Telegram as a Battlefield: Kremlin-Related Communications During the Russia-Ukraine Conflict

Apaar Bawa[1], Ugur Kursuncu[1✉], Dilshod Achilov[2], Valerie L. Shalin[3], Nitin Agarwal[4,5], Esra Akbas[1]

[1]Georgia State University, Atlanta, USA
[2]University of Massachusetts Dartmouth, USA
[3]Wright State University, Dayton, USA
[4]COSMOS Research Center, University of Arkansas - Little Rock, USA
[5]International Computer Science Institute, University of California, Berkeley, USA
abawa2@student.gsu.edu, ugur@gsu.edu, dachilov@umassd.edu, valerie.shalin@wright.edu, nxagarwal@ualr.edu, eakbas1@gsu.edu

Abstract

Telegram emerged as a crucial platform for both parties during the conflict between Russia and Ukraine. Per its minimal policies for content moderation, Pro-Kremlin narratives and potential misinformation were spread on Telegram, while anti-Kremlin narratives with related content were also propagated, such as war footage, troop movements, maps of bomb shelters, and air raid warnings. This paper presents a dataset of posts from both pro-Kremlin and anti-Kremlin Telegram channels, collected over a period spanning a year before and a year after the Russian invasion. The dataset comprises 404 pro-Kremlin channels with $4,109,645$ posts and 114 anti-Kremlin channels with $1,117,768$ posts. We provide details on the data collection process, processing methods, and dataset characterization. Lastly, we discuss the potential research opportunities this dataset may enable researchers across various disciplines.

Introduction

Social media has become a primary communication tool for political communication and influence, serving both state-affiliated actors and opposition groups in shaping public discourse and mobilizing support. In the Russian context, the government has leveraged online platforms to promote its policy agenda, manage its international image, and suppress dissenting narratives (Gunitsky 2015; Spaiser et al. 2017). These online efforts build upon the Kremlin's long-standing control over traditional media, including state-owned television channels, such as Rossiya 1 and NTV, which have been instrumental in disseminating official narratives (Aro 2016). Prior research has highlighted how these strategies extend to online communications where disinformation campaigns, surveillance measures, and restrictions on independent journalism have been employed to limit political pluralism and suppress pro-democratic discourse Polyanskaya, Krivov, and Lomko (2003); Aro (2016). This extensive ability to control communications enables the state to exert significant influence over public opinion.

On the other hand, the Russian opposition (i.e., anti-Kremlin) relies on social media platforms with more flexible content moderation policies, such as Telegram (Jurcevic 2019). Despite increased restrictions, digital platforms have amplified the voices of the Kremlin's critics (Glazunova and Amadoru 2023). Smyth and Oates (2017) have concluded that while state media's dominant narrative fosters regime support, alternative debates and government critiques on new media form the cognitive basis for opposition. For instance, intelligence on Russian troops' border movements and military plans was leaked and became trending on social media just before President Vladimir Putin announced the full-scale invasion (Karalis 2024). Because anti-Kremlin narratives can only be spread on such online platforms, it is crucial for researchers to investigate the dynamics and potential strategies of these communications.

Russian Invasion of Ukraine Portrayed in Online Communications

In February 2022, as Russian forces invaded Ukraine, the Russian government swiftly enacted legislation aimed at controlling the narrative around the conflict. This included a "fake news" law that specifically banned the use of the term "war" to describe the invasion, requiring instead the use of the government-approved phrase "special military operation" (Sherstoboeva 2024). Violations of this law led to severe punishment, including imprisonment for up to ten years, the shutdown of numerous news outlets, and the criminal prosecution of thousands of individuals for alleged misinformation. Pro-Kremlin channels often emphasized the legitimacy of the "special military operation," focusing on liberation and national security. Persisting anti-Kremlin channels highlighted war atrocities, humanitarian crises, and resistance efforts (Ramani 2023). The role of social media in this conflict is not merely passive; it actively attempts to shape perceptions and mobilize support or dissent. The rapid spread of information, from verified reports to misinformation, underscores the importance of digital literacy and critical analysis in contemporary conflicts (Pierri et al. 2023).

Telegram as an Essential Propaganda Medium

Major online platforms such as Facebook and X (formerly Twitter) have become critical sources of information for news, activism, business, and marketing in the last two decades (Kalsnes and Larsson 2018; Kursuncu et al. 2021). Similarly, Telegram, an instant messaging app, has gained substantial popularity, especially in Russia and Eastern Europe, where it serves as a crucial alternative to state-monitored communication channels for sharing news (Khaund et al. 2020). With Russia's invasion of Ukraine, Facebook and Instagram were banned over extremism charges by a Moscow court, and access to Twitter was restricted by the Russian censorship body Roskomnadzor (Glazunova and Amadoru 2023). In that light, Telegram's role as a vital source of information has become increasingly significant. According to a survey by the Levada Levada (2023), among Russian youth aged $18-24$, TV was found to be the least popular news source, with 41% relying on Telegram channels for news. Additionally, during the initial two months of the conflict, 76.6% of Ukrainians used social media for news, with Telegram being the most favored platform by 65.7% (OPORA 2022). Both Russian and Ukrainian government officials also actively use Telegram to rally international support, broadcast air raid warnings, and distribute maps of local bomb shelters.

Telegram distinguishes itself from other major online platforms through several unique features: (i) Telegram channels can be public or private, serving as feeds where text, photos, videos, audio, documents, and polls are broadcast to large audiences, including unlimited subscribers. (ii) Unlike other platforms, Telegram does not use algorithms to prioritize or restrict content visibility, giving users more straightforward access to all posted content. (iii) Users can easily disable comments and reactions, making channels a one-directional communication tool for mass information dissemination. These features have positioned Telegram as an effective platform for both anti- and pro-Kremlin propaganda within Russia, leading to its wide adoption for news consumption and dissemination. Over 40% of all Telegram channels are dedicated to news, with approximately 17% directly related to the war and ongoing events in Ukraine (Re:Russia 2024). This also makes Telegram a central medium for real-time updates and reliable information, influencing public opinion and information flow.

The Internet Research Agency (IRA), a Russian organization often associated with Kremlin-linked information operations, has been identified as playing a central role in coordinated online influence campaigns by disseminating polarizing narratives on Western online platforms (Alieva, Kloo, and Carley 2024; Linvill and Warren 2020; Starbird, Arif, and Wilson 2019). While much of the IRA's documented activity has focused on platforms such as X/Twitter and Facebook, growing evidence indicates its activity on Telegram as well (UK Government 2022). In February 2023, a Telegram channel affiliated with the Wagner Group published a statement from Yevgeny Prigozhin acknowledging his connection to the IRA (Krever and Chernova 2023). This development adds to concerns regarding Telegram's use as an instrument for coordinated influence operations, besides being an online space for independent communication.

Related Work

A growing body of research has examined online communications related to the Russian-Ukraine conflict across online platforms including X/Twitter, Reddit and Telegram. Much of this literature focuses on the spread of propaganda, disinformation, and strategic narratives by various state and non-state actors.

Hanley, Kumar, and Durumeric (2023a) conducted a comparative analysis of Western, Russian, and Chinese media ecosystems by examining 11,359 news articles and associated social media content published between January and April 15, 2022. Their work highlights distinct thematic patterns in how the conflict was framed in traditional and social media across these regions. In another study (Hanley, Kumar, and Durumeric 2023b), the same authors explored the dissemination of Russian state-sponsored narratives to English-speaking audiences by analyzing content from nine English-language Russian media websites and related Reddit communities, including r/Russia and 10 other political subreddits, to understand the most prominent narratives touted by the Kremlin for English-speaking audiences. In a subsequent work, Hanley and Durumeric (2024) curated 732 Telegram channels hyperlinked by 16 Russian media websites tracked by the U.S. State Department, providing insight into how narratives circulate between formal news sites and Telegram. However, this dataset largely comprises pro-Kremlin sources and is limited to a specific timeframe.

Beyond Telegram, multilingual social media analyses have been conducted. For instance, Lai, Toriumi, and Yoshida (2024) curated a X/Twitter dataset containing approximately 53 million tweets in six languages (English, Japanese, Spanish, French, German, Korean) during the early weeks of the invasion, identifying narratives and misinformation trends across linguistic communities.

Other researchers have focused on building datasets from Telegram to support broader interest areas. The Pushshift Telegram Dataset Baumgartner et al. (2020) is one of the largest publicly available collections, encompassing 317 million messages from 27,801 channels and over 2.2 million users. Though it predates the current conflict, its infrastructure informed subsequent data collection efforts. Gruzd et al. (2024), funded by Canada's Social Sciences and Humanities Research Council (SSHRC), compiled data from 55 English-language Telegram channels covering the conflict between February 2022 and February 2023, although details of their sampling strategy remain limited.

Telegram continues to be leveraged in research on specific narratives and movements. Herasimenka (2022) used Telegram data to study leadership strategies in anti-authoritarian protest movements, focusing on Alexei Navalny's 2017 mobilization in Russia. Kloo, Cruickshank, and Carley (2024) assembled a dataset of 6.7 million Telegram messages from over 18,000 channels to analyze disinformation narratives, particularly those portraying Ukraine as a "Nazi state." Ghasiya and Sasahara (2023) examined communication patterns on Telegram during the conflict by focusing on selected high-profile channels from both Ukrainian and Russian sides, such as UkraineNow and RT Russian.

The Current Study

This paper presents the first comprehensive Telegram dataset capturing both pro-Kremlin and anti-Kremlin Telegram channels within the context of the Russian political landscape. Our dataset is created from 404 pro-Kremlin and 114 anti-Kremlin Telegram channels, with over four million and one million posts, respectively. These channels have been instrumental in disseminating various narratives and counter-narratives within the ongoing Ukraine conflict (for Strategic Dialogue 2022; Bawa et al. 2024). A portion of this dataset, covering anti-Kremlin channels from January 2022 to March 2023, has been analyzed to examine the dynamics between offline events and online communications in anti-Kremlin channels over the seven distinct phases of the invasion (Bawa et al. 2024), as outlined by Murauskaite (2023). In the processes of data collection and curation, we ensure transparency and reproducibility in our approach. We provide details on data characteristics, including post volumes, views, forwarding and user engagement levels, and the use of multimodal content (text, images, videos). We also present n-gram analysis to provide insights into the lexical and topic characteristics prevalent in data, enabling parallel comparisons between the channels (Kursuncu et al. 2019). This unique dataset serves as an essential resource for understanding the dynamics of online political discourse in Russia, particularly in how digital platforms are used to influence public opinion and engage with the audience during periods of escalated political concern.

Methods

This section outlines our methodology for collecting and analyzing data from Russian political Telegram channels during the Russia-Ukraine conflict. We describe the process of selecting and categorizing channels, detail the data collection techniques, and explain how channel labels were validated. In addition, we present an overview of the dataset, including the types of attributes collected, covering the period before and after the invasion. Finally, we provide a content analysis highlighting the differences between Pro-Kremlin and Anti-Kremlin channels.

Data Collection and Annotation

We collected Telegram data using the open-source Python library Telethon (Telethon 2024), which provides client access to Telegram's public API. This API supports the retrieval of publicly available content such as messages, metadata, and user/channel information from public channels and groups, while excluding private or deleted content. The library adheres Telegram's rate limits, including 200 username per day. Our data collection was conducted retrospectively in batches, starting in May 2023 and continuing until June 2023.

To identify relevant Russian-language political Telegram channels, we used the TGStat platform (TGStat 2024), which catalogs channels by country, language, and thematic

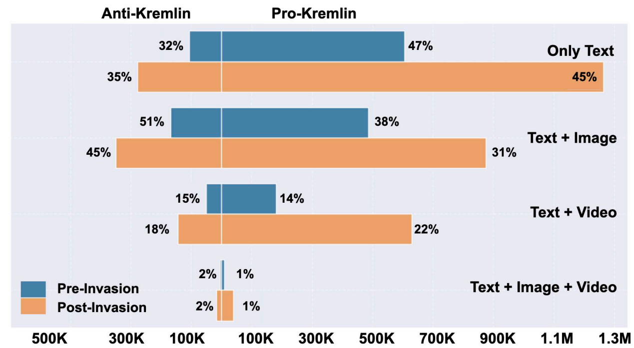

Figure 1: Comparison of Anti-Kremlin and Pro-Kremlin post volumes across different modalities (Only Text, Text Image, Text Video, and Text Image Video) during the Pre-Invasion and Post-Invasion phases.

content. We selected channels with at least $10,000$ subscribers to focus on influential accounts likely to contribute to public discourse. These channels were manually labeled by a native Russian-speaking coder into four categories: Pro-Kremlin, Anti-Kremlin, Neutral, and Others. Pro-Kremlin channels were identified based on their consistent dissemination and frequent alignment with official Kremlin narratives and support for state policies. In contrast, anti-Kremlin channels were those exhibiting critical stance toward the policies and actions of Russian government actions and leadership. Channels deemed "neutral" predominantly shared news without discernible political bias. The "Others" category included topics unrelated to the focus of our study, such as other countries' politics.

Label validation was conducted in two stages. First, a non-Russian-speaking coder reviewed the classifications using Telegram's built-in translation tools. Second, a native Russian-speaking political science expert reviewed a randomly selected subset of 100 channels to assess the label accuracy and reliability. To evaluate the inter-rater agreement, we calculated Cohen's kappa coefficient (Cohen 1960), κ. While κ provides a measure of consistency across coders, it does not confirm the accuracy of labels and should be interpreted accordingly.

Data Description

Following the annotation process, we identified 404 pro-Kremlin and 114 anti-Kremlin channels. Data collection from these channels spanned 430 days from December 21, 2020, to April 30, 2023—roughly more than a year pre-invasion and post-invasion. The dataset includes the following attributes of Telegram posts: (i) main text content of the posts, (ii) accompanying media (e.g., images, videos –represented as placeholders in the dataset), (iii) timestamps (e.g., date and time), (iv) number of view counts per post, (v) post forwarding counts (i.e., how often the post is forwarded), (vi) original or forwarded status (i.e., whether a post is original or forwarded), (vii) forwarding source (i.e., the origin source of the forwarded post, if applicable), (viii) emoji reactions (i.e., type of emoji reactions and their frequencies),

Modalities	Pro-Kremlin			Anti-Kremlin		
	Pre-Invasion	Post-Invasion	% Change	Pre-Invasion	Post-Invasion	% Change
Only Text	18,696	62,853	236.2%	25,006	101,597	306.3%
Text + Image	12,106	59,333	390.1%	25,218	103,115	308.9%
Text + Video	16,957	86,447	409.8%	26,355	112,429	326.6%
Text + Image + Video	20,529	97,394	374.4%	26,424	127,480	382.4%

Table 1: Distribution of the average number of views for different modalities, Pre- and Post-Invasion for both Pro-Kremlin and Anti-Kremlin channels.

and (ix) user replies to the post (i.e., each reply contains similar data as above). This comprehensive dataset contains $4,109,645$ posts from pro-Kremlin channels and $1,117,768$ posts from anti-Kremlin channels. The inclusion of such diverse data attributes allows for an in-depth analysis of communication patterns and user engagement on these channels.

Data Records

We performed a comparative exploratory analysis of both Pro-Kremlin and Anti-Kremlin channels in the context of pre- and post-invasion time frames. We examined the following attributes of the posts: (i) Multimodalities and Views (i.e., presence of various modalities of data and view counts), (ii) Post Forwarding, (iii) Reactions, and (iv) Replies.

Multimodality and Views

Research shows that multimodal content, a critical feature that Telegram users often utilize (Times 2023), boosts user engagement (Park et al. 2022; Voorveld et al. 2018). Both Pro-Kremlin channels and anti-Kremlin channels demonstrate notable increases in post volumes, from $1,289,173$ pre-invasion to $2,820,472$ post-invasion by 119%, and from $330,492$ to $787,276$ by 138%, respectively. Figure 1 presents a comparison across pre- and post-invasion post volumes between Pro- and Anti-Kremlin channels, highlighting changes in their multimodal characteristics. The content is various combinations of modalities as follows; (i) text and image, (ii) text and video, (iii) text, image and video, (iv) only text. The relative proportions of each modality type for a given group and period reveal cues on evolving communication strategies. For instance, text-only posts in Pro-Kremlin channels decreased slightly from 47% to 45% of total posts despite an absolute increase in count, while posts combining text and video increased significantly. In contrast, Anti-Kremlin channels experienced a modest 3% rise in both text-only and text-video posts, and a 6% decline in text-image posts, although the latter remained the most frequent modality overall.

After the invasion, average views per post for Pro-Kremlin channels rose from $15,929$ to $67,442$ (a 323% increase) and for Anti-Kremlin channels from $25,346$ to $104,778$ (313% increase). Table 1 presents the distribution of views across the aforementioned multimodal content for both Pro- and Anti-Kremlin channels, highlighting trends before and after the invasion. Notably, anti-Kremlin channels consistently garnered higher viewership than pro-Kremlin channels, especially post-invasion. Further, posts with videos have attracted more views in both channel categories, indicating a strong viewer preference for video content. These trends highlight the dynamic changes in content presentation to increase user engagement, emphasizing the importance of multimodal content.

Post Forwarding

Post forwarding on Telegram is a key contributor to user engagement (Ng et al. 2024). Pro-Kremlin channels saw an increase of 313.5% in the average number of post forwarding, rising from 37 pre-invasion to 153 post-invasion. Anti-Kremlin channels experienced a 252.4% increase in forwarding, with the average rising from 82 to 289 per post. Multimodal posts, which include text, images, and videos, showed more notable increases, from 171 to 600 in Pro-Kremlin channels by 250.9%, and from 311 to 1170 in Anti-Kremlin channels by 276.2%. Posts containing videos saw a 231.4% increase, from 140 to 464 in Pro-Kremlin channels, and a 283.5% rise, from 231 to 886, in Anti-Kremlin channels. Posts in anti-Kremlin channels were forwarded more frequently on average than those in pro-Kremlin channels, highlighting higher user engagement in channels opposing the Kremlin.

	Pro-Kremlin		Anti-Kremlin	
	Pre- **(∼8.4M)**	**Post-** **(∼2B)**	**Pre-** **(∼5M)**	**Post-** **(∼1.2B)**
👍	42.39%	51.83%	39.73%	39.12%
🤬	1.6%	6.55%	3.4%	11.14%
❤️	5.05%	7.33%	7.48%	10.11%
😁	15.14%	8.3%	9.77%	9.0%
😢	3.78%	3.65%	3.41%	7.09%
👎	3.85%	4.2%	10.4%	4.52%
🔥	8.86%	5.91%	5.36%	4.1%
💩	7.88%	0.95%	9.58%	2.77%
others	11.45%	11.28%	10.87%	12.15%

Table 2: *Pro-Kremlin* and *Anti-Kremlin* channels most prevalent user reactions in the content for *Pre-* and *Post-invasion*.

N-grams	Before Invasion	After Invasion
Unigrams	Russia, Ukraine, USA, Putin, Moscow, case, military, ruble, elections, political, war, coronavirus, governor, vaccination, court	Russia, Ukraine, military, USA, Ukrainian Armed Forces, war, Putin, Moscow, Kyiv, DNR - Donetsk People's Republic, shelling, destroy, Zelensky, sanction, opponent
Bigrams	United Russia, foreign agent, criminal case, Nord Stream, Rostov region, Alexei Navalny, DNR LNR, State Duma elections, Vladimir Zelensky, Sergey Sobyanin, law enforcement agency, price increase, Mikhail Mishustin, Minsk agreement, Alexander Lukashenko	Vladimir Putin, combat action, military operation, Kherson region, Kharkov region, White House, Wagner PMC, Ukrainian nationalist, result of shelling, Zelensky, liberate territory, DNR LNR, criminal case, nuclear weapon, Russian oil
Trigrams	performing the function of a foreign agent, mass information, initiate criminal case, Great Patriotic War, China Russia, armed forces of Ukraine, Anthony Blinken, movement forty forty, Russia NATO, detecting new cases of coronavirus	special military operation, war in Ukraine, territory of Russia, region launching caliber shells, performing the function of a foreign agent, citizen of Ukraine, reactive system of volley fire, horde of native evil, Great Patriotic War, initiating a criminal case

Table 3: Pro-Kremlin channels most prevalent n-grams in the content of before and after invasion.

Reactions and Replies

Telegram channels can opt to enable or disable user reactions and replies, which impacts public engagement. During our data collection process, distinguishing between posts without reactions due to low engagement and those where reactions were disabled posed a challenge. However, data shows a notable increase in the average number of reactions per post: from 7 pre-invasion to 991 post-invasion in Pro-Kremlin channels; from 15 to $1,533$ in anti-Kremlin channels. Table 2 compares the top 10 emoji used by users pre- and post-invasion for both Pro-Kremlin and Anti-Kremlin channels. In Pro-Kremlin channels, the "👍" emoji dominated over 50% of reactions, while a more diverse range of emoji use was observed in anti-Kremlin channels, including emoji representing strong disagreement, such as "🤬" (Kursuncu et al. 2019; Wijeratne et al. 2016), with larger volume in Anti-Kremlin channels, post-invasion. The sad emoji "😢" decreased slightly in Pro-Kremlin channels from 3.78% to 3.65%, while an increase was observed in Anti-Kremlin channels from 3.41% to 7.09%, indicating increasing emotional response post-invasion.

Before the invasion, both Pro-Kremlin and Anti-Kremlin channels averaged two replies per post, which rose to nine replies per post, indicating higher user engagement through replies post-invasion. In addition, we acknowledge the potential involvement of automated bots in replies or reactions to boost a post's popularity artificially.

Content Analysis

Our dataset shows specific qualitative differences between Pro-Kremlin and Anti-Kremlin channels, especially pre- and post-invasion. Anti-Kremlin channels typically focus on war atrocities, Russian troop movements, and humanitarian issues. In contrast, Pro-Kremlin channels often try to justify the *"special military operation"* (Hanley, Kumar, and Durumeric 2023a). To analyze the basic thematic characteristics of the data, we conducted an n-gram analysis of the text content from these channels both before and after the invasion. We extracted n-grams ($n = 1, 2, 3$) and performed text preprocessing, such as removing punctuation, hyperlinks, numbers, special characters, and Russian stopwords. Additionally, given the extensive use of inflections in Russian, we used lemmatization to standardize different forms of the same word, employing the MyStem package, a Python wrapper for the morphological analysis of the Russian language.

Our analysis of n-grams from both Pro-Kremlin (Table 3) and Anti-Kremlin (Table 4) channels reveals both common and unique phrases potentially distinct narratives. Common phrases include countries, such as *"Russia"*, *"Ukraine"*, *"USA"*; regional and military terms, such as *"Kherson Region"*, *"Kharkov region"*, *"Military"*, *"shelling"*; and leader names such as *"Vladimir Putin"*, *"Vladimir Zelensky"*, *"Joe Biden"*. On the other hand, unique to anti-Kremlin channels are phrases including *"Foreign Agent"*, *"War Crime"*, *"air raid alert"*, *"please keep an eye"*, *"Help"*, which potentially reflect critical perspectives on the conflict and requests for aid. In contrast, pro-Kremlin communications often use phrases unique to these channels, including *"liberate territory"* and *"special military operation"*, indicating a justification of military actions. Further, the term *"special military operation"* was used more frequently than *"war"*, and the phrase *"Great Patriotic War"* also appeared in Pro-Kremlin tri-grams, suggesting a historical framing of current events. This data potentially contains cues for contrasts in how each side portrays the conflict and mobilizes support through online content creation and dissemination.

N-grams	Before Invasion	After Invasion
Unigrams	Russia, Ukraine, Putin, Navalny, court, Moscow, USA, ruble, detain, police, Belarus, elections, law, criminal, coronavirus	Russia, Ukraine, war, military, region, USA, court, strike, missile, sanction, ruble, Kiev, resident, Vladimir, occupant, help, to perish
Bigrams	Foreign Agent, Alexey Navalny, Vladimir Putin, Criminal Case, State Duma Deputy, Russian Authority, Initiate Criminal, Vladimir Zelensky, White House, Supreme Court, State Duma Elections, House Arrest, DNR LNR, Alexander Lukashenko, Recognize Guilty, Ramzan Kadyrov	Foreign Agent, Ukraine War, Vladimir Putin, Vladimir Zelensky, Criminal Case, Missile Strike, Kherson Region, Kharkiv Region, War Crime, Joe Biden, Air Raid Alert, Political Circus, Nuclear Weapon, Russian Oil, Military Aid
Trigrams	foreign mass media, initiate a criminal case, Russia Ukraine, Criminal Code of the Russian Federation, cases of COVID actively in the region, United Russia Party, European Court of Human Rights, support Navalny, Anthony Blinken, during the last day in Ukraine, Great Patriotic War	Alexey Vladimirovich concerning activities, Russian invasion of Ukraine, special military operation, Donetsk region siren, civil defense shelter, attention air raid sirens, reactive system of volley fire, military aid to Ukraine, Donetsk region all-clear, occupied territory of Ukraine

Table 4: Anti-Kremlin channels most prevalent n-grams in the content of before and after invasion.

Data Quality and Qualitative Differences

The annotation comparison between a native Russian-speaking coder and a non-Russian-speaking coder yielded Cohen's kappa κ score of 0.97, suggesting strong inter-rater agreement. Further, the assessment between the expert and the Russian-speaking and non-Russian-speaking coders revealed kappa κ scores of 0.87 and 0.89, respectively, indicating substantial agreement with the expert's annotations. These results confirm substantial agreement with the expert's annotations, validating the reliability of our channel classification process.

Our dataset adheres to the *FAIR Data Principles*[1]. Each data point has a unique identifier accompanied by detailed metadata to enhance findability. The data is stored in a secure repository, which is openly accessible with clearly stated access guidelines. Standardized formats ensure that our dataset is interoperable. Comprehensive documentation provides detailed insights into the data's collection, processing, and context, supporting its reusability.

Usage Notes

The dataset can be accessed via the following Figshare repository[2]. The repository has files in CSV format. This dataset was constructed using public Telegram channels and does not involve any direct interaction with individuals or collection of personally identifiable private data. We have adhered to standard anonymization practices during the data collection and processing stages, ensuring that any personally identifiable information (PII), including user IDs, has been removed. This approach aligns with ethical guidelines to protect the privacy of individuals whose data may be part of the public channels used. It is important to note that this dataset does not include actual images and videos; instead, it contains placeholders indicating the presence of such media. The primary reason for this limitation is the practical challenge of collecting large video files, as Telegram allows users to send video files up to 2 GB in size each. Consequently, this dataset focuses on textual data and metadata, which may limit certain types of multimedia analysis but facilitates more accessible and scalable data handling and processing.

Discussion & Research Opportunities

This dataset provides a unique vantage point for analyzing the dynamics of online discourse during the Ukraine-Russia conflict, opening various avenues for answering research questions. Its temporal depth and breadth, which spans over a year before and after the invasion, presents researchers a foundation to explore longitudinal changes in online discourse. By differentiating between pro-Kremlin and anti-Kremlin Telegram channels, the dataset enables comparative analyses of divergent propaganda and counter-narrative strategies, fostering questions into how these narratives shape public perception, political behavior, and transnational information flows.

One major opportunity could be a systematic analysis of narrative formation and transformation in response to real-world geopolitical events. This dataset, enriched with detailed metadata (e.g., timestamps, forwarding activity, emoji reactions), allows researchers to examine how information propagates and evolves across Telegram communities. Contextual content analysis coupling message-level data with known political developments, can reveal how rhetorical strategies respond to, reinforce, or challenge dominant media frames and state-sponsored messaging. This can help scholars investigate phenomena such as agenda-setting, frame alignment, and narrative breach (Colleoni, Rozza, and Arvidsson 2014; Entman 2008; Bawa et al. 2024).

Further, the dataset can support the application of ad-

[1]`https://force11.org/info/the-fair-data-principles/`

[2]`https://doi.org/10.6084/m9.figshare.28785449`

vanced natural language processing techniques, including topic modeling, stance detection, and sentiment analysis, to identify emergent themes and shifts in thematic patterns (AlDayel and Magdy 2021; Kursuncu et al. 2019; Kim et al. 2024; Shu and Liu 2022). These tools can help uncover the subtle linguistic and semantic patterns embedded in propaganda or dissenting discourse (Rashkin et al. 2017). When supplemented with entity recognition and relationship extraction, the resulting insights can be integrated with knowledge graphs that structurally map the key actors, themes, and their interrelations across time (Hogan et al. 2021). These knowledge-driven approaches can deepen our understanding of coordinated information operations and opposition narratives by embedding them in a semantically enriched representations (Padhi et al. 2024; Garg et al. 2024).

The potential applications of this dataset extend beyond academic research. Journalists, civil society organizations, and policy analysts may leverage the data to track disinformation campaigns, monitor grassroots mobilization, or evaluate the effectiveness of online censorship. Additionally, by studying the temporal interplay between online discourse and offline events, such as military escalations, international diplomacy, or protests—researchers, can better understand how online platforms serve as mirrors or amplifiers of real-world tensions.

Lastly, given Telegram's unique role as a semi-private yet mass-broadcast platform with limited algorithmic moderation, this dataset facilitates examination of how political actors and communities navigate hybrid media ecologies. It invites further inquiry into the affordances and constraints of encrypted platforms in authoritarian and semi-authoritarian settings (Gorwa and Guilbeault 2020), and raises important ethical considerations around the governance and transparency of digital public spheres.

Acknowledgments

Author Nitin Agarwal acknowledges the support from the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Office of Scientific Research (FA9550-22-1-0332), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program.

References

AlDayel, A.; and Magdy, W. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management*, 58(4): 102597.

Alieva, I.; Kloo, I.; and Carley, K. M. 2024. Analyzing Russia's propaganda tactics on Twitter using mixed methods network analysis and natural language processing: a case study of the 2022 invasion of Ukraine. *EPJ Data Science*, 13(1): 42.

Aro, J. 2016. The cyberspace war: propaganda and trolling as warfare tools. *European View*, 15: 121–132.

Baumgartner, J.; Zannettou, S.; Squire, M.; and Blackburn, J. 2020. The pushshift telegram dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, 840–847.

Bawa, A.; Kursuncu, U.; Achilov, D.; and Shalin, V. L. 2024. the adaptive strategies of anti-kremlin digital dissent in telegram during the Russian invasion of Ukraine. *arXiv preprint arXiv:2408.07135*.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20: 37–46.

Colleoni, E.; Rozza, A.; and Arvidsson, A. 2014. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of communication*, 64(2): 317–332.

Entman, R. M. 2008. Theorizing mediated public diplomacy: The US case. *The International Journal of Press/Politics*, 13(2): 87–102.

for Strategic Dialogue, I. 2022. A false picture for many audiences: How Russian-language pro-Kremlin Telegram channels spread propaganda and disinformation about refugees from Ukraine. https://www.isdglobal.org/digital_dispatches/a-false-picture-for-many-audiences-how-russian-language-pro-kremlin-telegram-channels-spread-propaganda-and-disinformation-about-refugees-from-ukraine/. Accessed: 2024-12-28.

FORCE11. 2023. The FAIR Data Principles. Accessed: 2025-01-14.

Garg, R.; Padhi, T.; Jain, H.; Kursuncu, U.; and Kumaraguru, P. 2024. Just KIDDIN: Knowledge Infusion and Distillation for Detection of INdecent Memes. *arXiv preprint arXiv:2411.12174*.

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.

Ghasiya, P.; and Sasahara, K. 2023. Messaging Strategies of Ukraine and Russia on Telegram during the 2022 Russian invasion of Ukraine. *First Monday*.

Glazunova, S.; and Amadoru, M. 2023. "Anti-Regime Influentials" Across Platforms: A Case Study of the Free Navalny Protests in Russia. *Media and Communication*, 11: 187–202.

Gorwa, R.; and Guilbeault, D. 2020. Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12(2): 225–248.

Gruzd, A.; Li, Y.; Mai, P.; Ribeiro, T.; and Lab, S. M. 2024. [Dataset] Shaping the Narratives of the Russia-Ukraine War for Western Audiences: An Exploration of English-language Telegram Channels. Available via Figshare.

Gunitsky, S. 2015. Corrupting the cyber-commons: Social media as a tool of autocratic stability. *Perspectives on Politics*, 13: 42–54.

Hanley, H. W.; and Durumeric, Z. 2024. Partial mobilization: Tracking multilingual information flows amongst russian media outlets and telegram. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 528–541.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2023a. " A Special Operation": A Quantitative Approach to Dissecting and Comparing Different Media Ecosystems' Coverage of the Russo-Ukrainian War. In *Proceedings of the International AAAI Conference on Web and social media*, volume 17, 339–350.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2023b. Happenstance: utilizing semantic search to track Russian state media narratives about the Russo-Ukrainian war on Reddit. In *Proceedings of the international AAAI conference on web and social media*, volume 17, 327–338.

Herasimenka, A. 2022. Movement leadership and messaging platforms in preemptive repressive settings: Telegram and the Navalny Movement in Russia. *Social Media+ Society*, 8(3): 20563051221123038.

Hogan, A.; Blomqvist, E.; Cochez, M.; d'Amato, C.; Melo, G. D.; Gutierrez, C.; Kirrane, S.; Gayo, J. E. L.; Navigli, R.; Neumaier, S.; et al. 2021. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4): 1–37.

Jurcevic, K. 2019. Social media-The only voice for oppositional media in Russia?

Kalsnes, B.; and Larsson, A. O. 2018. Understanding news sharing across social media: Detailing distribution on Facebook and Twitter. *Journalism Studies*, 19: 1669–1688.

Karalis, M. 2024. Russia-Ukraine through the eyes of social media. https://gjia.georgetown.edu/2024/02/02/russia-ukraine-through-the-eyes-of-social-media/. Accessed: 2024-12-28.

Khaund, T.; Hussain, M. N.; Shaik, M.; and Agarwal, N. 2020. Telegram: Data collection, opportunities and challenges. In *Proceedings of the Annual International Conference on Information Management and Big Data*, 513–526. Springer International Publishing.

Kim, N.; Mosallanezhad, D.; Cheng, L.; Mancenido, M. V.; and Liu, H. 2024. Robust Stance Detection: Understanding Public Perceptions in Social Media. In *International Conference on Advances in Social Networks Analysis and Mining*, 21–37. Springer.

Kloo, I.; Cruickshank, I. J.; and Carley, K. M. 2024. A Cross-Platform Topic Analysis of the Nazi Narrative on Twitter and Telegram During the 2022 Russian Invasion of Ukraine. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 839–850.

Krever, M.; and Chernova, A. 2023. Wagner chief admits to founding Russian troll farm sanctioned for meddling in US elections. https://www.cnn.com/2023/02/14/europe/russia-yevgeny-prigozhin-internet-research-agency-intl/index.html. Accessed: 2025-04-13.

Kursuncu, U.; Gaur, M.; Lokala, U.; Thirunarayan, K.; Sheth, A.; and Arpinar, I. B. 2019. Predictive analysis on Twitter: Techniques and applications. In *Emerging research challenges and opportunities in computational social network analysis and mining*, 67–104. Springer Nature.

Kursuncu, U.; Purohit, H.; Agarwal, N.; and Sheth, A. 2021. When the bad is good and the good is bad: Understanding cyber social health through misc behavioral change. *IEEE Internet Computing*, 25(1): 6–11.

Lai, C.; Toriumi, F.; and Yoshida, M. 2024. A multilingual analysis of pro Russian misinformation on Twitter during the Russian invasion of Ukraine. *Scientific Reports*, 14(1): 10155.

Levada, C. 2023. Main sources of information and popular journalists. https://www.levada.ru/2023/04/20/osnovnye-istochniki-informatsii-i-populyarnye-zhurnalisty/. Accessed: 2024-12-28.

Linvill, D. L.; and Warren, P. L. 2020. Troll factories: Manufacturing specialized disinformation on Twitter. *Political Communication*, 37(4): 447–467.

Murauskaite, E. 2023. *U.S. Arms Transfers to Ukraine: Impact Assessment*. College Park, MD: START.

Ng, L. H. X.; Kloo, I.; Clark, S.; and Carley, K. M. 2024. An exploratory analysis of COVID bot vs human disinformation dissemination stemming from the Disinformation Dozen on Telegram. *Journal of Computational Social Science*, 1–26.

OPORA, S. 2022. Media Consumption of Ukrainians in a Full-Scale War. https://oporaua.org/report/polit_ad/24068-mediaspozhivannia-ukrayintsiv-v-umovakh-povnomasshtabnoyi-viini-opituvannia-opori. Accessed: 2024-12-28.

Padhi, T.; Kursuncu, U.; Kumar, Y.; Shalin, V. L.; and Fronczek, L. P. 2024. Enhancing Cross-Modal Contextual Congruence for Crowdfunding Success using Knowledge-infused Learning. In *2024 IEEE International Conference on Big Data (BigData)*, 1844–1853. IEEE.

Park, C. Y.; Mendelsohn, J.; Field, A.; and Tsvetkov, Y. 2022. Challenges and opportunities in information manipulation detection: An examination of wartime Russian media. In *Findings of the Association for Computational Linguistics (EMNLP)*, 5209–5235.

Pierri, F.; Luceri, L.; Jindal, N.; and Ferrara, E. 2023. Propaganda and misinformation on Facebook and Twitter during the Russian invasion of Ukraine. In *Proceedings of the 15th ACM web science conference 2023*, 65–74.

Polyanskaya, A.; Krivov, A.; and Lomko, I. 2003. Virtualnoye Oko Stareshevo Brata [The Virtual Eye of Big Brother]. *Vestnik misc*, 9: 320.

Ramani, S. 2023. *Putin's War on Ukraine: Russia's Campaign for Global Counter-Revolution*. Hurst Publishers.

Rashkin, H.; Choi, E.; Jang, J. Y.; Volkova, S.; and Choi, Y. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, 2931–2937.

Re:Russia. 2024. What does the Kremlin fear? An analysis of protest potential in Russian society. https://re-russia.net/en/review/263/. Accessed: 2024-12-28.

Sherstoboeva, E. 2024. Russian Bans on 'Fake News' about the war in Ukraine: Conditional truth and unconditional loyalty. *International Communication Gazette*, 86: 36–54.

Shu, K.; and Liu, H. 2022. *Detecting fake news on social media*. Springer Nature.

Smyth, R.; and Oates, S. 2017. Mind the gaps: Media use and mass action in Russia. In *State Against Civil Society*, chapter 4, 127–147. Routledge.

Spaiser, V.; Chadefaux, T.; Donnay, K.; Russmann, F.; and Helbing, D. 2017. Communication power struggles on social media: A case study of the 2011–12 Russian protests. *Journal of Information Technology and Politics*, 14: 132–153.

Starbird, K.; Arif, A.; and Wilson, T. 2019. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on human-computer interaction*, 3(CSCW): 1–26.

Telethon. 2024. Telethon's Documentation. https://docs.telethon.dev/en/stable/index.html. Accessed: 2024-12-28.

TGStat. 2024. Telegram channels and groups catalog. https://tgstat.ru/en/politics. Accessed: 2024-12-28.

Times, T. M. 2023. Telegram surpasses WhatsApp traffic volume in Russia. https://www.themoscowtimes.com/2023/01/23/telegram-surpasses-whatsapp-traffic-volume-in-russia-a80012. Accessed: 2024-12-28.

UK Government. 2022. UK exposes sick Russian troll factory plaguing social media with Kremlin propaganda. Accessed: 2025-04-13.

Voorveld, H. A. M.; Van Noort, G.; Muntinga, D. G.; and Bronner, F. 2018. Engagement with social media and social media advertising: The differentiating role of platform type. *Journal of Advertising*, 47: 38–54.

Wijeratne, S.; Balasuriya, L.; Sheth, A.; and Doran, D. 2016. Emojinet: Building a machine readable sense inventory for emoji. In *Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part I 8*, 527–541. Springer.

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 dataset consists of publicly available Telegram posts from open channels, and all personal identifiers have been removed to ensure privacy and ethical compliance.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, abstract and introduction clearly outline the dataset's scope, including its focus on pro-Kremlin and anti-Kremlin Telegram channels, the data collection methods, and the potential research opportunities enabled by the dataset. These claims are consistent with the paper's contributions and findings.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, the paper explains the methodology in detail, including channel selection, data collection using the Telethon package, and annotation processes validated by experts.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, possible artifacts, such as the potential impact of automated bots on user engagement metrics like replies and reactions, are addressed.
   (e) Did you describe the limitations of your work? Yes, the paper describes limitations, including the challenges of identifying automated bot activity (see Subsection Reactions and Replies), the absence of actual multimedia content in the dataset (see Section Usage Notes).
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? Yes
   (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
   (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)? NA

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

(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)? NA

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

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

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? NA

(b) Did you mention the license of the assets? Yes, see https://doi.org/10.6084/m9.figshare.28208729.v1

(c) Did you include any new assets in the supplemental material or as a URL? Yes. The DOI and its URL are included.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes. The data was collected from publicly available sources, and all personal identifiers have been removed to ensure privacy and ethical compliance.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, the paper discusses that all personally identifiable information has been removed to ensure privacy. While the dataset may include offensive content due to the nature of the conflict-related discussions, this is inherent to the public data being analyzed and is addressed in the context of ethical use.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see (FORCE11 2023))? Yes, see Usage Notes section.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? Yes. If the paper is accepted, the datasheet will accompany the camera-ready version, as including the datasheet would exceed the submission page limit.

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? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

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

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

Ethical Considerations

In developing this dataset, we have carefully considered potential negative societal impacts. We recognize that the study of politically sensitive content, especially in the context of conflicts such as the Russia-Ukraine situation, requires a rigorous ethical framework. Hence, we anonymized any personally identifiable information, adhering to ethical standards. We emphasize that any interpretations and conclusions that may be drawn from the dataset are aimed at understanding online communication patterns for academic purposes and should not be used to exacerbate any political conflicts. We are committed to monitoring and addressing any potential misuse that may arise from disseminating this work, ensuring that the research contributes positively to the research community and beyond.