

# Disinformation Demasked- Project Report

## Table of Contents

Introduction .....	1
Literature review .....	2
Antecedents of Information Disorder in the Global South .....	2
Disinformation in the MENA region .....	3
Benefiting from Global South insights .....	4
Methodology.....	4
Connecting to Expected Themes .....	5
Manual coding of disinformation and propaganda.....	6
Methodological Discussion.....	9
Findings.....	10
Conclusion.....	12
References .....	14

## Introduction

In an era where online information wields significant influence on attitudes, political actions and power relations, a comprehensive project was undertaken to decode and analyze Russian propaganda and disinformation targeting the Arabic-speaking world. The focus was centered on Russia Today (RT Arabic) and Sputnik Arabic, chosen for their direct affiliations with the Russian state, thus offering a transparent window into official Russian narratives.

The project's methodology hinged on leveraging Telegram, a platform gaining traction, especially in Russian contexts, for its less restrictive content dissemination. Unlike mainstream social media platforms prone to filtering and censorship, Telegram provided an unfiltered conduit to the data. This approach yielded a substantial corpus, comprising approximately 170,000 messages from Sputnik Arabic and 277,000 from RT Arabic.

Advanced AI-based Natural Language Processing (NLP) techniques were employed to parse through the vast dataset and identify dominant themes. However, the analysis was not solely reliant on technology. A team of experts, well-versed in Arabic-language news and Russian state media narratives, conducted a manual qualitative inspection of the most viewed and forwarded Telegram posts, including those containing videos. This human element ensured that every step of the analysis was cross-verified, guaranteeing the reliability of the findings.

## Literature review

Worldwide, there is increasing concern about an ‘information disorder’, which includes phenomena such as misinformation, hate speech, and orchestrated deceit. This phenomenon arises from the interplay between evolving digital media landscapes and polarized political climates. The burgeoning interest in disinformation studies, while criticized for lacking historical context, underscores its significance (Bernstein, 2021). There is however a disproportionate focus on Northern perspectives, particularly triggered by events like the 2016 US election, and consequently, there is a need for a more inclusive approach to understanding disinformation. The Global South’s extensive experience with information disorder predates contemporary scholarly interest, urging a recalibration of research agendas toward broader geographical, cultural, and geopolitical diversity (Lewis & Molyneux, 2018).

## Antecedents of Information Disorder in the Global South

Historically, journalism in the Global South has grappled with manipulation, state censorship, and political pressures. Alternative communication channels outside of formal journalism have often served progressive purposes, using humor, oral traditions, and contextually specific cultural expressions. Present-day manifestations of disinformation in the Global South thus reflect deep-seated socio-political dynamics and historical legacies (Bebawi, 2022; Ellis, 1989; Mano, 2007).

In recent years, concerns have been growing over the pervasive influence of digital disinformation, particularly in fragile democracies of the Global South. Research has indicated high levels of exposure to fabricated political narratives in these contexts, even more so than in Western democracies. Societies in the Global South are often economically highly unequal, and marked by ethnic tensions going back to the colonial era. These inequalities and tensions are exploited in hate speech that is spread on social media and mobile messaging platforms. The COVID-19 pandemic has brought these problems into the spotlight, and local political leaders have often leveraged anti-Western sentiments to propagate harmful disinformation (BBC, 2020; Davis, 2014; Steinberg, 2007). Disentangling the complex web of disinformation in the Global South necessitates a detailed knowledge of context. Studying these contexts also requires a multi-dimensional approach, integrating socio-technical analyses with historical and geopolitical insights. By contextualizing media ecologies within broader socio-political dynamics, scholars can discern the underlying drivers of information disorder and formulate more nuanced responses (Gagliardone, 2016; Marwick, 2018).

## Disinformation in the MENA region

Disinformation in the MENA region spans a range of areas, including political, economic, social, security, health, cultural, artistic, sports and technological sectors (Mortada et al., 2022, p. 26).

Disinformation often originates from authoritarian regimes which ‘upgrade their capacity to resist dissent and adapt to new challenges (such as social media’ and regional power plays between state actors in the region, and the invocation of external enemies in an attempt to deflect criticism from within their countries (Jones, 2022, 13). Disinformation in the region may also relate to sectarian religious tensions, or fake news disseminated to undermine or redirect activism against political leaders (Mortada et al., 2022:25).

Although disinformation is a common feature of online communication in the MENA, the region has received comparatively little attention in the academic literature on geopolitical disinformation, which, according to Jones (2022, p.8) has tended to focus on ‘transatlantic security concerns’.

Social media companies have also tended to neglect the region when it comes to moderating disinformation and dealing with platform manipulation (Jones, 2022, p.8). To make things worse, the region has also been experiencing increased suppression of free expression, which makes it difficult for journalists and independent media outlets to do their work and counter disinformation and authoritarianism in the information space (Mortada et al., 2022, p. 25).

There has been a growing network of researchers, activists, policymakers, journalists, and fact-checkers working on understanding and countering disinformation in the MENA region (Jones, 2002, p. 9). Several fact-checking platforms have been established in recent years, notably in response to the COVID-19 pandemic. As Mortada et al. (2022, p. 28) point out, these platforms have played an important role in verifying Arab-language information on social media. More than fifty fact-checking organizations exist across the region, although some of these have become active, and there is no regional networking framework or umbrella organization for these initiatives (Mortada et al., 2022, p. 28). Apart from this lack of coordination, funding remains a challenge for fact-checking platforms across the region. These organizations are also hampered by a lack of training, little public interest, a lack of legal protection, and inadequate technological support for their work (Mortada et al., 2022, p. 33).

These challenges that confront counter-disinformation efforts in the region are further exacerbated by poor regulation and the commercial orientation of social media companies, at the same time as there is a growing uptake and spread of digital technology in the region, making it possible for disinformation and deception – often a result of collusion between regional state actors, hackers, PR agencies, political consultants and influencers – to spread within and beyond the MENA region (Jones, 2022, p. 69).

## Benefiting from Global South insights

Studying disinformation in the Global South offers invaluable lessons for addressing similar challenges worldwide. Early manifestations of geopolitical interference and weaponization of disinformation, observed in regions like Ukraine and the Philippines, foreshadowed later developments in Western contexts. The interconnected nature of disinformation ecosystems, transcending geographical boundaries, makes it important to study the problem of information disorder from a global perspective (Mozur, 2018). It is important to understand the range of factors, including media landscapes, socio-cultural dynamics, political systems, and geopolitical relations, that influence the problem of disinformation in the Global South. It is equally important to understand the historical factors that continue to shape the social dynamics of these regions and the more contemporary global forces that play out in the information landscape. Not only political dynamics but also sociocultural factors, including community orientations and economic disparities, influence information consumption patterns and media users' susceptibility to disinformation (Madrid-Morales et al., 2021; Tully et al., 2021).

Political upheavals, often intertwined with media dynamics, create fertile ground for disinformation campaigns, exploiting ethnic and socio-economic fault lines. The problem is often made worse by the heavy-handed responses by governments, which range from censorship to crackdowns on activists and protesters. These responses, frequently justified by authoritarian governments' claims to counter disinformation, have resulted in a contested and fraught information landscape (RSF, 2021). Addressing disinformation in the Global South therefore demands tailored responses that acknowledge its intricate socio-political contexts. The organizations engaged in combating information disorder in this region therefore often have to combine different strategies, such as political advocacy, media literacy, and rights-based interventions (Wasserman, 2022).

By integrating multiple imperatives, these organizations epitomize the need for holistic, context-sensitive approaches to counter disinformation. Embracing user agency and leveraging real-world networks are crucial steps towards fostering resilience against information disorder in the Global South (Gagliardone et al., 2021).

## 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.

<b>Main theme</b>	See Table 2 for values. “Multiple themes” is chosen if more than one theme applies
-------------------	--

<b>Subtheme</b>	See Table 2 for values.
<b>Confidence level</b>	How confident the coder is in the labeling of the themes (Likert scale is used 'Very high', 'High', 'Moderate', 'Low', 'Very low')
<b>Multiple themes</b>	If Main theme is used as 'Multiple themes', this is where all the applicable themes from Table 2 are entered
<b>Comments</b>	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:**

<b>Main Theme</b>	<b>Example</b>
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')
<b>Sub Theme</b>	<b>Example</b>
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	Russia's coverage of MENA events. Example: Russian media



popular events (culture, sports, etc.)	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 Bio Labs on own Territory	Ukraine's alleged biolabs. Example: Report on suspected biolabs 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.

## Conclusion

It is clear that understanding disinformation globally and finding successful strategies to counter it, will require more attention to the Global South perspectives. More diversity in disinformation studies, policy design, and regulatory responses can benefit the field as a whole and ultimately result in more effective global responses to the scourge of disinformation (Tumber & Waisbord, 2021).

This project is a significant contribution to this aim of understanding influence operations and disinformation strategies outside the Global South, with a particular focus on how Russia's disinformation strategies manifest in the Middle East and North Africa region. Not only did the project contribute new data that identified the dominant narratives in Russian influence operations, but it also employed a novel methodology that incorporates advances in artificial intelligence to identify and analyse disinformation narratives at scale.

The study collected data from Russia's state-backed media outlets, specifically the Sputnik and RT Arabic Telegram channels and analyzed this data through a rigorous manual coding process to identify major narratives within these datasets. This project stands as a significant endeavor in the landscape of media analysis and narrative detection. This initiative is of importance for several reasons, each highlighting the need to understand and counteract the influence of state-backed propaganda efforts, particularly in the context of the Middle East and North Africa.

Firstly, the project represents a joint effort to systematically gather and scrutinize content from media outlets known to be influenced by the Russian state, thereby providing a structured approach to understanding how such entities operate and disseminate information. This work is crucial in the era of information warfare, where state actors often employ media outlets to shape public opinion, sow discord, and advance their geopolitical interests.

The importance of the database created as a result of this project lies not only in its function as a repository of analyzed content but also in its potential as a tool for researchers, policymakers, and the general public. By making this database accessible, the project empowers various stakeholders to engage with the data, fostering a more informed understanding of the narratives being pushed by these media channels. This, in turn, enhances our collective ability to discern truth from state-crafted narratives.

This database will also be useful to the fact-checking community, as understanding the main themes of narratives serving Russian propaganda in the Middle East will help them draw patterns between these narratives spread by Russian official media outlets and the disinformation spread by unattributed social media pages.

The most common keywords discovered in the database will also serve as search terms for fact-checkers to navigate through the flow of disinformation to find suspicious material.

Moreover, this project plays a pivotal role in challenging Russia's state-backed media outlets. RT and Sputnik have a long-established presence in the Middle East, portraying their reporting as professional journalism. By unveiling the strategies and narratives employed by these outlets, the project contributes to a broader effort of challenging the perception that RT and Sputnik are merely news media outlets, with no hidden state-backed agenda of spreading propaganda and disinformation. This study equips stakeholders with evidence-based insights, enabling them to critically assess the information disseminated by these channels and to question their credibility and intentions.

The identified narratives within the project's dataset are particularly revealing, shedding light on the nature of Russia's propaganda efforts in the Middle East. This aspect of the project is instrumental in understanding the specific themes and messages that are being targeted at Middle Eastern audiences, thereby uncovering the subtleties of Russia's influence operations in the region. By doing so, the project not only adds to our understanding of the geopolitical aims behind such narratives but also aids in formulating responses to mitigate their impact.

Finally, the project's exploration into building knowledge on testing Large Language Models (LLMs) for narrative detection marks a significant advancement in the field of computational linguistics and artificial intelligence. This innovative approach highlights the project's contribution to the development of tools and methodologies that can automate the detection of narratives within large datasets. The implications of this are far-reaching, offering the potential to scale up efforts to monitor and analyze propaganda across different languages and media platforms, thus broadening the scope of research and intervention in the fight against disinformation.

## References

Anderson, C.W. (2021). Propaganda, misinformation, and histories of media techniques. Harvard Kennedy School Misinformation Review. April 15. <https://misinforeview.hks.harvard.edu/article/propaganda-misinformation-and-histories-of-media-techniques/> Accessed 29 October 2021

Bachmann, I., Grassau, D., Labarca, C. (2022, forthcoming) Aliens, Spies, And Staged Vandalism: Disinformation in the 2019 Protests in Chile. In: Wasserman, H. & Madrid-Morales, D. (Eds). (2022, forthcoming). *Disinformation in the Global South*. New York: Wiley-Blackwell

BBC (2020) Coronavirus: What misinformation has spread in Africa? <https://www.bbc.com/news/world-africa-51710617> Accessed 30 October 2021

Bebawi, S. (2022, forthcoming) Disinformation in Arab Media: cultural histories and political dynamics. In: Wasserman, H. & Madrid-Morales, D. (Eds). (2022, forthcoming). *Disinformation in the Global South*. New York: Wiley-Blackwell.

Berger, G. (2022, forthcoming). Foreword. In: Wasserman, H. & Madrid-Morales, D. (Eds.) (2022, forthcoming). *Disinformation in the Global South*. New York: Wiley-Blackwell.

Bernstein, J. (2021). Bad News: Selling the Story of Disinformation. *Harpers Magazine*. September. <https://harpers.org/archive/2021/09/bad-news-selling-the-story-of-disinformation/> Accessed 11 October 2021.

Carlson, M. (2020). Fake news as an informational moral panic: The symbolic deviancy of social media during the 2016 US presidential election. *Information, Communication & Society*, 23(3), 374–388. <https://doi.org/10.1080/1369118X.2018.1505934>

Cowan, K. (2020) FACT CHECK: How a lie about Bill Gates and a vaccine for the coronavirus exploded on Twitter. *News 24* 7 April. <https://www.news24.com/news24/Analysis/fact-check-how-a-lie-about-bill-gates-and-a-vaccine-for-the-coronavirus-exploded-on-twitter-20200407> Accessed 30 October 2021.

Damstra, A., Hajo G. Boomgaarden, Elena Broda, Elina Lindgren, Jesper Strömbäck, Yariv Tsfati & Rens Vliegthart (2021). What Does Fake Look Like? A Review of the Literature on Intentional Deception in the News and on Social Media, *Journalism Studies*, DOI: 10.1080/1461670X.2021.1979423

Davis, B. (2014). South Africa's Health Minister's Misguided Belief. Duke Global Health Institute. 8 July. <https://globalhealth.duke.edu/news/south-africas-health-ministers-misguided-belief> Accessed 30 October 2021.

Ellis, Stephen. 1989. "Tuning in to Pavement Radio." *African Affairs* 88 (352): 321–330.

Farhall, K., Carson, A., Wright, S., Gibbons, A., Lukamto, W. (2019) Political Elites' Use of Fake News Discourse Across Communications Platforms. *International Journal of Communication* 13(2019): 4353-4375.

Freedland, J. (2021). Is Facebook the tobacco industry of the 21st century? *The Guardian* 8 October. <https://www.theguardian.com/commentisfree/2021/oct/08/facebook-tobacco-industry-research-harm-causes> Accessed 31 October 2021.

Gagliardone, Iginio. 2016. *The Politics of Technology in Africa*. Cambridge: Cambridge University Press.

Gagliardone, I., Diepeveen, S., Findlay, K., Olaniran, S., Pohjonen, M., Tallam, E. (2021). Demystifying the Covid-19 infodemic: Conspiracies, Context and the Agency of Users. *Social Media + Society* July-September: 1-16

Internetlab. (2022, forthcoming). Efforts to tackle disinformation in Latin America and the Caribbean. In: Wasserman, H. (Ed.) (2022, forthcoming). *Confronting the Challenge of Information Disorder in the Global South*. Ottawa: International Development Research Center.

Jones, M.O. (2022). *Digital Authoritarianism in the Middle East*. London: Hurst & Co.

Lewis, S. C., & Molyneux, L. (2018). A decade of research on social media and journalism: Assumptions, blind spots, and a way forward. *Media and Communication*, 6(4), 11–23.

LIRNEasia (2022, forthcoming). Mapping the Information Disorder in Asia. In: Wasserman, H. (Ed.) (2022, forthcoming). *Confronting the Challenge of Information Disorder in the Global South*. Ottawa: International Development Research Center.

Madrid-Morales, D., Wasserman, H., Gondwe, G., Ndlovu, K., Sikanku, E., Tully, M., Umejei, E., Uzuegbunam, C. 2021. Motivations for Sharing Misinformation: A Comparative Study in Six Sub-Saharan African Countries. *International Journal of Communication* 15(2021):1200-1219

Mano W (2007) Popular music as journalism in Zimbabwe. *Journalism Studies* 8(1): 61–78.

Marwick, A.E. (2018). Why Do People Share Fake News? A Sociotechnical Model of Media Effects. *Georgetown Law Technology Review* 474(2018).

Maweu, J. M. (2021) *Media, Ethnicity and Electoral Conflicts in Kenya*. Lanham: Lexington Books.

Mortada, S., Ashfour, A. & Nofal, A. (2022). Middle East and North Africa. In: Wasserman, H. (Ed.). *Meeting the Challenges of Information Disorder in the Global South*. Ottawa: International Development Research Center. <https://idrc-crds.ca/en/project/confronting-challenges-information-disorder>

Mozur, P. (2018). A Genocide Incited on Facebook, With Posts from Myanmar's Military. *New York Times*. 15 October. Accessed 30 October 2021.

Newman, N. (2021). Reuters Institute Digital News Report 2021. Accessed 29 September 2021. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021>

Nothias, T. (2020). Access granted: Facebook's free basics in Africa. *Media, Culture & Society* 42(3): 329-348

RSF (Reporters Without Borders), 2021. RSF 2021 Index: Covid makes African journalism more vulnerable than ever. <https://rsf.org/en/rsf-2021-index-covid-makes-african-journalism-more-vulnerable-ever-0> Accessed 1 November 2021.

Segal, D. (2018). How Bell Pottinger, P.R. Firm for Despots and Rogues, Met Its End in South Africa. *New York Times*. 4 February. <https://www.nytimes.com/2018/02/04/business/bell-pottinger-guptas-zuma-south-africa.html> Accessed 1 November 2021.

Shumba, E. (2021) An investigation into 'Wolf Warrior' diplomacy in Africa via Twitter and Facebook at the height of the COVID-19 pandemic. *South African Journal of International Affairs* 28(3): 471-483

Steinberg, J. (2007). *Notes from a Fractured Country*. Johannesburg & Cape Town: Jonathan Ball

Subedar, A. (2017). Why people believe the myth of 'plastic rice'. *BBC News*. 5 July. <https://www.bbc.com/news/blogs-trending-40484135> Accessed 1 November 2021.

Tandoc, E. C. (2019). The facts of fake news: A research review. *Sociology Compass*, 13(9), e12724. doi:10.1111/soc4.12724

Tandoc, E.C., Jenkins, J. & Craft, S. (2019). Fake News as a Critical Incident in Journalism, *Journalism Practice*, 13:6, 673-689, DOI:10.1080/17512786.2018.1562958

Tully, M., Madrid-Morales, M., Wasserman, H., Gondwe, G. & Ireri, K. (2021): Who is Responsible for Stopping the Spread of Misinformation? Examining Audience Perceptions of Responsibilities and Responses in Six Sub-Saharan African Countries, *Digital Journalism*, DOI: 10.1080/21670811.2021.1965491

Tumber, H. & Waisbord, S. (Eds.) (2021). *The Routledge Companion to Media Disinformation and Populism*. London: Routledge

Udupa, S., Gagliardone, I. & Hervik, P. (Eds.) 2021. *Digital Hate: The Global Conjunction of Extreme Speech*. Bloomington: Indiana University Press



Velluet, Q. (2020). France/Russia: Propaganda War on Facebook targets Mali & CAR. *The Africa Report*. 21 December. <https://www.theafricareport.com/56013/france-russia-propaganda-war-on-facebook-targets-mali-the-car/> Accessed 1 November 2021.

Wardle, C. & Derakshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policymaking*. Strasbourg: Council of Europe

Wasserman, H. & Madrid-Morales, D. (2019) An Exploratory Study of “Fake News” and Media Trust in Kenya, Nigeria and South Africa. *African Journalism Studies*, 40:1, 107-123, DOI: 10.1080/23743670.2019.1627230

Wasserman, H. & Madrid-Morales, D. (2021). South African government’s handling of COVID-19: study shows declining trust. *The Conversation* July 1. <https://theconversation.com/south-african-governments-handling-of-covid-19-study-shows-declining-trust-163372> Accessed 31 October 2021

WHO (World Health Organization) (2021). Infodemic. [https://www.who.int/health-topics/infodemic#tab=tab\\_1](https://www.who.int/health-topics/infodemic#tab=tab_1) Accessed 11 October 2021.