figure 17: Annotator disagreement estimated by month via fixed-effects regression. Post volume on a monthly basis denoted via orange dashed line.

Figure 18: Mentions of "Ukraine" plus "NATO," "Fascist," "Nazi," or "Casualty" over time in both the VK and Facebook data sets. In total, 126,976 posts mention Ukraine.

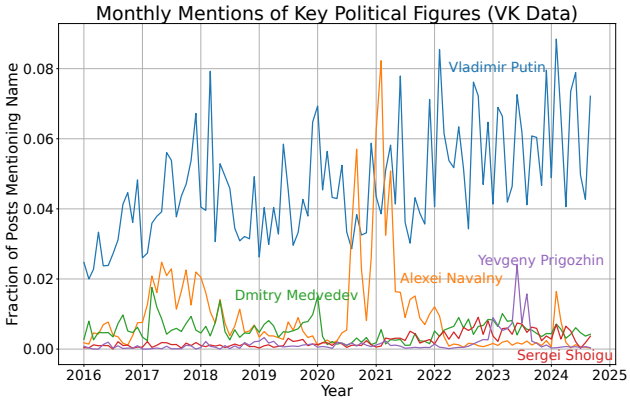

Figure 19: Mentions of key political figures in the VK data over time.

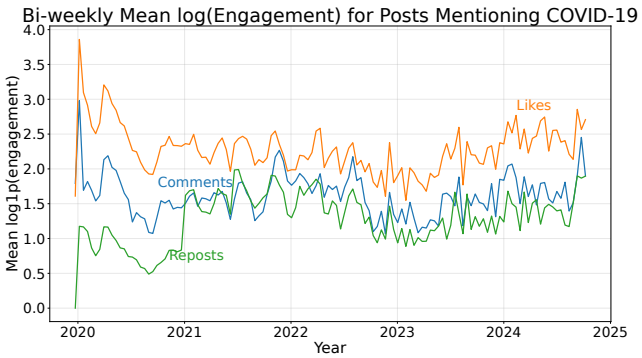

Figure 20: Correlations between mentions of COVID-19 and "pandemic" from keyword searches and logged engagement on VK since 2020. n=58,413.

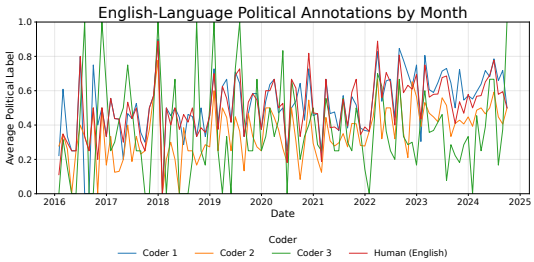

Figure 21: POLITICAL label over time for individual English-language human coders. Observations: Coder 1=1,999, Coder 2=1,900, Coder 3=698, Human (English)=1,900.

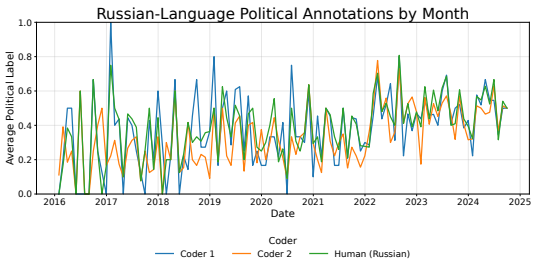

Figure 22: Time series of individual Russian-language human coders on POLITICAL. Obs: Coder 1=1298, Coder 2=1900, Human (Russian)=1850.

Table 15: Events shifting ownership, influence, and/or control for 20 major Russian-language media outlets with brief descriptions and post-event political alignment.

Outlet	Date	Event / Transaction Details	Alignment
AIF.ru	11 Mar 2014	Moscow City Government purchases controlling stake; publisher folded into city media holding	Pro-Kremlin
Gazeta.ru	2012	Alexander Mamut acquires SUP Media (incl. Gazeta.ru) from Alisher Usmanov	Pro-Kremlin
	2013	SUP Media merges with Rambler-Afisha to form Rambler&Co	Pro-Kremlin
	29 Oct 2020	Sberbank buys remaining 45% of Rambler&Co	Pro-Kremlin
Izvestia.ru	3 Jun 2005	Gazprom-Media acquires 50.19% stake from Prof-Media	Pro-Kremlin
	2008	Shares transferred to National Media Group (NMG)	Pro-Kremlin
Kommersant.ru	Sep 2006	Alisher Usmanov buys 100% of Kommersant Publishing House	Pro-Kremlin
Lenta.ru	2013	Bought by Mamut's Afisha–Rambler–SUP holding	Pro-Kremlin
	12 Mar 2014	Editor Galina Timchenko fired after Roskomnadzor warning; staff exodus	Pro-Kremlin
	2020	Rambler stake sold to Sberbank	Pro-Kremlin
Life.ru	2009	Launch of LifeNews/Life.ru by Aram Gabrelyanov's News Media	Pro-Kremlin
	2017	TV channel shuttered; newsroom restructured as digital tabloid	Pro-Kremlin
MK.ru	1991	Editor Pavel Gusev leads buy-out of *Moskovsky Komsomolets*	Pro-Kremlin
Pravda.ru	1999	Vadim Gorshenin registers Pravda.ru, separate from *Pravda*	Pro-Kremlin
Vedomosti.ru	2005	Sanoma sells stake to Demyan Kudryavtsev & partners	Mixed→Pro
	Apr 2015	Kudryavtsev group buys FT & WSJ stakes	Mixed→Pro
	29 May 2020	Yeremin's Sapport/Arbat Press acquires 100% stake	Pro-Kremlin
mediazona.ru	2013	Founded by Maria Alyokhina and Nadezhda Tolokonnikova after their release incarceration for anti-Putin activism	Independent
	March 2022	Government blocked Mediazona in Russia	Independent
meduza.ru	2014	Founded by former *Lenta.ru* employees	Independent
Newizv.ru	24 Oct 1997	Founded by journalists leaving *Izvestia* after Berezovsky takeover	Independent
	20 Feb 2003	Majority purchased by Oleg Mitvol	Independent
NewTimes.ru	2007	Relaunched by Yevgenia Albats; remains critical despite fines	Independent
NG.ru	1993	Editorial board buys *Nezavisimaya Gazeta* from state	Neutral
	1995	Boris Berezovsky acquires controlling interest	Pro-Kremlin (then)
	2005	Sold to Konstantin Remchukov	Neutral/Mixed
Novayagazeta.ru	1 Apr 1993	Founded by ex-*Komsomolskaya Pravda* reporters	Independent
	28 Mar 2022	Suspends print/online after Roskomnadzor warning	Independent
	2023	Courts revoke print license; editions continue outside Russia	Independent
Republic.ru	May 2009	Online magazine *Slon.ru* created by Leonid Bershidsky	Independent
	Oct 2016	Rebrands; Dozhd Media and Alexander Vinokurov take minority stake	Independent
tvrain.ru	2010	Founded by Natalya Sindeyeva and Vera Krichevskaya	Independent
	March 2022	Government blocked access to TV Rain in response to its coverage of the Ukraine invasion	Independent
	December 2022	Began broadcasting from the Netherlands	Independent
Vesti.ru	1 Jul 2006	VGTRK launches Vesti.ru as online arm of state TV news	Pro-Kremlin (state)
RG.ru	1990	Government establishes *Rossiyskaya Gazeta* as official record	Pro-Kremlin (state)
RBC.ru	16 Jun 2017	Grigory Berezkin buys RBC Media	Pro-Kremlin
	Nov 2018	Investigative desk purged; editor resigns after Kremlin pressure	Pro-Kremlin