

Clemson University

TigerPrints

Media Forensics Hub Reports

Watt Family Innovation Center Media Forensics
Hub

6-2024

Old Despots, New Tricks - An AI-Empowered Pro-Kagame/RPF Coordinated Influence Network on X

Morgan Wack

Darren Linvill

Patrick Warren

Follow this and additional works at: https://tigerprints.clemson.edu/mfh_reports



CLEMSON® UNIVERSITY
MEDIA FORENSICS HUB

Old Despots, New Tricks

An AI-Empowered Pro-Kagame/RPF Coordinated
Influence Network on X

Morgan Wack, Darren Linvill, Patrick Warren
June 20, 2024



Report Overview

Under President Paul Kagame, who will be elected to a fourth consecutive term in office next month, actors aligned with the ruling Rwandan Patriotic Front (RPF) have routinely used digital platforms to harass and intimidate critics. Prior reporting on the behavior of these actors has primarily focused on specific attacks, accounts, or targets due to their potential for harm. In examining recent extensions of this work, in this report we identify a large coordinated RPF-aligned influence campaign and conduct a thorough analysis of its scale, scope, and techniques.

Key Findings

- *Involved at least 464 accounts and issued over 650,000 messages since January 2024, with output accelerating over time.*
 - *Made extensive use of AI, through the integration of large-language models (ChatGPT) and AI-generated imagery.*
 - *Centralized and coordinated messaging, but decentralized account operation to a large set of individual posters. Published posts that mostly addressed one or more of five broad themes: The DRC (52%), Kagame/RPF (20%), critics of Kagame/RPF (16%), government successes (9%), and sports/tourism/events (8%).*
 - *Used targeted hashtags, for promotion and/or demotion of narratives. At certain times, such as in response the Rwanda Classified report, these accounts produced an overwhelming share of the messages on a hashtag, successfully flooding out the conversation.*
-

Behind the thinly-veiled veneer of legitimacy sustained in part by connections to international and corporate partners, the Rwandan state has developed a capacity for digital repression with the likely goals of mitigating criticism and increasing control. Prior work has illustrated how the state has incorporated several interconnected strategies to accomplish these goals, including use of state propaganda,¹ harassment of critics,² and targeted assassinations.³ Despite the persistent use of these authoritarian tactics, Rwanda has maintained an elevated global standing through the promotion of its economic accomplishments, regional peacekeeping efforts, and prolific international outreach, including its recent commitment to serve as a destination for deported asylum seekers to the United Kingdom.⁴

Bilateral relations are not the only form of reputational maintenance coordinated through Kigali. In recent years, it has been suspected that the Rwandan government has expanded its online influence through connections with global PR firms to control the bounds of appropriate

¹ <https://www.fidh.org/en/region/Africa/rwanda/civic-and-democratic-space-under-total-control-of-rpf>

² https://forbiddenstories.org/actualites_posts/statement/

³ <https://www.hachettebookgroup.com/titles/michela-wrong/do-not-disturb/9781610398435/?lens=publicaffairs>

⁴ <https://www.bbc.com/news/explainers-61782866>



discussion of the state and its achievements.⁵ One prominent layer of this reputation management machine involves the use of a vast network of pro-government social media accounts, which have been linked to efforts to promote the state as well as campaigns to harass journalists who have attempted to investigate the government’s extrajudicial actions.⁶ A second strategy involves the promotion of cultural events, including its continued hosting of the NBA’s Basketball Africa League (BAL),⁷ and external partnerships with prominent global sports clubs, including as Arsenal FC,⁸ Bayern Munich FC,⁹ and Paris Saint-Germain.¹⁰ Building on the attention provided by these partnerships, Rwanda has maintained its image of credibility among global partners, likely due in no small part to strategic ignorance.

In May 2024, the international journalism network Forbidden Stories reported on Rwanda’s use of coordinated social media accounts for online propaganda and harassment.¹¹ This report extends the work of Forbidden Stories and shares the tactics, techniques, and procedures (TTPs) employed by the campaign. Specifically, we detail for the first time in the Rwandan context how government-aligned actors have attempted to amplify the influence of coordinated influence operations (CIOs), which look at influence audiences, through the integration of large language models (LLMs) and generative artificial intelligence (GAI). The use of these methods, which are explicitly banned on X/Twitter and most other social media platforms, provide evidence that these artificial intelligence tools have begun to play a primary role in the use of CIOs by pro-government actors.

By exposing the persistent patterns of behavior enabled by these novel tools, we illustrate in this report high levels of coordination among pro-government social media accounts focused on specific narratives of interest to the Kigali government. Moreover, in detailing our evidence related to this coordinated behavior, we show how the pro-government campaign has recently begun to amplify targeted harassment and propaganda through the integration of generative artificial intelligence (GAI), large language models (LLMs), and visual synthetic media.

Pro-Kagame/RPF Campaign

We identified a network of over 460 accounts that operated on X/Twitter and purported to be in Rwanda (see *Appendix A*). These accounts have been highly active since the start of 2024, although several are much older. Nearly all of the identified accounts remain active as of the publication of this report. As with previous examples of government-aligned coordinated networks, these accounts pushed narratives favorable to Paul Kagame and the RPF while

⁵ <https://www.theguardian.com/world/2024/jan/27/well-camouflaged-dictatorship-rwandans-fear-for-safety-while-dirty-tricks-campaign-undermines-critics>

⁶ <https://forbiddenstories.org/in-the-west-and-online-rwandas-influence-machine-keeps-churning/>

⁷ <https://visitrwanda.com/basketball-africa-league/>

⁸ <https://www.nytimes.com/athletic/5065518/2023/11/15/arsenal-visit-rwanda/>

⁹ <https://sportspromedia.com/news/bayern-munich-visit-rwanda-sponsorship-fan-criticism-sportswashing-reaction/>

¹⁰ <https://onefootball.com/en/news/paris-saint-germain-and-visit-rwanda-celebrate-three-years-of-success-and-extend-their-partnership-37440150>

¹¹ <https://forbiddenstories.org/in-the-west-and-online-rwandas-influence-machine-keeps-churning/>



targeting perceived enemies of the state, particularly those related to the ongoing conflict in the DRC and the release of Rwanda Classified reporting by Forbidden Stories. The campaign, which exhibits several markers of coordinated inauthentic behavior, seems to be trying to affect discourse about the performance of the Kagame regime, the reliability of prominent critics of that regime, and the level of the threat posed by external forces perceived to be opposed to state interests.

By influencing the prominence of discussion related to Rwanda and its government's actions, this campaign has several paths to potential influence. These include impacts for residents of the region as well as journalists and media members attempting to hold the government accountable. To determine the primary targets of the campaign, we coded the 200 most used hashtags produced by the 460 coordinated accounts. Each hashtag was coded (see *Appendix B*) to correspond to its association with one of five themes:

- **Democratic Republic of the Congo:** Discussion of the ongoing conflict, including targeted attacks on the country's president as well as calls for disarmament (see *Narrative: DRC*).
- **Kagame – RPF:** Includes hashtags supporting Paul Kagame or the RPF.
- **Deflection of Criticism:** Addresses perceived critics of the regime, which include specific journalists/reports as well as organizations such as the UN. These discussions employ both flooding and direct attacks (see *Narrative: #RwandaClassified*).
- **Sports – Tourism – Events:** These hashtags are used to promote the state by highlighting its associations with prominent sports teams and destination offerings as well as hosted events (see *Narrative: Sportswashing*).
- **Government Successes:** The promotion of the government through the promotion of its domestic successes.

The total posts produced containing hashtags related to each of these five prominent themes are presented in *Figure 1*.

Output over Time

While some of the identified accounts were created in late 2023, presumably as part of this coordinated effort, many of the accounts in the network are the accounts of real people. Although many of the posts associated with these authentic accounts prior to 2024 were deleted, these accounts remain active beyond their involvement in the campaign.



Figure 1: Total Posts Including Each Prominent Theme by Members of the Coordinated Network

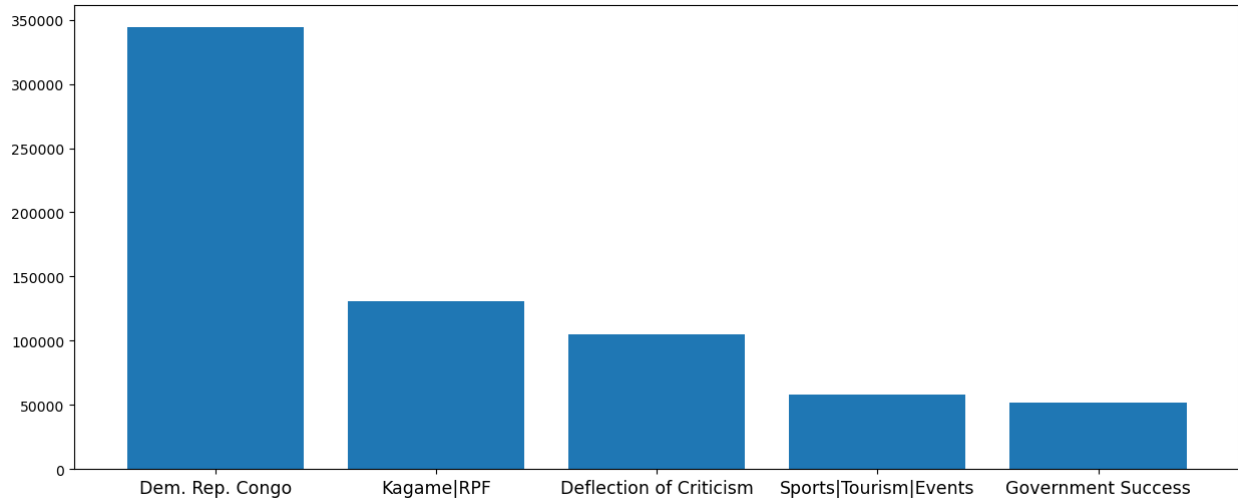
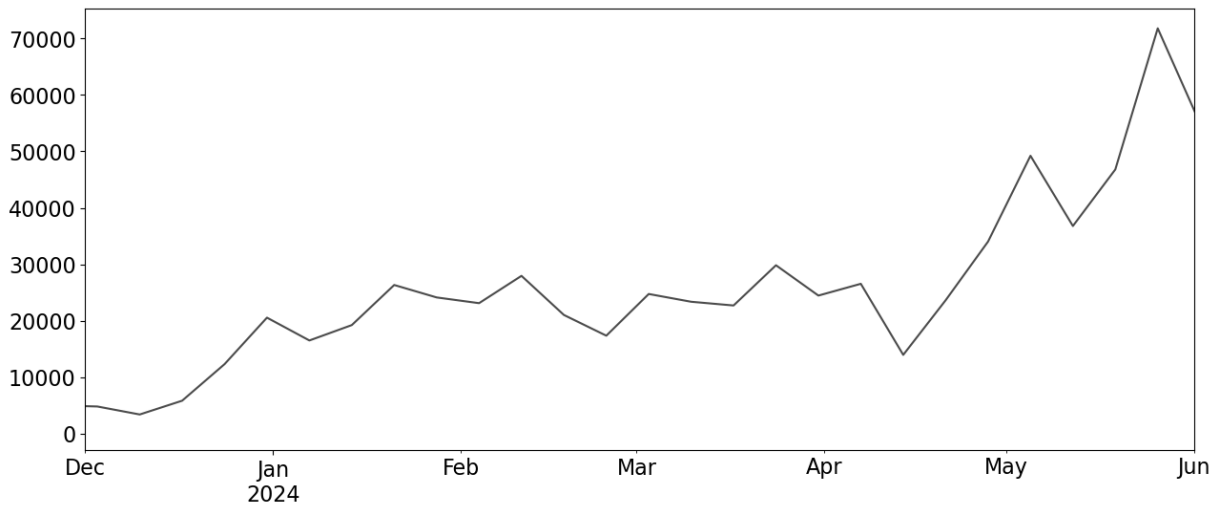


Figure 2: Total Campaign Output on Twitter per Week from Dec 1, 2023 – May 31, 2024



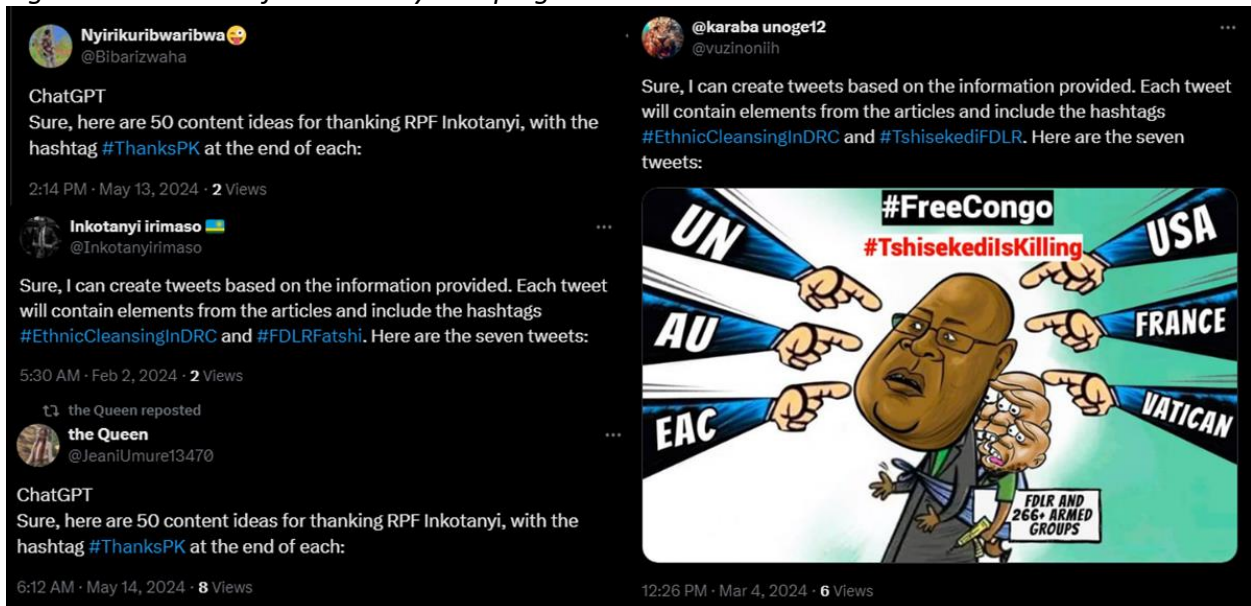


Tactics, Techniques, and Procedures

I. LLM-Based Text Creation

The campaign's most prominent feature is its use of AI-generated text reposted across the campaign's accounts. Specifically, the AI was provided with some background and a prompt to create several variants of messages expressing the same idea, and those variants were posted within and across accounts. The most direct evidence for this tactic comes from several leaks of the prompt from the AI's response that were accidentally included in the posts. Figure 3 includes several examples of these leaks. Each reveals the task assigned to the AI, always to create many variants of posts expressing some specific idea, often including a hashtag. One of the leaks also reveals the specific Large Language Model used for this task, ChatGPT.

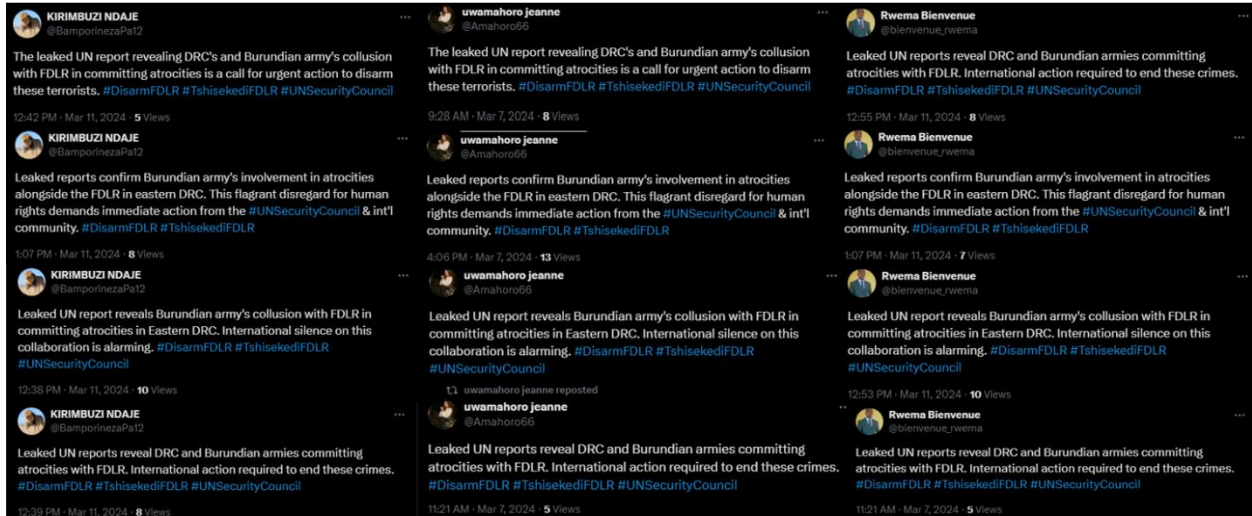
Figure 3: Evidence of LLM Use by Campaign Accounts



The output of these tasks was then shared between and within accounts. Each row of Figure 4 shows a different message, all expressing the same core idea—that the international community should intervene to call out the Burundian army's collaboration with the FLDR in the DRC. All four messages also use a common set of hashtags. The columns represent three different accounts that were part of the campaign, each of which shared all four messages, at slightly different times.



Figure 4. Evidence of LLM Use and Repeated Text



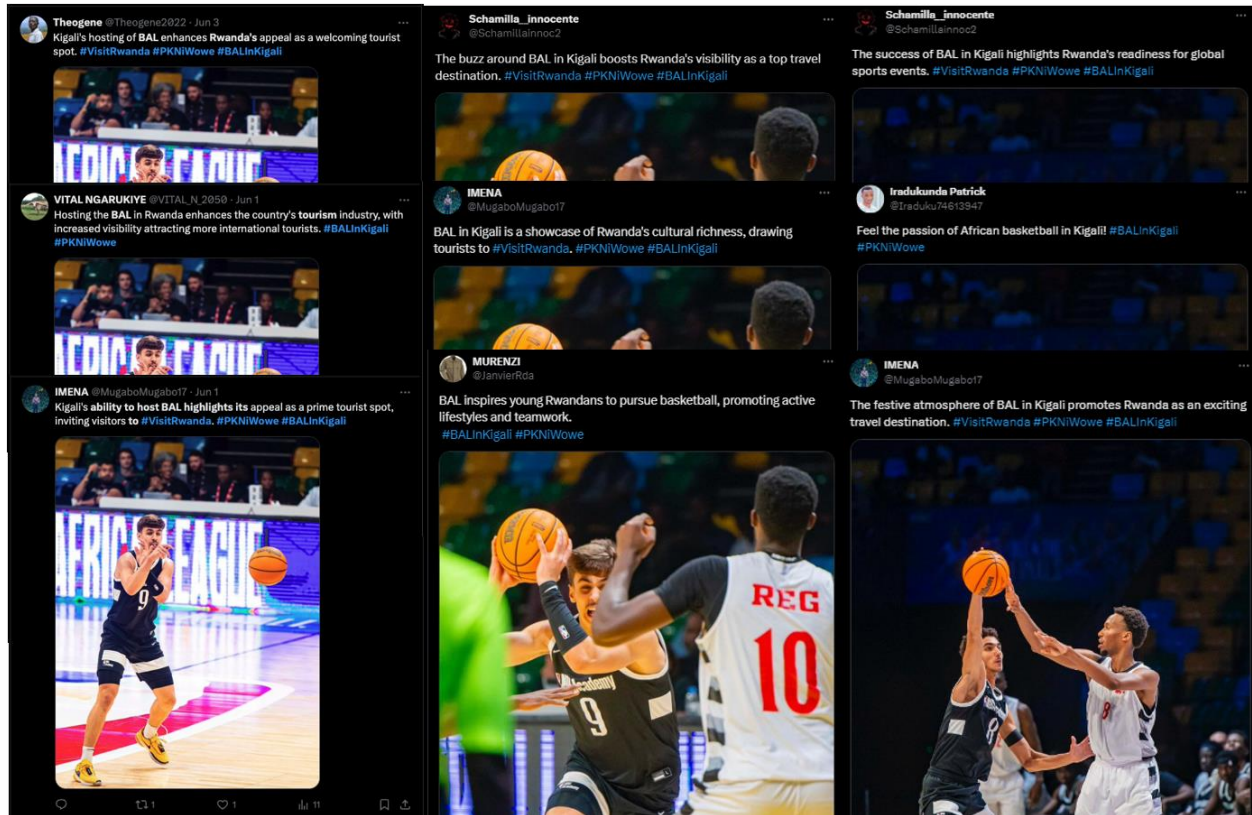
While this strategy still leaves behind evidence that enables it to be tracked, the use of both new (LLM) and old (copy-paste) technology allowed the creation of a massive amount of content quickly, across languages. The increased variety also allowed the campaign to split the variants across different accounts (not including all variants on all accounts), which may have helped avoid detection. Further variation was introduced in the interlacing text variants with media (see below), for combinatorial increases in variety.

In addition to the impact of variety on operational security, it may have also been useful to overwhelm readers. By crowding out or flooding critical discussions of the state with similar, but not identical posts, the speed and diversity enabled by the use of LLMs has allowed the campaign to grow in relevant discussions beyond prior user-only systems of obfuscation.



II. Repeated Imagery

Figure 5: Coordinated Promotion of Images alongside AI-drafted Text



Another prominent feature of the campaign is its reuse of a relatively small corpus of related images. *Figure 5* illustrates this idea in the context of a particular event. Each column shows three different accounts using the same image, often with the same set of hashtags and general message, but with unique text. The text variety was enabled by the integration of LLMs. By spacing out the images across accounts, the network is able to repeat sentiments without the financial/temporal constraints which exist for campaigns reliant on the generation of novel visual media.



III. Use of Visual Forms of AI/Synthetic Media

Figure 6: Use of Synthetic Photos

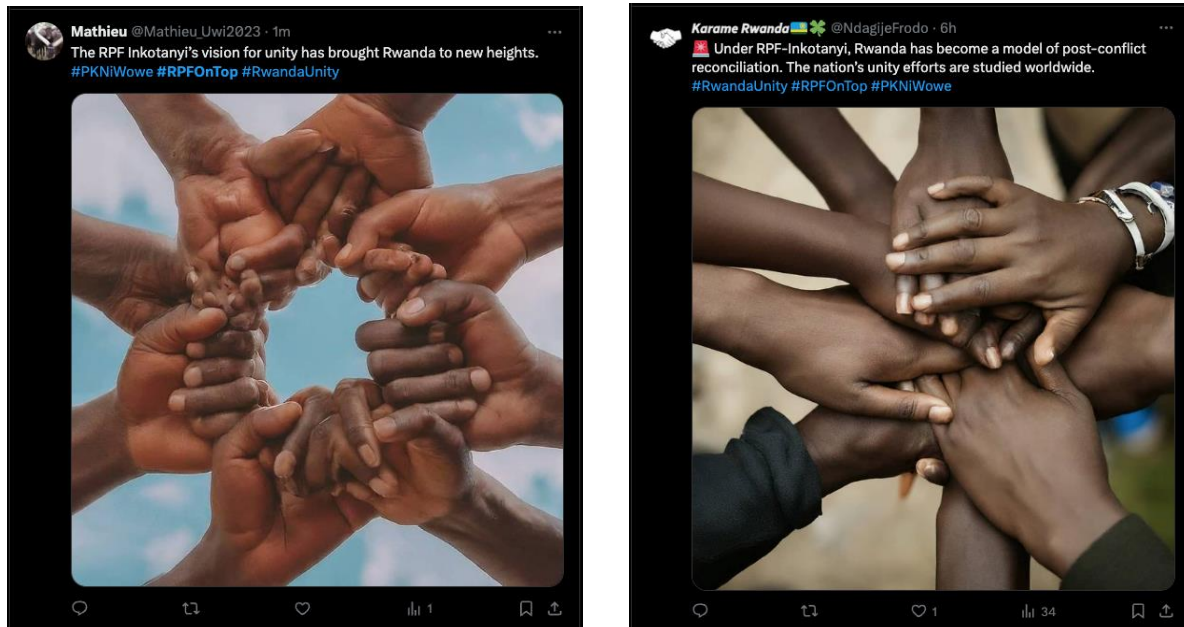
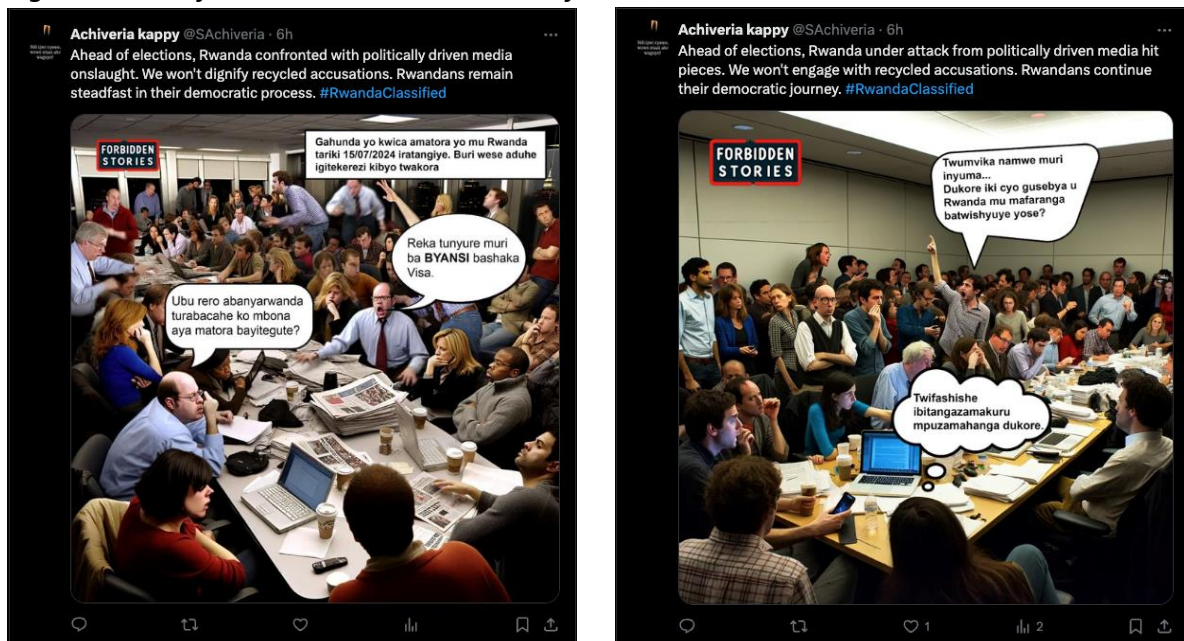


Figure 7: Use of GAI to Create Caricatures of State “Enemies”



Unsurprisingly, given evidence of the prominent use of LLMs and the desire to integrate novel media, the campaign has on several occasions relied on publicly accessible forms of generative



artificial intelligence (GAI) to create images to promote their campaign. This has included the generation of imagery aimed at expressing support for the state and its actions (see *Figure 6*) as well as the development of critical caricatures and scenes which are often further edited to include the target of a pointed attack (see *Figure 7*). Appendix D offers additional ways in which the campaign employed AI text and images.

IV. Centralized Messaging and Media with Decentralized Account Operation

Accounts in the network move together to share content on each targeted theme using common media, but the off-theme messages and other behaviors are idiosyncratic, as are the technical characteristics of the accounts (specific hour of operation, posting client, account creation dates, profile image). This pattern is consistent with individual participants setting up and operating the accounts participating in the campaign with limited supervision and guidance. This structure is consistent with the dual use of personal accounts for coordinated influence practiced for many years by distributed disinformation campaigns associated with China’s “Fifty Cent Party”¹² or Venezuelan astroturfing,¹³ as well as the large collection of small idiosyncratic pods of accounts contributing to campaigns associated with the Indian Army Chinar Corp.¹⁴

V. Hashtag Targeting – Promotion and Demotion

A large share (82%) of the messages from accounts participating in the campaign included hashtags. More specifically, 68% of messages (536,346 out of 787,261) included one of the top-20 hashtags. This is strong evidence for hashtag targeting. Many of the targeted hashtags represented narratives aligned with the goals of the campaigns, but not all. Several, however, represented topics or narratives critical of the Rwandan government (#ForbiddenStories, #RwandaClassified, #UNSecurityCouncil), the targeting of which suggests an effort to demote the prominence of these narratives through flooding and/or insertion of counter narratives into the discourse.

VI. Work-Week Posting

One additional piece of evidence that posting on this campaign occurred as part of a publicly organized effort is the timing of hashtag-targeted posts across days of the work week. We can separate the messages produced by the accounts in the network into two groups, those using hashtags related to the key campaign narratives (see *Appendix B*) and those without. Since many of these accounts are dual-use (personal and campaign, see *Section V*), we expect the set of tweets outside those key narratives to consist primarily of personal tweets.

¹² <https://www.theatlantic.com/technology/archive/2017/01/trolling-by-distraction/514589/>

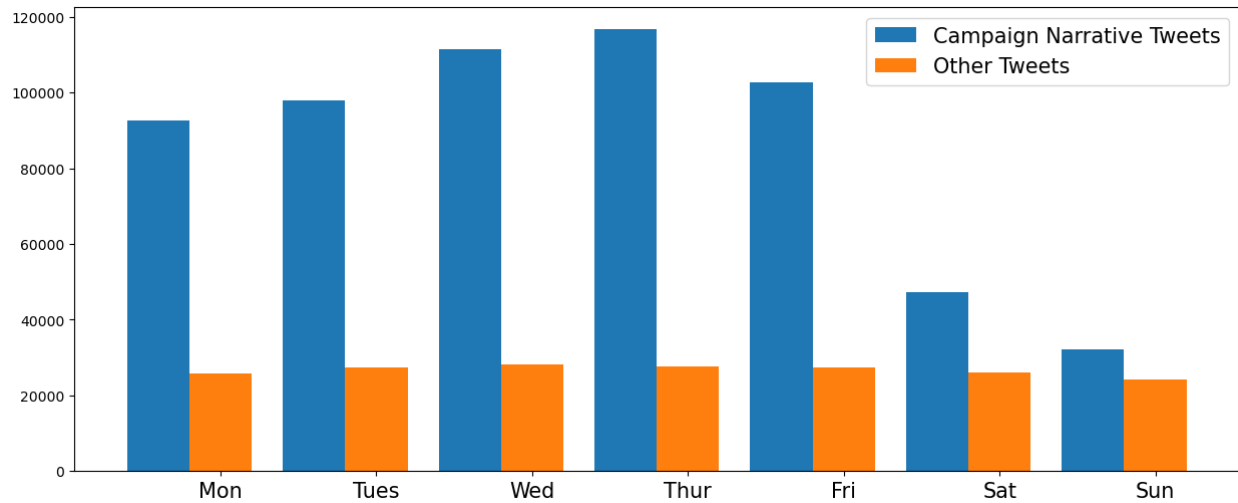
¹³ <https://www.cazadoresdefakenews.info/sin-rt-no-hay-paraiso-maquinaría-de-propaganda-en-twitter-de-nicolas-maduro/>

¹⁴ https://tigerprints.clemson.edu/mfh_ci_reports/9/



Figure 8 shows how those two sets of tweets are distributed across the days of the week. There is a very apparent differential pattern across these two sets of tweets. The tweets that we can attach to a campaign narrative are very disproportionately produced during the work week (Monday-Friday), while the other tweets are quite evenly produced throughout all seven days. This pattern suggests that the campaign tweets were organized around formal work.

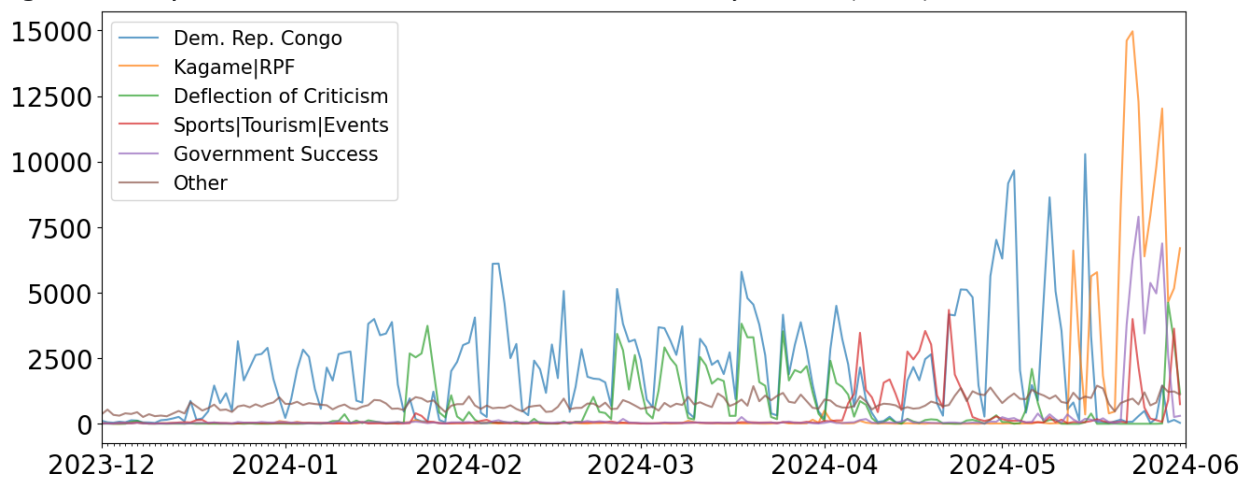
Figure 8: Distribution of Tweets across Days of Week by Inclusion of Theme-Related Hashtags.



Campaign Narratives and Promoted Accounts

The coordinated network has used each of these strategies to either promote/deflect several recent narratives since the start of 2024. To identify the target of the campaign over time, each of the top-200 hashtags used by campaign accounts throughout the past year was coded into five categories detailed in Figure 9.

Figure 9: Daily Tweets within the Coordinated Network by Theme (2024)



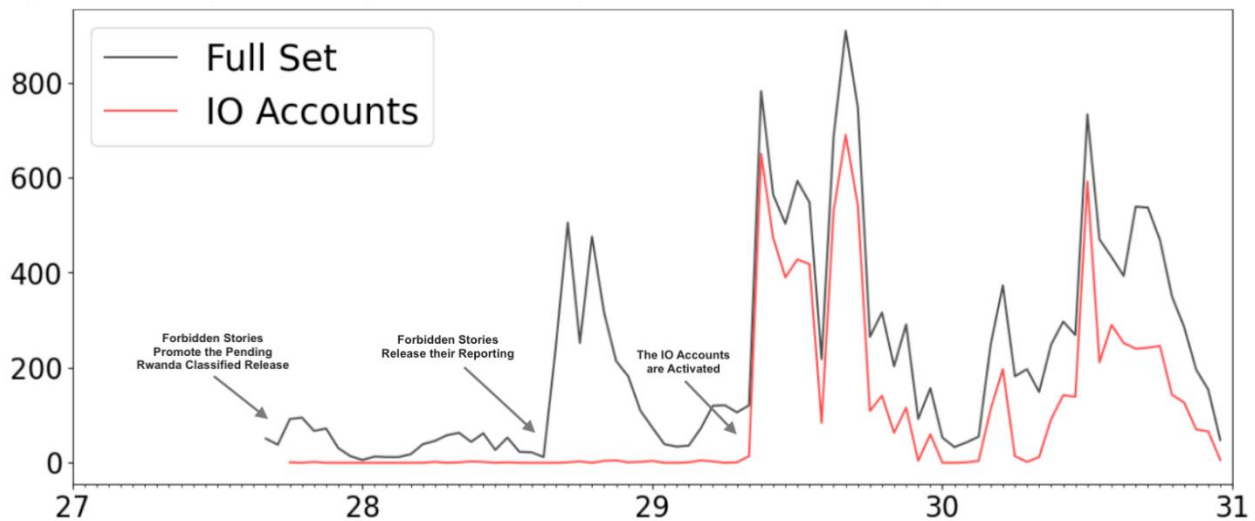


The accounts in the coordinated network also serve to promote the accounts and statements of other accounts both within and outside the network. The list of the top-20 accounts by amplification (retweet totals by in-network accounts) is presented in *Appendix C*.

Narrative: #RwandaClassified

Our initial investigation focused on identifying the characteristics of an apparent coordinated response to the release of the Rwanda Classified reporting led by Forbidden Stories. Based on this research, it became clear that accounts associated with the identified coordinated network were engaging with the hashtag to both discredit and deflect attention from the reporting. Our overview of this coordinated response details these efforts, their growing prominence over time, and the behavior exhibited by the network. *Figure 10* provides an initial overview of the proportion of the online discussion using the hashtag associated with the reporting (#RwandaClassified) linked to the accounts in the coordinated network (red), compared to all accounts which used the hashtag (black).

Figure 10: Hourly Posts using #RwandaClassified Published between May 27th and 31st, 2024



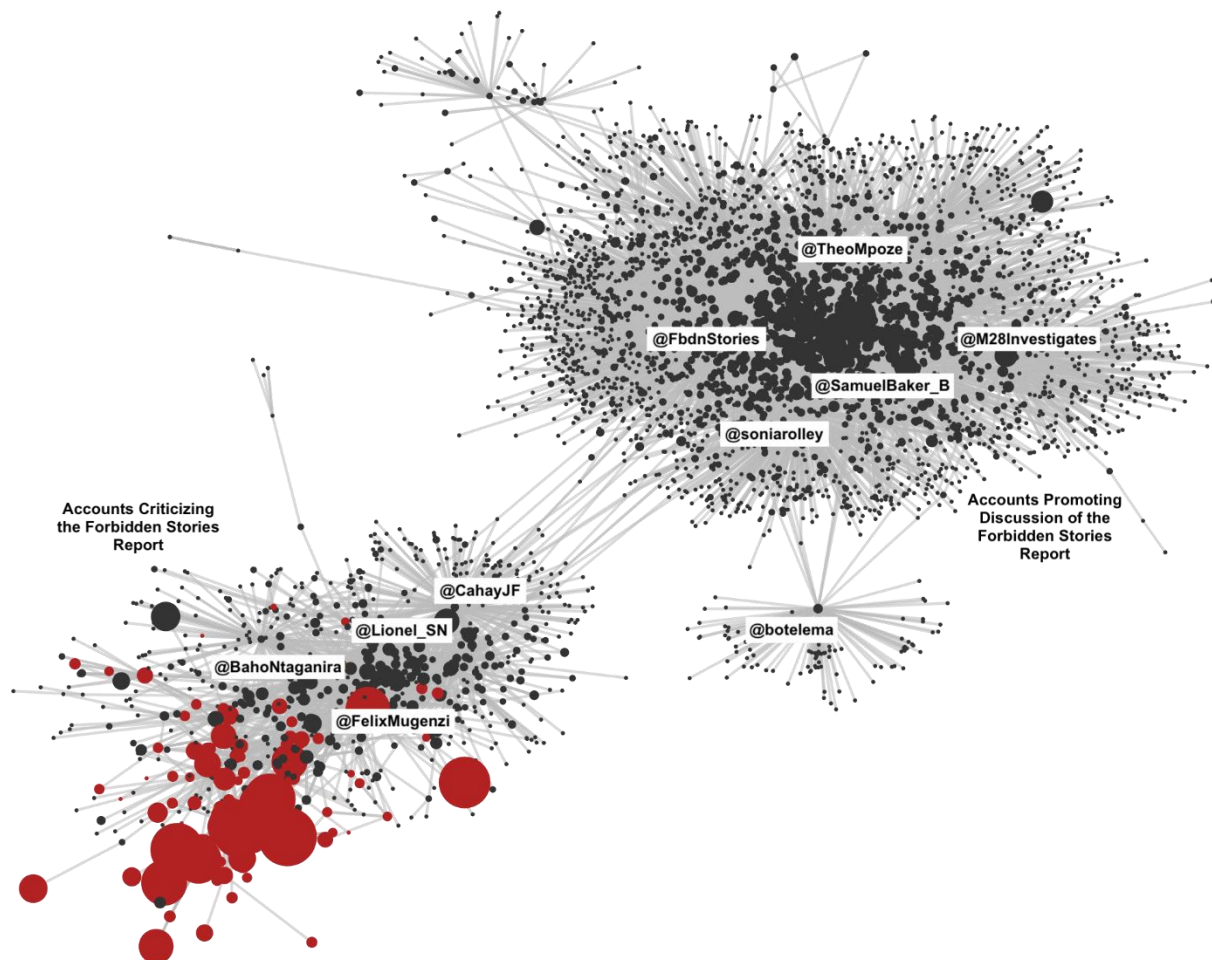
Though in total the network accounted for only 11% of the accounts involved in the discussion, these accounts produced 44% of all posts using #RwandaClassified in the study period. This proliferation of posts was also highly concentrated in time, with the initial release likely catching the operation off-guard (as evidenced by the lack of response in the first 24 hours by the coordinated accounts). Once the group had time to generate a list of text and image-based responses, their post rates rose precipitously. From that point forward, their posts made up the majority of the discussion.



Network Positioning and Behavior

To illustrate the influence of these accounts on the online discursive space, we can analyze the position of coordinated accounts within the wider #RwandaClassified narrative we are able to provide further insight regarding their behavior and intended impact.

Figure 11: Network Diagram of Accounts Contributing to in the #RwandaClassified Discussion



The diagram presented in *Figure 11* uses a force-directed algorithm to provide an overview of the connections between accounts in the coordinated network (red) and the remaining accounts (black) which used the hashtag “#RwandaClassified.”¹⁵ The size of each node denotes the total number of tweets each account (node) contributed to the overall #RwandaClassified discussion. In assessing the patterns underlying these interactions, we find the network to be visually split into two central communities: one promoting the report and one criticizing the report, which is highly intertwined with the coordinated network.

¹⁵ Each node represents an individual X/Twitter account, with the edges (grey) signifying interactions, such as retweets or mentions, between accounts using the hashtag. For visual clarity, the max node size was set to 10. As a result, the network understates the extent of the coordinated network’s control of the overall discussion. Finally, nodes are linked by retweet relationships between accounts (grey).



The labeled nodes represent the accounts with the greatest betweenness centrality ($n=10$). We include labels on these top accounts to illustrate the key participants and potential opinion leaders located in each cluster. In full, the smaller cluster (*bottom left*) is characterized by a tight-knit collection of highly active pro-Rwanda accounts, providing further evidence of partial coordination. As evidenced by the positioning of the prominent pro-Rwandan actors in this cluster, including influencer accounts @FelixMugenzi and @BahoNtaganira, the coordinated accounts appear to be embedded near these actors in the cluster which aimed to undermine the credibility of Forbidden Stories. In contrast, the larger cluster (*top right*), which contains Forbidden Stories (@FbdnStories) as well as several journalists involved in the project (including @SamuelBaker_B), is larger and more diffuse. As expected, this discussion in this cluster promoted the reporting.

Taken collectively, this visual depiction of the discussion surrounding the report reiterates the prominence of the coordinated network and the use of these accounts to “flood” the online discussion of the Rwanda Classified documents.

Narrative: Democratic Republic of the Congo

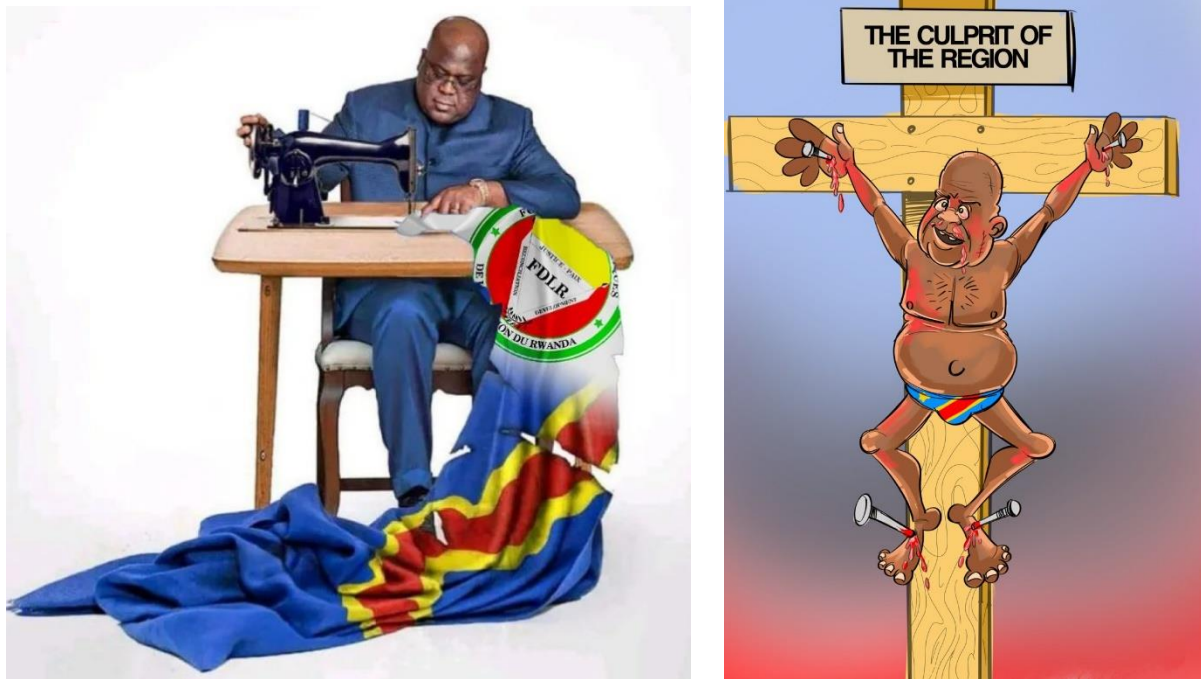
In tracing the origins of the coordinated accounts attacking/deflecting Forbidden Stories’ reporting, evidence points to the ongoing conflict between Rwandan-backed forces and local rebels in Eastern DRC.¹⁶ One version of this narrative included calls to action paired with hashtags calling for peace, primarily through unilateral disarmament. The hashtag #DisarmFDLR was the most commonly referenced term in the dataset, with the hashtag present in over 145,000 separate posts associated with the network.

A separate strategy aimed at influencing online discussion of the conflict involved the direct targeting of individuals associated with the DRC forces, including President Felix Tshisekedi. Of the top twenty most cited hashtags, seven directly implicated President Tshisekedi, including #TshisekediFDLR, #TshisekediCriminel, #TshisekediAgainstPeace, and #TshisekediIsKilling. These efforts not only involved repetitive AI-generated posts, but also generated images aimed at promoting Tshisekedi’s complicity.

¹⁶ <https://acleddata.com/2024/03/04/acleddata-brief-rwanda-backed-m23-rebels-advance-toward-goma-in-eastern-dr-congo/>



Figure 12: Several Edited and Artificial Images Depict Tshisekedi as Complicit in the Conflict



The use of generative AI (GAI) to produce/modify images was not solely used to produce targeted cartoons, but also to promote images purported to tie enemies of the state to the conflict. On more than one occasion, images from the DRC were modified to present opposition candidate Victoire Ingabire Umuhiza as a member of the rebel groups.

Finally, while a smaller proportion of total posts relate to the conflict, over 3,000 posts actively promoted the efforts and territorial gains of the M23 armed movement (#M23, #M23ARC, #ViveM23), which the government maintains it does not sponsor.



Figure 13: Use of Transposed Imaging – Modified Image (left) and Original (right)



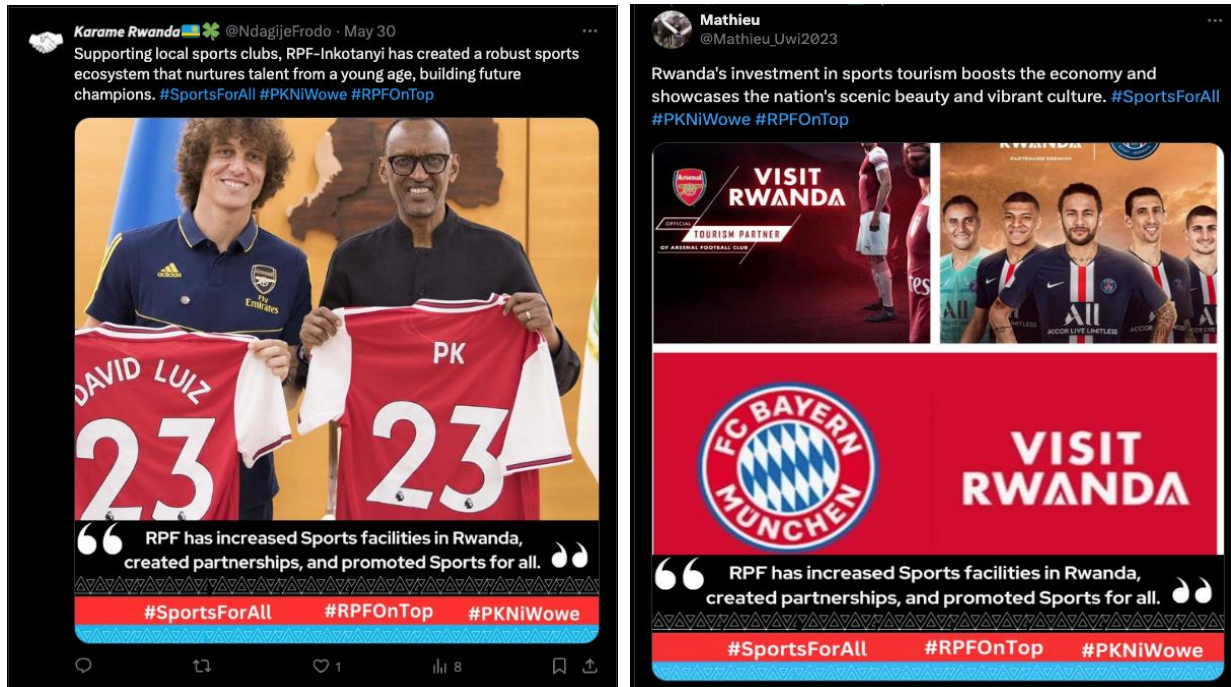
Narrative: Sportswashing

One of the most prominent pro-Rwanda topics of the network, as opposed to topics related to targeted criticism/harassment, is sports.

As illustrated in *Figure 14*, this multi-pronged effort involves not only the promotion of widely discussed partnerships with global football clubs such as Arsenal FC (#AFC2024 was promoted 230 times by the network), but also partnerships with global sporting federations and events such as the African Nations Championship (CAF), FIFA, the global Badminton Federation, and the National Basketball Association's (NBA) Basketball Africa League (BAL). The majority of this discussion incorporated the hashtag #SportsForAll (4,695 promotions) as a general reference alongside mentions of Kagame and the RPF.



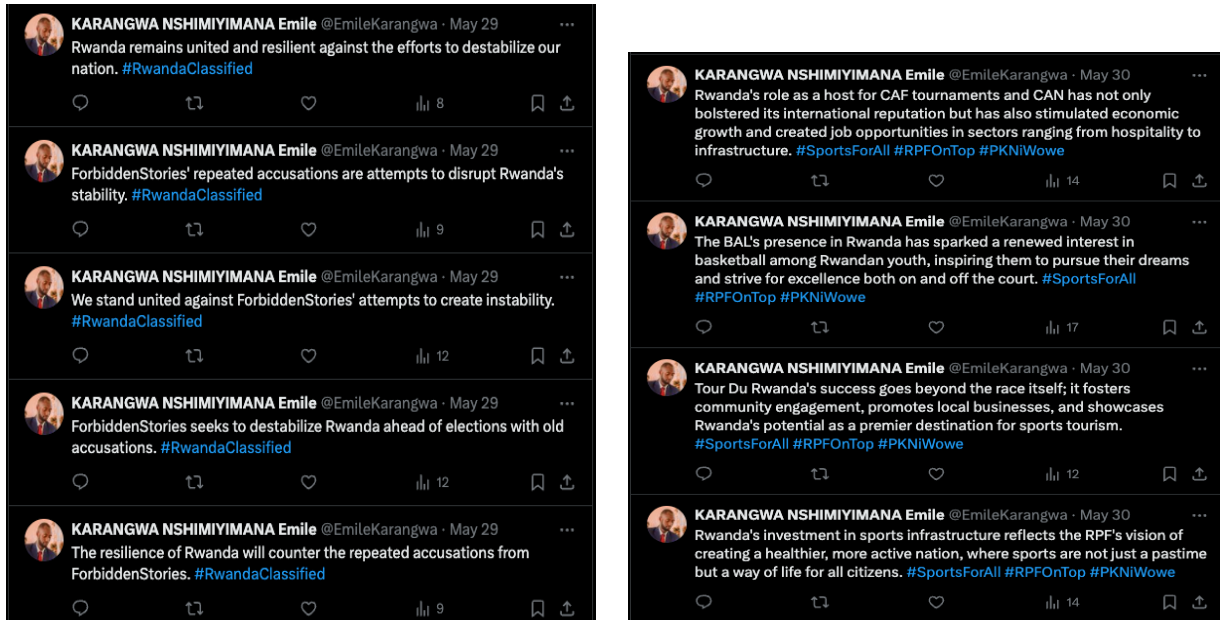
Figure 14: Use of Sports Events and Partnerships to Promote Paul Kagame and the RPF



BAL's recent final tournament, hosted in Kigali for the fourth successive season, was used directly as a foil to criticism of the state in the aftermath of the release of the Forbidden Stories report. Built on discussion involving the hashtag #BALinKigali, the same network of accounts used to harass journalists and deflect from the Rwanda Classified narrative seamlessly shifted to the generation of AI-altered discussions of the tournament, teams, and, of course, Rwanda's involvement with the league. In several instances, accounts shifted directly from attacks on Forbidden Stories to the promotion of BAL and related sporting achievements associated with the state (see Figure 15).



Figure 15: An example of a single user using generative text to promote Rwandan sporting collaborations alongside attacks on the credibility of Forbidden Stories reporting



Conclusion

This research details the recent advances in the sophistication of the tools used by pro-Rwandan actors to influence the prominence of narratives of interest to the state, including both those presenting the RPF/Kagame in good light and attacking critics of the state’s authoritarian actions. Unlike previous work on the influence of individual propagandists associated with the East African county, the evidence presented here illustrates a high level of AI-augmented coordination among pro-government social media accounts, suggesting an organized distributed disinformation effort to manipulate public perception and stifle dissent.

Our findings indicate that these operations have begun to increasingly integrate modern AI tools which may serve to further project the state’s sanitized image and empower efforts to harass critical voices. The voices include the brave reporters who worked with Forbidden Stories to produce their comprehensive report. The troll network we have identified, comprising at least 460 accounts, pushes narratives favorable to Kagame and the RPF, while discrediting critics and minimizing the visibility of unfavorable content. This includes several targets beyond those which we have highlighted in our report, including a focus on journalists like Michela Wrong, members of Human Rights Watch, and the UN Security Council. Amidst this criticism, by promoting positive narratives involving cultural events and prominent sports partnerships, this network has attempted to deflect attention from the persistence of state harassment and targeted assassinations.



While evidence of coordination is clear, we must note that this research is intended to be illustrative rather than exhaustive. The accounts and behaviors certainly do not represent the full extent of Rwanda's influence operations, nor do they explicitly tie the Rwandan state to the coordination of these accounts or their actions. As such, while this report provides insights building on the work of Forbidden Stories on the tactics used in this recent campaign, including evidence that a coordinated campaign is actively pushing pro-Rwanda content online, additional research is needed to fully scope the funding and organizational mechanisms sustaining the network and other coordinated actions.

In conclusion, by integrating novel forms of GAI and synthetic media into online campaigns to promote Rwandan propaganda, this report underscores the evolving nature of state-affiliated influence operations in Africa's authoritarian states. At the same time, while more sophisticated, the markers produced by this integration of coordinated action effectively eliminates the plausible deniability defense. There is clear evidence of coordination among pro-government accounts that cannot be written-off as the behavior of individual supporters acting independently. However, the methods used by the identified network will continue to develop, just as the perpetrating actors will likely continue to improve their efforts to mask their deployment. In order to counter the use of GAI, synthetic media, and other evolving technologies by authoritarian regimes, collaboration between digital platforms and policymakers is needed to uphold global democratic principles.



Appendices

Appendix A: Campaign Identification and Inclusion Criteria

Researchers began with a careful inspection of the social media information environment across several platforms and examined narratives which were likely targets for manipulation. Having identified clear signs of CIO using the TTPs outlined in previous analysis of Indian campaigns,¹⁷ we proceeded to identify specific accounts attributable to the campaign. The criteria for inclusion as part of the Indian-aligned CIO varied a bit by platform, as the sets of traces available vary. For an account to be included, it had to:

- Share either #RwandaClassified or #RPFOnTop between 5/13/2024 and 5/31/2024
- Share at least two suspicious hashtags 6 or more times between 5/13/2024 and 5/31/2024.¹⁸
- Have one or more series of **original** tweets with the following common characteristics:
 - All in the same language (French or English).
 - Any included media were pulled from a small common pool shared across accounts.
 - Similar length, usually 1-3 sentences.
 - Terminate with one or more hashtags.
- Not be excluded by manual inspection for posting substantial idiosyncratic content.

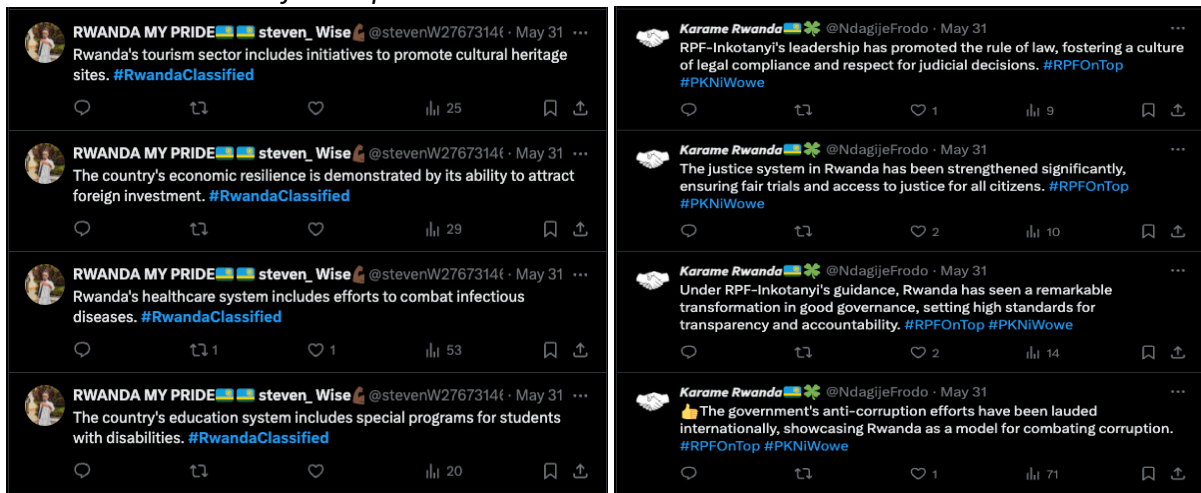
This process identified 464 accounts for inclusion on X/Twitter.

¹⁷ https://tigerprints.clemson.edu/mfh_ci_reports/9/

¹⁸ Hashtags were deemed suspicious if they were among the top-20 shared by accounts that shared #RwandaClassified or #RPFOnTop at least 10 times. See the full list in Appendix D.



Figure A1: Examples of inauthentic posts “flooding” the discussion of Rwanda, Paul Kagame, and the Rwanda Classified Report



The majority of the posts by the identified accounts follow this characteristic pattern of rapid publication. As seen in *Figure A1*, this involved the standard template of copied AI text affixed with whichever hashtag was preferred in the moment. In the aftermath of the release of the Forbidden Stories reporting, for instance, this took the form of two main targets – the reporting itself (#RwandaClassified) – and the promotion of both Paul Kagame (#PKNiWowe) and the Rwandan Patriotic Front (#RPFOnTop). Grounded in widely followed “flooding” tactics,¹⁹ these publications often did not address the content of the hashtag, but instead aimed to shift the conversation and disrupt the online discussion through mass production of government propaganda.

Original Data Pull: The original pull included all tweets that mentioned either #RwandaClassified or #RPFOnTop between 5/13/2024 and 5/31/2024. This included 149,246 messages from 5,537 accounts.

Suspicious Hashtags: Suspicious hashtags were defined as the top-20 most common hashtags in messages that were sent by accounts that sent at least 10 messages in the original data pull. These hashtags, with counts, were: ('#RPFOnTop', 127695), ('#PKNiWowe', 95100), ('#ThanksPK', 27609), ('#AgricultureAndFarming', 26879), ('#ECDInRwanda', 10919), ('#RwandaClassified', 10285), ('#TourismInRwanda', 6346), ('#SportsForAll', 4541), ('#PKNiWowe', 3946), ('#RPFInkotanyi', 3720), ('#RPFInkotanyis', 2391), ('#RwandaEducation', 2218), ('#Rwanda', 2060), ('#RPFOnTop', 1839), ('#ForbiddenStories', 1571), ('#RwandaAgriculture', 1356), ('#RPFs', 1047), ('#AgricultureInRwanda', 961), ('#ICTInnovation', 852), ('#Rwandas', 723)

¹⁹ <https://press.princeton.edu/books/hardcover/9780691178868/censored>



Winnowing of Initially Suspicious Accounts: 610 accounts in the original data pull shared 2 or more of the suspicious hashtags 6 or more times. Of those, 4 were suspended and 460 others qualified under the hand-cleaning criteria.

Final Data Collection: On 6/1/24, we pulled the smaller of the last year and the latest 10,000 messages from each qualifying account that we could still find on Twitter/X. This included 787,261 posts from 463 accounts. One account changed its name before we could collect it.



Appendix B: Hashtags, Counts, and Classification

Hashtag	Count	Classification
#DisarmFDLR	145,08	DRC
#TshisekediFDLR	137,34	DRC
#RPFOnTop	121,57	Kagame RPF
#PKNiWowe	95,138	Kagame RPF
#FDLRFatshi	80,294	DRC
#UNSecurityCouncil	61,152	Deflection
#FreeCongo	53,580	DRC
#TshisekediKilling	49,613	DRC
#Kwibuka30	37,219	Sports Tourism Events
#FDLRIsKilling	33,246	DRC
#ThanksPK	25,113	Kagame RPF
#AgricultureAndFarming	24,105	Governance
#TshisekediCriminel	20,734	DRC
#Tshisekedi	19,670	DRC
#TshisekediAgainstPeace	19,577	DRC
#EthnicCleansingInDRC	16,158	DRC
#EvaristeMarionnette	16,062	Deflection
#BilanDeTshisekedi	12,245	DRC
#Rwanda	11,759	Governance
#FDLR	10,291	DRC
#ECDInRwanda	10,181	Governance
#UNsecurityCouncil	9,140	Deflection
#DRC	9,134	DRC
#RwandaClassified	8,975	Deflection
#FatshiAgainstPeace	7,825	DRC
#TourismInRwanda	6,668	Sports Tourism Events
#Tshisekedis	6,581	DRC
#RDC	6,234	DRC
#FatshiFDLR	5,487	DRC
#SportsForAll	4,695	Sports Tourism Events
#HateSpeechDRC	4,587	Deflection
#DésarmFDLR	4,568	DRC
#VisitRwanda	4,547	Sports Tourism Events
#RacismIsWrong	3,847	Deflection
#FatshiFaible	3,791	DRC



#DR Congo	3,478	DRC
#BendeleEkweyaTe	3,456	DRC
#PKNiWOwe	3,428	Kagame RPF
#RPFInkotanyi	3,200	Kagame RPF
#RwandalsOpen	3,139	Sports Tourism Events
#RDCongo	2,785	DRC
#M23	2,684	DRC
#NettoyageEthniqueEnRDC	2,386	DRC
#Kagame	2,311	Kagame RPF
#Tshisekedilskilling	2,277	Sports Tourism Events
#InkuruYa30	2,149	Sports Tourism Events
#Ndayishimiye	2,139	Other
#RwandaEducation	2,116	Governance
#EthnicCleansing	2,112	Deflection
#RwandaWorks	2,100	Governance
#RPFInkotanyis	2,027	Kagame RPF
#FDLRWazalendo	1,984	DRC
#FARDC	1,823	DRC
#Bilandetshisekedi	1,778	DRC
#CrimesOfVictoireIngabire	1,730	Deflection
#PerezidaWacu	1,721	Kagame RPF
#Congo	1,705	DRC
#TshisekediCriminal	1,661	DRC
#ForbiddenStories	1,653	Deflection
#Freecongo	1,638	DRC
#TshisekeidFDLR	1,620	DRC
#Goma	1,596	DRC
#RPFonTop	1,592	DRC
#DesarmementFDLR	1,590	DRC
#RwOT	1,581	Other
#DésarmerFDLR	1,576	DRC
#SomeoneTellMatthew	1,532	DRC
#DisarmFDRL	1,464	DRC
#Fatshi	1,459	DRC
#StopFDLR	1,389	DRC
#PaulKagame	1,339	Kagame RPF
#UNSecurityCouncill	1,300	Deflection
#RBAAmakuru	1,282	Sports Tourism Events
#Wazalendo	1,277	DRC



#RwandaAgriculture	1,265	Governance
#FatshilsKilling	1,248	DRC
#NRCSymposium2024	1,240	Sports Tourism Events
#GOAfestival	1,229	Sports Tourism Events
#Paris	1,218	Other
#FARDC=	1,186	DRC
#UNReport	1,169	Deflection
#UNSecurityCouncil	1,144	Deflection
#IngabireVictoireFDLR	1,075	DRC
#Rwandas	1,056	Other
#StopGenocidIdeology	1,050	Other
#TshisekediAgainstCongolese	1,034	DRC
#FatshiFdlr	1,018	DRC
#Kwikuba30	1,017	Sports Tourism Events
#Burundi	987	Other
#DRCIsKilling	986	DRC
#RPFs	938	Kagame RPF
#FDLRskilling	911	DRC
#AgricultureInRwanda	891	Governance
#Umushyikirano2024	890	Sports Tourism Events
#kwibuka30	872	Sports Tourism Events
#freecongo	859	DRC
#InterahamweWazalendo	854	DRC
#ICTInnovation	819	Governance
#DesarmFDLR	815	DRC
#DRCs	781	DRC
#ChangementDeNarratif	778	DRC
#UNGA2023	767	Sports Tourism Events
#HateSpeech	741	Deflection
#Ndayishimiyes	720	Other
#GenocidIdeology	709	Other
#RPF	698	Kagame RPF
#UNGA78	682	Sports Tourism Events
#TunyweLess	681	Other
#Kwitalzina2023	674	Sports Tourism Events
#TshisekedilsKilling	657	DRC
#desarmFDLR	650	DRC
#HatespeechDRC	625	Deflection
#TwibukeTwiubaka	624	Sports Tourism Events



#FATSHIFAIBLE	619	DRC
#PK	606	Kagame RPF
#Kigali	593	Other
#Musanze	573	Governance
#EchecDeFatshi	555	DRC
#DecadeOfImpact	546	Governance
#GenocideAgainstTutsi	527	Sports Tourism Events
#FDLRFATSHI	521	DRC
#PKNiwowe	514	Kagame RPF
#Congolese	499	DRC
#GOA20years	499	Governance
#FDRLFatshi	499	DRC
#Kwitalzina	495	Sports Tourism Events
#TwinLakes	491	DRC
#AERG27	476	Sports Tourism Events
#RFPOntop	464	Kagame RPF
#WeForPK2024	462	Kagame RPF
#RDF	443	Kagame RPF
#RPFInkotanyi	441	Kagame RPF
#TraceAwardsRwanda2023	440	Sports Tourism Events
#Inkotanyi	426	Kagame RPF
#ConferenceInternationale	420	Sports Tourism Events
#SmartRwandaMasterPlan	418	Governance
#TukurilnyumaPK	414	Kagame RPF
#Kagame2024	410	Kagame RPF
#YouthInResearch	406	Governance
#IctInnovation	404	Governance
#TshisekediTue	402	DRC
#FatshiAzaNanu	401	DRC
#PK2024	400	Kagame RPF
#InternationalConference	396	Sports Tourism Events
#SADC	387	DRC
#OurPresident	384	Kagame RPF
#FactsOnRwanda	381	Governance
#GOAfest23	369	Sports Tourism Events
#TdRwanda2024	367	Sports Tourism Events
#AmayagaTuribuka	365	Sports Tourism Events
#Kwibohora29	362	Sports Tourism Events
#PRPFOntop	358	Kagame RPF



#RwandaDay2024	353	Sports Tourism Events
#UN	349	Deflection
#BiasedHRW	349	Deflection
#Tshisekediiskilling	347	DRC
#RPFOnTOP	346	Kagame RPF
#Kinshasa	337	DRC
#NYCMUSANZE	337	Governance
#InkotanyiNiUbuzima	334	Kagame RPF
#FPR	328	Kagame RPF
#EvaristeMarionnetteBurundi	326	Other
#TshisekediFDLRUN	325	DRC
#RBAnews	315	Sports Tourism Events
#FDRLWazalendo	312	DRC
#Tshisekediiskilling	309	DRC
#SomeOneTellMatthew	305	DRC
#Rwandaworks	297	Governance
#DRCAgainstPeace	297	DRC
#FDRs	296	DRC
#NyagatareYacu	287	Governance
#RwOX	285	Governance
#InternationalNursingDay2024	285	Sports Tourism Events
#NyagatareKulsonga24	281	Governance
#NewProfilePic	281	Other
#FDRL	272	DRC
#Kwibuka29	271	Sports Tourism Events
#Tutsi	270	Other
#M23ARC	268	DRC
#disarmFDLR	264	DRC
#Africa	263	Other
#WazalendoFARDCFDRLMAIMAIMERCENARIES	262	DRC
#Rwandan	261	Other
#Rebounce	259	Sports Tourism Events
#NeverAgain	258	Sports Tourism Events
#IdéologieGénocidaire	254	Sports Tourism Events
#Rwandans	250	Governance
#ConseildeSécuritédeIONU	246	DRC
#FDLRFatsh	246	DRC
#TshisekediContreCongolais	244	DRC
#NdayishimiyeVyaranse	242	Other



#VisitRwanda ^{rw}	242	Sports Tourism Events
#ViveM23	241	DRC
#CongoIeseM23ARC	241	DRC
#Fatshifailed	234	DRC
#wazalendo	231	DRC
#ACF2024	230	Sports Tourism Events
#GénocideDesTutsi	229	Sports Tourism Events
#AgricultureFarming	229	Governance

Appendix C: Amplified Accounts

The accounts in the coordinated network also promoted several prominent activists and party members. While not conclusive of a connection, this relationship further clarifies the aims of the network. *Table C1 presents the Top-20 accounts retweeted by the network.*

Account Handle	Network Retweets	In Campaign
1. <i>Mathieu_Uwi2023</i>	11331	TRUE
2. <i>JUSTIN_NSENGIMA</i>	9576	FALSE
3. <i>Claudeniyonku10</i>	4050	TRUE
4. <i>Fiona35690</i>	3843	TRUE
5. <i>DANIELHABIMANA9</i>	3754	TRUE
6. <i>UrugwiroVillage</i>	3665	FALSE
7. <i>KayinamuraJose8</i>	3123	FALSE
8. <i>FelixMugenzi</i>	2994	FALSE
9. <i>rbarwanda</i>	2960	FALSE
10. <i>LeonPMuhire</i>	2861	FALSE
11. <i>Amahoro66</i>	2581	TRUE
12. <i>BalisaYvette</i>	2504	TRUE
13. <i>NGillaine</i>	2469	TRUE
14. <i>IGIHE</i>	2464	FALSE
15. <i>FNdangije</i>	2307	FALSE
16. <i>Ndagije__sam</i>	2298	TRUE
17. <i>kura_jadox</i>	2284	TRUE
18. <i>harerimana_tito</i>	2089	FALSE
19. <i>RwandaisOpen</i>	2070	FALSE
20. <i>albcontact</i>	1884	FALSE



Appendix D: Additional Examples of AI-Use

In several cases, the posters failed to remove the topic headers prior to pushing out the automated posts. The resulting sets of posts provide a glimpse into the instructions provided to the LLM used to vary the published text.

Figure D1: Figures with Topic Headings Provide Further Detail of Targets

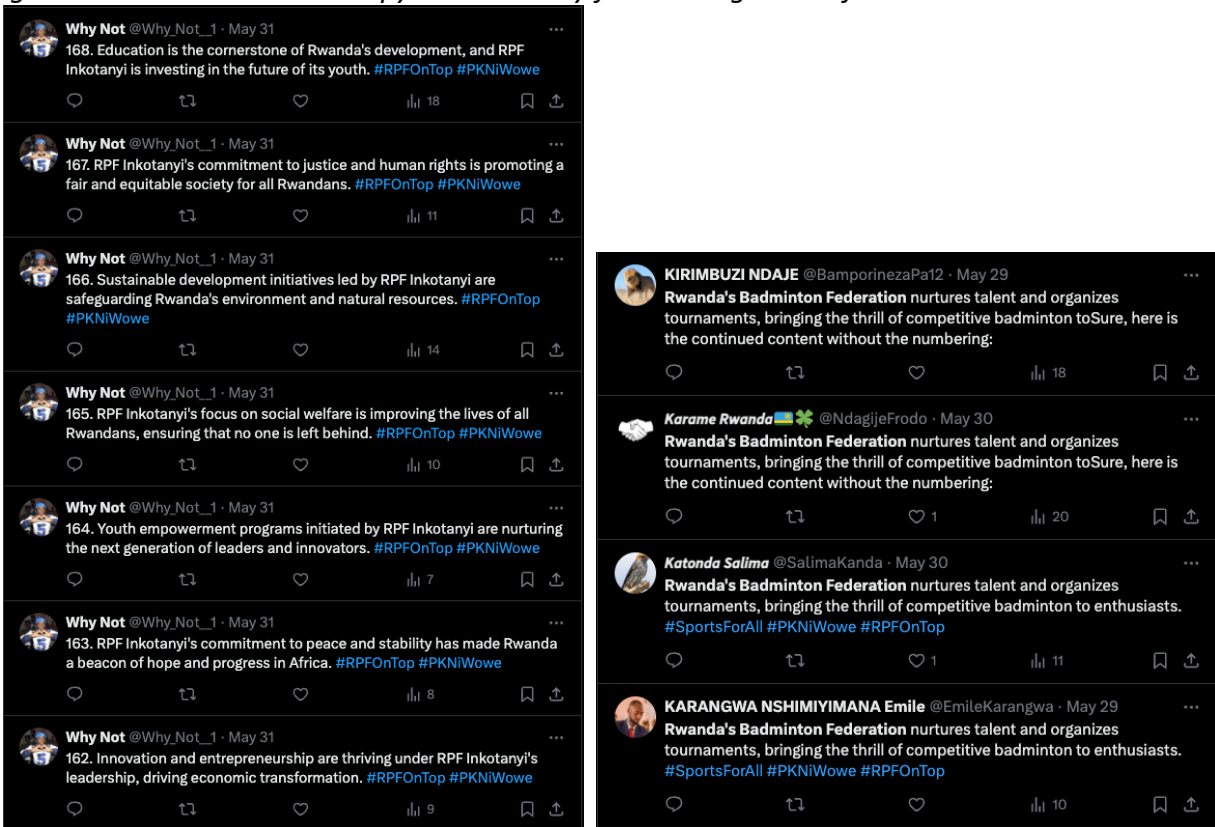


In other instances, we can see improper execution of copy-paste efforts from the dashboard of the LLM to X/Twitter. In one case, we can see the transposed text promoted by four accounts in the network. For the first account, the expected output is replaced by text from the LLM responding to a query by the user to have the numbering of the posts removed.

Providing further evidence of the use of number outputs, in other cases we find that posts were copied and reposted without first removing the numbering. When paired with posts like the one presented on the right in Figure B2, which contains undeleted instructions to the LLM to remove numbers from generated lists, we are given evidence of the link between the use of LLMs to generate text and the process for sharing/copying the output prior to posting online.



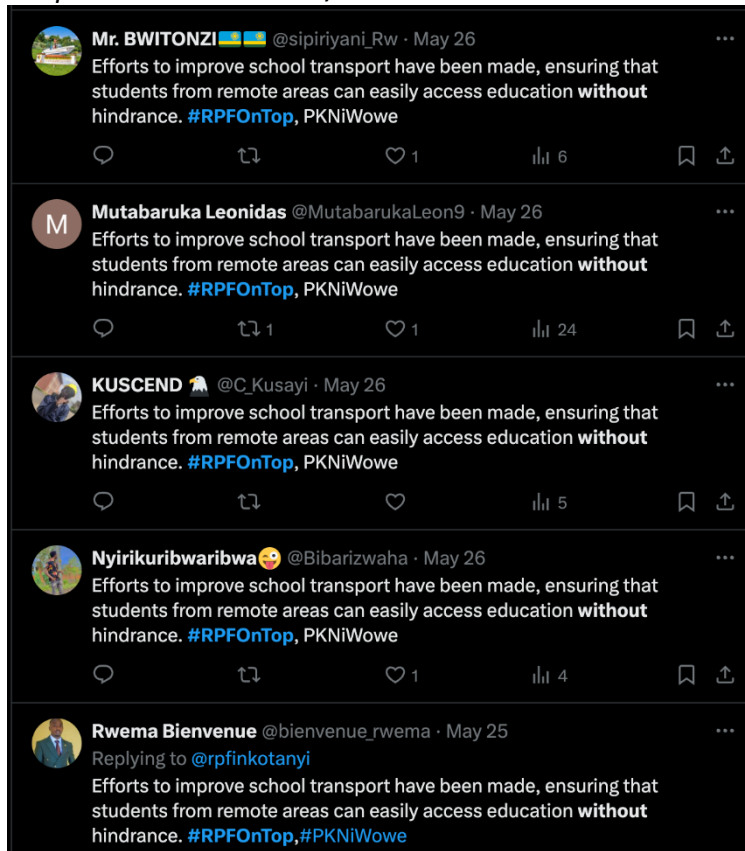
Figure D2: Numbered Posts Copy-Pasted Likely from a Larger Set of Generated Texts



Due in part to the post-generation process used by the network, mistakes in the use of copy-paste techniques carried over to their promotion on X/Twitter.



Figure D3: Mistakes Copied Across Accounts/Posts



Finally, though not present in the majority of accounts, several of the accounts had replaced profile images with posts generated using artificial intelligence. Some of these posts were available online when we dove further into the details of the accounts (see Figure B4).

Figure D4: Use of Fictional Profile Photos

