

Disinformation Demasked- Research methodology

Methodology

In this study, we systematically analyze 20,000 posts extracted from the Russia Today and Sputnik Arabic channels on Telegram using its official Application Programming Interface (API) to understand how the Russian Federation uses Telegram, as a rapidly growing platform, to promote certain Russia-related narratives to Arabic audiences.

Given the large size of the dataset, we attempt to use Artificial Intelligence, specifically Large Language Models, to do the coding. In order for us to provide sufficient data to be used for prompting, we manually code the top 275 most forwarded messages from the channels (in a combined dataset) and run the process through the research team in a way that ensures consistency and inter-coder reliability.

To find narratives, the coders start by categorizing the posts into themes and subthemes. We relied exclusively on the textual content that is provided in the post and have not done any further systematic analysis of any multimedia content such as images, videos, or links to external web pages.

The approach requires two stages. The first requires manual coding (or labeling) and thereafter, the data is passed to an AI research team to run it through the LLM and proceed with labeling the rest of the data. Two validation stages are done with LLM-based labeling; the first batch of 100 labeled posts are reviewed and validated by the human coders before providing feedback. Thereafter, a second batch of an additional 100 posts are labeled by the LLM followed by a second validation to minimize errors and increase output quality. At that point, LLMs are used to label the remainder of the dataset. This scalability is crucial for understanding the broader patterns and strategies in Russian influence operations, as larger datasets can provide more comprehensive insights and strengthen the robustness of our conclusions.

This process generates a dataset that directly reflects the themes and sub-themes extracted from the textual content of the 20,000 posts we analyze. The AI-LLM technique is rather experimental but promising since the formulation of the instructions (or prompts) are critical for higher quality outputs. The labeling is done on multiple LLMs to compare quality at the initial stage before proceeding with the remainder of the dataset.

From the top manually coded 275 posts, 20 narratives were extracted with example posts provided for reference. This manual process involves an in-depth examination of each piece of content to identify and label the themes and sub-themes present. The produced datasets are distinct through the data labeling with reference to the coder as one of the variables in the dataset. This allows for comparison between human and AI-labeled entries.

Connecting to Expected Themes

The output representing the themes generated through this method is then compared to the expected themes based on prior research (Janadze, 2022), which include:

- **Russia and Energy:**

Exploring narratives around Russia's role in global energy supply.

- **Russia and Islam:**

Assessing how Russia's relationship with Islam and Muslim minorities is portrayed.

- **Putin as a Strong Leader:**

Evaluating portrayals of Vladimir Putin's leadership qualities.

- **Russia as a Great Power:**

Analyzing depictions of Russia's military and geopolitical strength.

- **Russia as an Ally to Arab Countries:**

Investigating narratives of Russia's alliances and diplomatic efforts in the Arab world.

- **Russia Challenging the West:**

Understanding how Russian media frames its opposition to Western policies and values.

- **Russia and the War in Ukraine:**

Examining the portrayal of the conflict in Ukraine and Russia's role in it.

The LLM-labeling approach is meant to be experimental to explore possible uses and applications in labeling large volumes of textual data if provided with sufficient contextual information in the instructions. This technology remains in its infancy so it is not expected to be bullet proof. Nonetheless, this approach aims at providing factual information to minimize potential bias to affect the overall outcome.

Manual coding of disinformation and propaganda

The themes and sub-themes used in coding each individual Telegram code are provided as part of the instructions to coders who underwent a pilot phase with 50 per coder and revise the themes to ensure high inter-reliability reliability. To ensure coders understand their application we drafted a list of codes linked to the list of expected themes mentioned, we used the pilot phase to add further variables/values as required to achieve maximum consistency across the coders. This would not guarantee that some new possible themes/sub-themes would emerge at a later stage, however. In this case, coders labeled the theme as ‘Other’ and used the ‘Comments’ variable to add description as to what the additional/new theme represents exactly. In cases where multiple themes or sub-themes are spotted, there is a special variable called ‘Multiple themes’ where those themes are explicitly added. Table 1 shows the codebook for the values entered by the coders. There are metadata variables whose values are extracted automatically using the Telegram API, which are available on their website at <https://core.telegram.org/constructor/message>. Table 1 shows the values that coders need to enter for each unique Telegram message upon checking the text and link to the message in Arabic language, and which is provided with the API.

Main theme	See Table 2 for values. “Multiple themes” is chosen if more than one theme applies
Subtheme	See Table 2 for values.
Confidence level	How confident the coder is in the labeling of the themes (Likert scale is used ‘Very high’, ‘High’, ‘Moderate’, ‘Low’, ‘Very low’)
Multiple themes	If Main theme is used as ‘Multiple themes’, this is where all the applicable themes from Table 2 are entered
Comments	An open-text explanation of the rationale behind the coder’s decision to code in the way he/she did
<i>Table 1: Variables that coders enter for each Telegram message</i>	

The themes and sub-themes are as shown in Table 2 below:

Main Theme	Example
0-NOT RELATED TO RUSSIA	if the message contains no Russia-related narratives. Example: Neutral news report on a particular event.
1-Russia and energy	Promotes Russian energy. Example: Report on Russia increasing oil exports.
2-Russia and Islam	Relates to Russia and Islam. Example: Story on mosque opening in Moscow.
3-Russia has allies	Showcasing support for Russia from other allies. Example: China endorses Russia’s position in the peace negotiations with Ukraine.
4-Russia as a great power	Shows Russia as a global force. Example: Commentary on Russian influence in global politics in the BRICS efforts.

5-Russia as an ally to Arab countries	Depicts Russia as a friend to Arab nations. Example: Analysis of Russia-Arab trade relations.
6-Russia challenging the West	Russia opposes Western countries. Example: Editorial on Russia's stance against NATO.
7-Russia and the war in Ukraine	Russia's role in Ukraine. Example: Report justifying Russian actions in Ukraine.
8-Other	Doesn't fit any theme/multiple themes. Explain in 'Comments'. Example: Article talking about a dispute between Russia and Georgia.
Multiple themes	Post contains multiple themes (ensure you mention the main and sub themes detected in the column 'Multiple themes')
Sub Theme	Example
1.1-Russia delivering sufficient gas and oil	Russia's energy supply capabilities. Example: News on Russia fulfilling energy contracts.
1.2-Russia exporting nuclear power (building new nuclear plants)	Russia's nuclear energy exports. Example: Article on Russia building nuclear plant abroad.
1.3-Other sub-theme	Doesn't fit other energy sub-themes. Example: Unique report on Russian renewable energy.
2.1-Islam as an official religion in Russia	Portrays Russia as protective of Muslims. Example: Story on Muslim cultural festivals in Russia.
2.2-Russia protects/respects Muslim minorities in Russia (10% of the population)	Portrays Russia as protective of Muslims. Example: Story on Muslim cultural festivals in Russia.
2.3-Other sub-theme	Doesn't fit other Islam sub-themes. Example: Report on interfaith dialogues in Russia.
3.1-China supporting Russia	Focuses on the support for Russia from China. Example: China criticizes Western efforts to isolate Russia in the UN Human Rights Council.
3.2-Iran supporting Russia	Focuses on the support for Russia from Iran. Example: Article on Iran's President's call for ending the embargo against Russia.
3.3-India supporting Russia	Focuses on the support for Russia from Belarus. Example: Modi calls Putin to express support to Russia.
3.4-Belarus supporting Russia	Focuses on the support for Russia from Belarus. Example: Belarus providing support for Russia.
3.5-Other sub-theme	Doesn't fit other sub-themes. Example: Op-ed on how African countries are standing with Russia.
4.1-Russia as an unbeatable military power	Russia's military might. Example: Report on Russian military exercises.
4.2-Russia as a nuclear power	Russia's nuclear capabilities. Example: Analysis of the destructive capacity of Russia's nuclear arsenal.

4.3-Russia as a stabilizer in MENA	Russia as a stabilizing force in MENA. Example: Commentary on Russian peacekeeping.
4.4-Russia delivering Lucrative Military Weapons	Russia's arms trade. Example: News on Russian arms deals with MENA countries.
4.5-Russia's role within the BRICS states (MENA: Saudi Arabia, UAE)	Russia in the BRICS context. Example: Report on Russia-BRICS economic summit.
4.6-Other sub-theme	Doesn't fit other great power sub-themes. Example: Unique insight into the Russian space program.
5.1-Russia and the Palestinian cause	Russia's stance on Palestine. Example: Editorial on Russia's support for Palestine.
5.2-Russia promoting peace in Syria	Russia's role in Syrian peace. Example: Report on Russian mediation in Syria.
5.3-Russia guaranteeing food security in MENA	Russia aiding MENA food security. Example: Analysis of Russian grain exports to MENA.
5.4-Russia and Cultural Diplomacy in MENA	Russia's cultural outreach in MENA. Example: Feature on Russian cultural festivals in MENA.
5.5-Russia reporting on popular events (culture, sports, etc.)	Russia's coverage of MENA events. Example: Russian media report on major MENA sports events.
5.6-Other sub-theme	Doesn't fit other ally sub-themes. Example: Story on Russian technology transfers to MENA.
6.1-Russia as a moral beacon for conservative values	Russia's conservative stance. Example: Article on Russia's family values campaigns.
6.2-Russia opposing LGBTQ+ rights	Russia's stance on LGBTQ+ rights. Example: News on Russian policies against LGBTQ+.
6.3-Russia as anti-imperialist	Russia is portrayed as anti-imperialist. Example: Op-ed on Russia's foreign policy independence.
6.4-Russia promoting a multi-polar world (esp. with China)	Russia's vision of a multi-polar world. Example: Analysis of Russia-China relations.
6.5-Other sub-theme	Doesn't fit other challenging West sub-themes. Example: Report on Russian internet censorship.
7.1-Ukraine as a Failed State	Depicting Ukraine as unsuccessful. Example: Commentary on Ukraine's political instability.
7.2-Ukraine not a sovereign state	Denying Ukraine's sovereignty. Example: Expert quote claiming Ukraine is controlled by the West.
7.3-Ukraine led by Nazis and drug addicts	Specific claims about Ukraine's leadership. Example: Article alleging extremist leaders in Ukraine.
7.4-Ukraine Developing	Ukraine's alleged biolabs. Example: Report on suspected biolabs

Bio Labs on own Territory	in Ukraine.
7.5-Ukraine committing war crimes	Accusing Ukraine of war crimes. Example: News on alleged Ukrainian attacks on civilians.
7.6-Ukraine spreading disinformation	Alleging Ukrainian disinformation. Example: Story debunking Ukrainian news reports.
7.7-Other sub-theme	Doesn't fit other Ukraine sub-themes. Example: Unique perspective on Ukrainian culture.
8.1-Other sub-theme	Doesn't fit any sub-theme/ fits multiple. Explain in 'Comments'. Example: Mixed analysis of the Russian economy and global politics.
<i>Table 2: Possible values for themes and sub-themes for each Telegram post</i>	

Methodological Discussion

One of the key strengths of our approach lies in its comprehensive and multi-levelled nature. By integrating AI-LLM labeling with manual coding, we are able to analyze a very large dataset to get insights of narratives and themes. The combination of human, manual coding with AI coding is a form of triangulation that strengthens the overall methodology and allows us to capture a wide range of themes and narratives that might be present in the data. Furthermore, the use of the Telegram API for data extraction enhances the efficiency and systematic nature of our data collection process, ensuring a robust dataset for analysis.

Another significant advantage is the scalability of our methodology. The study starts with an analysis of 20,000 posts, but the approach is designed to be applied to much larger datasets, potentially extending to millions. This scalability is crucial for uncovering broader patterns and strategies in Russian narratives, as larger datasets can provide more comprehensive insights and bolster the robustness of our conclusions.

However, there are several limitations and weaknesses inherent in our methodology that must be acknowledged. One of the primary concerns is the limitations of AI-LLM tools. While these technologies are powerful, they may not always capture the nuances and complexities of human language, particularly in the context of intricate and evolving narratives that require a comprehensive understanding of context as they may contain allusions, references or innuendos which may not be captured by AI-LLM. This limitation could affect the accuracy of the themes and subthemes identified through AI-driven analysis.

In addition to the limitations of AI, the subjectivity inherent in the manual analysis of non-textual content poses its own set of challenges. This part of the analysis depends heavily on the interpretations and judgments of the analysts, which may introduce biases or inconsistencies. This subjectivity is a crucial factor to consider, especially when analyzing content like videos and images, where meanings can be more ambiguous and open to interpretation. Furthermore, our focus on two specific Telegram channels may limit the generalizability of our findings to other Russian news platforms or different linguistic contexts. This limitation must be taken into account when extrapolating our findings to broader discussions about Russian disinformation strategies.

In terms of validity, reliability, and generalizability, our study faces several challenges. Ensuring the validity of both AI models and manual interpretations is critical for the accuracy of our findings. The reliability of our methodology hinges on the consistency and repeatability of our analytical procedures across both AI-driven and manual analyses. As for generalizability, while our study offers valuable insights into the specific channels and content types analyzed, extending these findings to other contexts or media outlets requires careful consideration and further investigation. While a similar methodology may be applied to other influence operations, the specificity of context will always have to be taken into account.

This methodological framework, combining manual and AI-driven labeling of data to identify narratives, ensures a thorough analysis of Russian narratives in Telegram. It not only uncovers the dominant themes but also allows us to delve deeper into specific high-engagement content that might carry significant weight in shaping public perception. The insights gained from this study will be crucial for understanding the dynamics of disinformation campaigns on social media platforms like Telegram. While our methodology provides a detailed and nuanced framework for analyzing Russian narratives, it is essential to be cognizant of its limitations. The study's findings should be contextualized within the broader landscape of media narratives and disinformation campaigns. Acknowledging and addressing these methodological considerations will be crucial for the interpretation and application of our results.

Findings

Findings highlighted various narratives. Among the top 100 most viewed messages, themes ranged from promoting RT Arabic's live broadcasts to discussions on Russia's historical interactions with global figures like Saddam Hussein. Notably, messages often portrayed Russia in a positive or defensive light, particularly in the context of international criticism or conflict. An example includes a message by Ramzan Kadyrov, head of the Chechen Republic, warning Ukrainians against joining nationalist groups, a narrative potentially aligning with Russian political strategies.

Sputnik Arabic analysis revealed a consistent focus on military and geopolitical themes. Several messages featured footage from the Russian Ministry of Defense, suggesting an intent to showcase Russian military operations positively. For instance, a message highlighting the destruction of Ukrainian military equipment by the Russian army stood out as a clear example of propaganda, aimed at demonstrating Russian military prowess.

The findings indicate a strategic use of these channels to disseminate narratives that favor Russian perspectives, especially in geopolitical conflicts. The emphasis on military successes and the portrayal of opposing forces, particularly in the context of the Ukraine conflict, align with typical propaganda tactics. This approach seems designed to shape public opinion in the Arabic-speaking world, potentially influencing perceptions towards Russia and its international stance.

With themes ranging from the promotion of RT Arabic's broadcasts to nuanced discussions about Russia's global diplomatic interactions and military engagements, a clear contour of Russian state messaging emerges. The portrayal of Russia as a resilient force and a beacon of stability, especially in the context of the Russian-Ukrainian conflict, is recurrent. The presence

of esteemed Russian figures in these narratives further accentuates the strategic dissemination of viewpoints intended to shape public perception.

As the AI team delves into topic modeling, this project aims to revisit and refine the expected themes, ensuring a robust and nuanced understanding of the disinformation and propaganda landscape. The upcoming classification findings will illuminate how these content types are crafted, enabling a clearer linkage and identification of propagandistic and disinformation elements within the data.

In terms of data preservation and structure, general data management principles will be applied. The database will be designed to capture a variety of data points such as source credibility, post interactions, and user profiles. Labels will be meticulously assigned, reflecting a spectrum from clear disinformation to verified information, with crowd-sourced and expert annotations enriching the dataset. The database will serve as a bedrock for training machine learning models that can discern patterns of disinformation, analyze network behavior, and predict the likelihood of content being propagandistic in nature.

The project's database will be structured with flexibility and scalability in mind, accommodating various data types and allowing for sophisticated analysis. This structure will enable the deployment of machine learning models tailored to specific tasks, such as text analysis, timing of replies, or full network analysis, offering insights into how disinformation and propaganda proliferate across social networks.

This initiative stands at the forefront of a comprehensive effort to dissect and understand state-sponsored disinformation and propaganda campaigns. The intricate dance between AI and human analysis, underpinned by a robust database structure, will provide a granular view of the strategies employed to sway public opinion and shape geopolitical narratives.