mozilla

INSIDE THE SHADOWY WORLD OF DISINFORMATION FOR HIRE IN KENYA

CONTENTS

Executive Summary

This research examines how Kenyan journalists, judges, and other members of civil society are facing coordinated disinformation and harassment campaigns on Twitter — and that Twitter is doing very little to stop it. The research provides a grim window into the booming and shadowy industry of Twitter influencers for political hire here in Kenya.

We conducted the research over the course of three months using tools like Sprinklr, Gephi and Trendinalia. We also interviewed influencers who participated in the campaigns, and collected a vast trove of screenshots, memes, and other evidence. In total, our research uncovered at least nine different disinformation campaigns consisting of over 23,000 tweets and 3,700 participating accounts.
Highlights of the investigation include:
Disinformation campaigns are a lucrative business. One interviewee revealed that disinformation influencers are paid roughly between $10 and $15 to participate in three campaigns per day. Payments are made directly to the influencers through the mobile money platform MPESA.

Twitter's trending algorithm is amplifying these campaigns — and Twitter is also placing ads amid all this misinformation. 8 of the 11 campaigns examined reached the trending section of Twitter. The campaigners we spoke to told us that this is their number one target, as it affords them the amplification they seek.

These campaigns run like a well-oiled machine. One of the influencers we spoke to explained a complex system of using Whatsapp groups to coordinate and synchronize tweets and messaging. Anonymous organizers uses these groups to send influencers cash, content and detailed instructions

These campaigns are increasingly targeting individuals. No longer focusing on just broad issues and events, disinformation campaigns are increasingly targeting individuals, like members of the Linda Katiba movement and the Kenyan judiciary. This work is also beginning to border on incitement, which is against Kenyan law

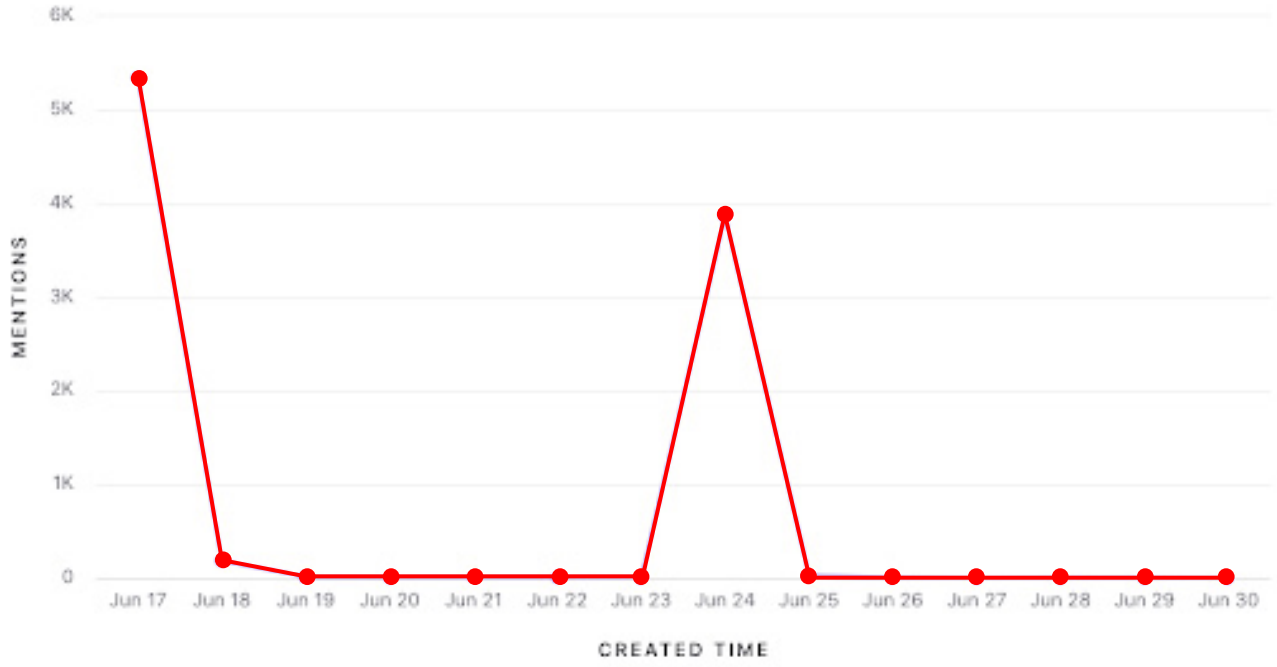
Verified accounts are complicit. One influencer we spoke to mentioned that the people who own coveted "blue check" accounts will often rent them out for disinformation campaigns. These verified accounts can improve the campaign's chances of trending. Says one interviewee: "The owner of the account usually receives a cut of the campaign loot"

These campaigns are chilling good-faith activism and making the platform harder to use for activists. Good-faith activists are now self-censoring on Twitter. One activist said she significantly reduced her Twitter activity thanks to all the trolling she experienced: "What was once a place where one could have some semblance of a healthy discussion on topics has now been completely poisoned."
Another activist mentioned that she had to spend a significant amount of effort countering narratives that were being seeded by disinformation campaigns.

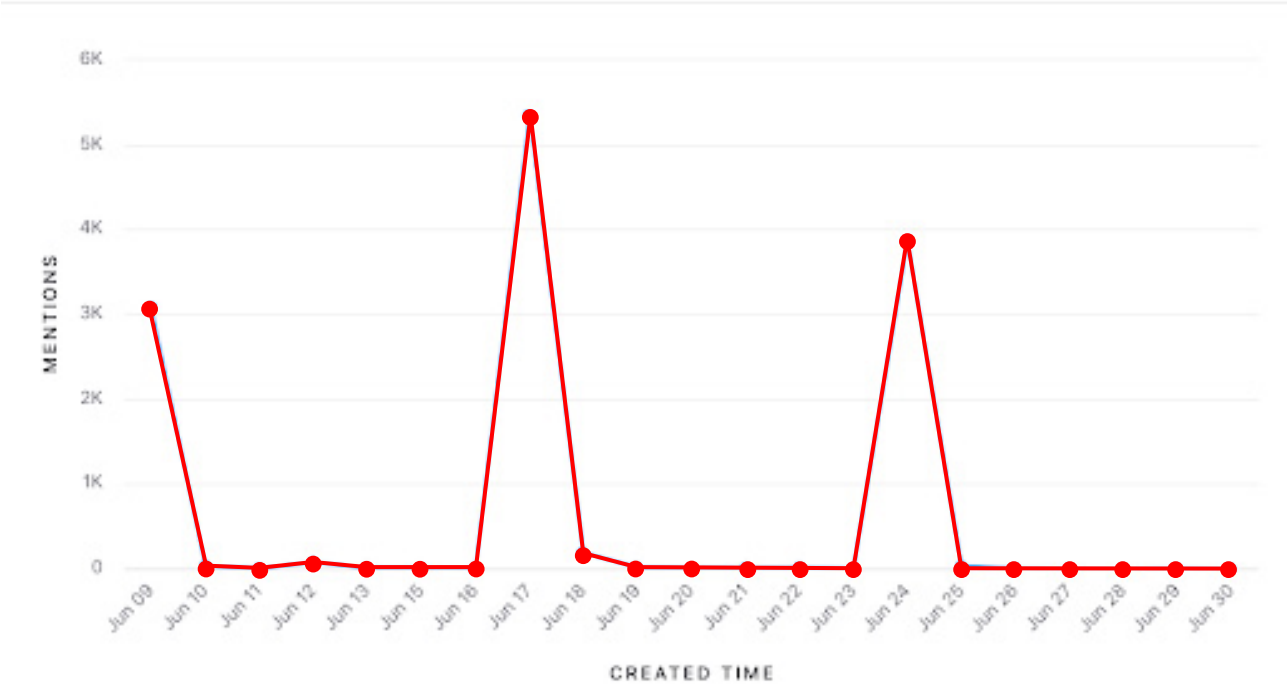
Introduction

Malicious, coordinated, inauthentic attacks which seek to silence members of civil society, muddy their reputations and stifle the reach of their messaging is a growing problem in Kenya. Twitter especially has been central to these operations due to the influence it has on the country's news cycle. The proliferation of social media platforms in Kenya carries the promise of a renewed definition of freedom of speech. Moreover, Twitter is a vital tool of expression for many Kenyan citizens, several of whom use it to hold their government to account and call out its failures. But, civil society members and journalists have increasingly come under attack thanks to disinformation operations in the country.

This report, a culmination of two months of data collection between May 1 2021 to June 30 2021, examines this problem in depth as relates to the Building Bridges Initiative, a constitutional review process that was going on in Kenya. We gathered data on these attacks by identifying and analyzing the hashtags the perpetrators used on Twitter with the aid of Twint, Sprinklr, and Trendinalia. The criteria of identification of a campaign, was done by assessing the content within the hashtag to check if it contains malicious content targeted towards activists and members of the judiciary; synchronized timestamps of publishing within the metadata and a lack of content in subsequent days, indicating one very sharp burst of activity and then fizzling out. In total, through Sprinklr, which has access to Twitter's full historical archive, we found that 23,606 tweets and retweets were released by 3,742 accounts participating in the 11 hashtags that we uncovered. Additionally, we were able to obtain 15,350 of these tweets using the Twint package on Github to carry out additional analysis of the content within the tweets.



[Tweet volumes for 2 of the hashtags we identified Source:Sprinklr]



[Tweet volumes for 2 of the hashtags we identified Source:Sprinklr]

Additionally, we gained access to the inner workings of how these campaigns are carried out. This was through interviews with influencers who participated in them. The individuals also provided us with evidence such as screenshots from their internal communications, and details about how they are organized which we were able to review.

Examining the campaigns provide a window into the booming, shadowy industry of Twitter influencers for political hire in Kenya. Many of the accounts and individuals involved promote brands, causes and political ideologies without disclosing that they are part of paid campaigns. This is a lucrative, well-oiled machine with very clear targets and as a result it is chilling good faith activism. Twitter's features are being exploited to achieve the goals of these campaigns. Its trending algorithm is amplifying these campaigns and accounts verified by the platform are complicit in leading these attacks. The goal of these campaigns is to exhaust critical thinking and poison the information environment by annihilating truth.

1. How disinformation is spread

The strategies and tactics used in Twitter disinformation campaigns within Kenya.

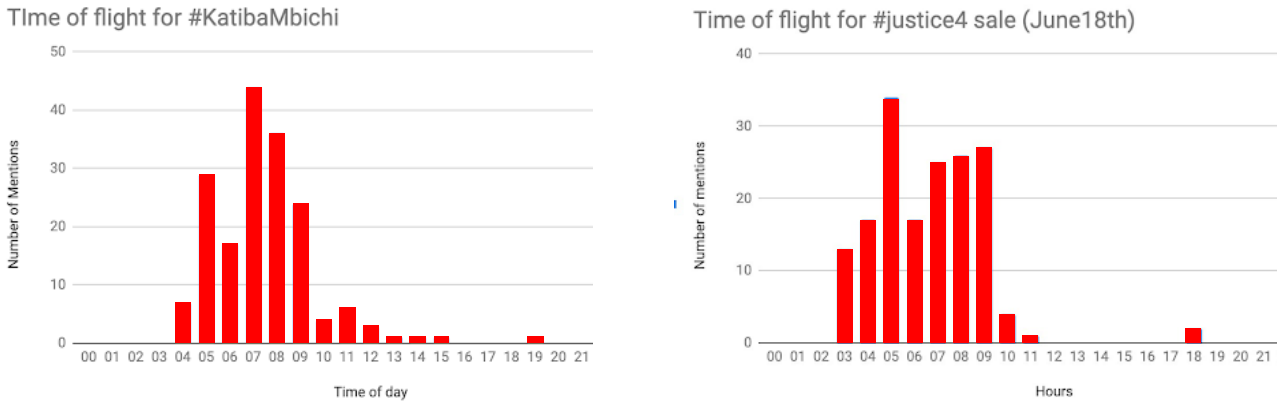
Overall, Twitter disinformation attacks within Kenya successfully use a large number of accounts to tweet using predetermined hashtags. This ensures that their narrative trends on Twitter and gains significant visibility.

Our investigation shows that this particular network's main goal is to sway public opinion during high pressure political instances such as by-elections and protests. However, there has been a specific interest by these accounts in influencing the constitutional review process (known as the Building Bridges Initiative, or BBI in short) that the country has been going through.

The Twitter campaigns were being used to inauthentically promote the BBI to citizens and attack prominent civil society activists that voiced their opposition to it. Furthermore, they sought to discredit CSOs and activists, portraying them as villains who were being funded by Kenya's Deputy President William Ruto. Ruto is known to oppose the BBI process.

[Screenshots of content used to target CSO members]

In our interviews with one of the influencers, they informed us of the agile tactics they use to organize and avoid detection. For example, when it's time to carry out the campaign the influencers would be added to a Whatsapp group. Here, they received direction about what to post, the hashtags to use, which tweets to engage with and who to target. Synchronizing the tweets was also incredibly important for them. It's what enables them to achieve their goal of trending on Twitter and gain amplification.



[Time of posting of content for the hashtags we tracked showed signs of synchronized activity. A sharp burst of activity then a flat line]

Recently, however, we were informed that as the work they do has gotten much more sensitive, they are increasingly identifying and targeting individuals. Their work is also beginning to border on incitement which is against Kenyan Law. As a result, they are being sent briefs individually so as to cover their tracks better.

"We stopped joining groups sometimes to make it less easier for us to be found out and keep things from leaking. We also really don't know who specifically we're working for sometimes. Nowadays the organizer just sends us cash, content and the instructions individually and tells us to post."
—Disinformation influencer

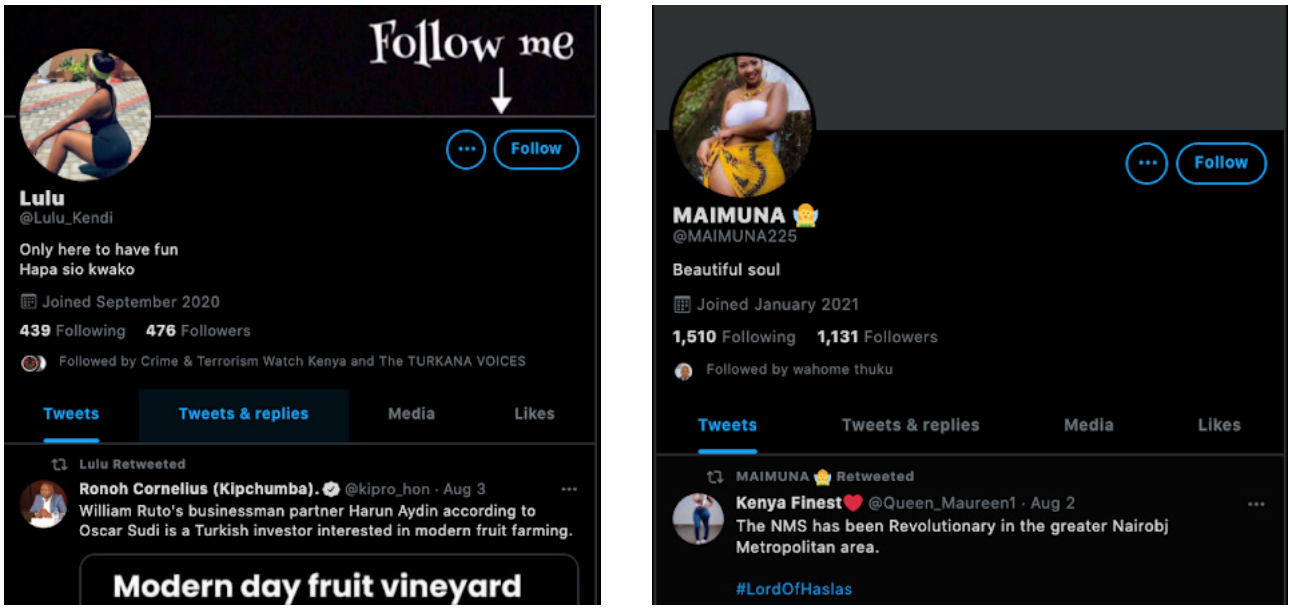
As the influencer suggests, there is also money to be made in attacking civil society. They revealed to us that those participating in the exercise are paid roughly between $10 and $15 to participate in three campaigns per day. Each campaign execution involves tweeting about the hashtags of the day until it appears on the trending section of Twitter. Additionally, some individuals have managed to reach retainer level and get paid about $250 per month. Their job is to make sure the campaigns are executed on a day-by-day basis with different hashtags.

In both cases, payments are made directly to the influencers through the mobile money platform MPESA. The individuals we spoke to, however, refused to reveal to us who was paying them when probed about it.

"Sometimes the money comes before the campaign and sometimes it comes after the campaign. They're usually careful not to delay for fear of us exposing them on Twitter for lack of payment."
—Disinformation influencer

Inauthentic amplification also served the goal of getting the hashtags to trend on Twitter. It improved the odds that people saw them. Another influencer, who participated in the campaigns (but did not wish to be named), also informed us that many of the tweets in these hashtags were amplified by other sock puppet accounts. The aim is to trick people into thinking that these opinions are popular. This type of inauthentic engagement borrows from an age-old strategy. It is the online equivalent of many Kenyan politicians' long standing tradition of paying crowds to show up at political rallies to create the appearance of popularity. Retweets, however, are much easier to obtain.

Many of the accounts we examined appear to give an aura of authenticity, but in reality they are not authentic. Simply looking at their date of creation won't give you a hint as to their purpose. We had to dig deeper. The profile pictures and content of some of the accounts gave us the answers we were looking for. A common tactic these accounts utilize is using suggestive pictures of women to bait men into following them, or at least pay attention. In terms of content, many of these accounts tweeted off the same hashtags for days on end and will constantly retweet a specific set of accounts.



Examples of typical accounts used in these campaigns

Another tactic is to use pictures of famous individuals to create accounts. One account impersonated Omar Lali, a man who became infamous in Kenya after allegedly killing his wife on Kenya's coast. He denied any association with the account when we contacted him.

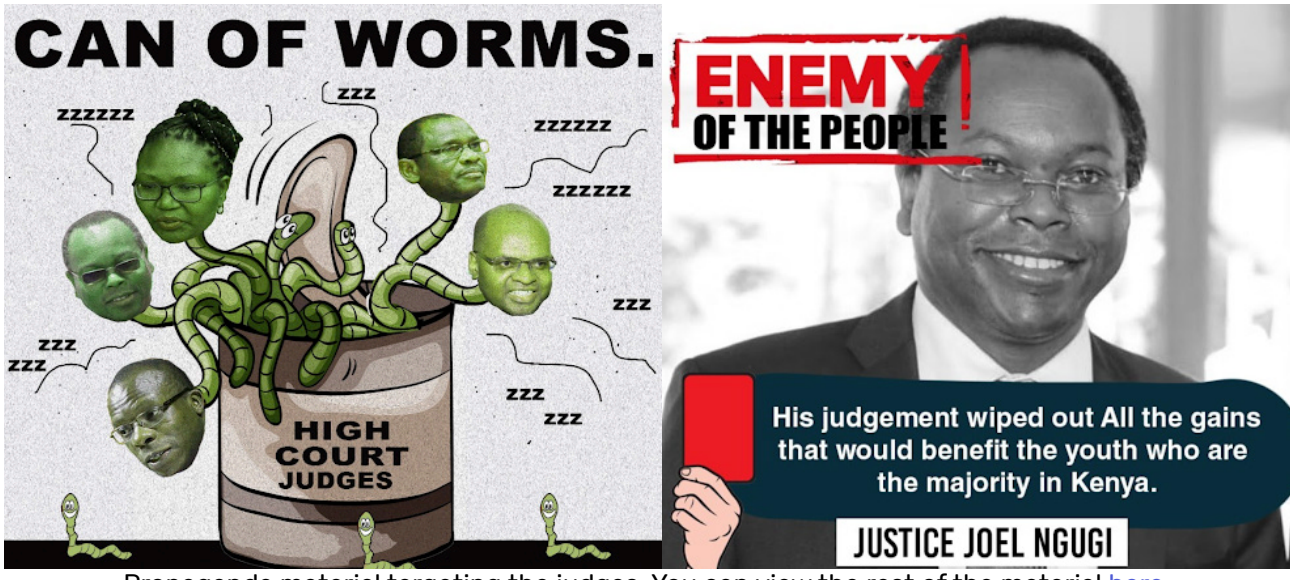
2. Who disinformation targets

Kenyan journalists, judges, and activists are the most frequent victims

Prominent activists under the Linda Katiba movement voiced their opposition towards the BBI and petitioned in court against it. As a result, they faced acute attacks against them in early May which had the specific goal of discrediting their campaign. Jerotich Seii is one of the members of the Linda Katiba campaign who was targeted. She said in an interview with us that she had to spend significant time trying to prove that her activism efforts were genuine and not a front for someone else because of the attacks against her.

"The disinformation attacks against me focussed on painting me as someone with ulterior motives who isn't interested in the welfare of Kenyans. I had to spend a good chunk of my time defending my position as someone who is actually a patriot who does what they do out of love for their country."
—Jerotich Seii, Kenyan activist

The Kenyan High court struck down the BBI on May 14th on the grounds that it was an unconstitutional initiative, a ruling which exacerbated the strained relationship between Kenya's Judiciary and its executive. Since then, wave after wave of disinformation attacks were launched in a bid to discredit the independence of the judiciary and question the accuracy of their decision. It is in these anti-judiciary campaigns where the aesthetic of the content within the campaigns began to shine through. They often employed the use of caricatures (which mimic the style of newspaper editorial cartoons) and memes with repetition of particular templates using the likeness of the judges.
All this serves the aim of making the content more palatable and shareable.



Propaganda material targeting the judges. You can view the rest of the material here.

One example of an anti-judiciary campaign is #Weshallrevisit, a hashtag that alludes to Uhuru Kenyatta's warning shot at the Judiciary after they nullified his election in 2017. It was used to propagate the idea that the Judges had a personal vendetta against the president and were seeking revenge against him for budget cuts they had experienced in his second term. In another instance, the hashtag #Justiceforsale was used by the trolls to push the idea that Kenya's courts are corrupt and they were bribed to make their decision.

3. What impact disinformation has

Censorship and the chilling of free expression are rampant

Data we gathered from Trendinalia (which collects data on Twitter trends in Kenya) shows that sufficient amplification was gained for some of the hashtags we identified to become trending topics. Part of how this amplification was achieved was through the use of verified accounts to also tweet using the hashtags. One influencer we spoke to mentioned that the people who run verified accounts who take part in such kinds of campaigns will often rent them out. Verified accounts are used to improve the campaign's chances of trending.

"The owner of the account usually receives a cut of the campaign loot from the person that rented it from them once it's over."
—Disinformation influencer

We however did try to reach out to the blue check accounts they identified in the campaigns to verify if indeed these claims were true. All of them denied that their accounts were rented. Reiterating that they were publishing content with the hashtags in their own capacity.

The demand for this service from the political class in Kenya is very strong. We counted at least 31 artificial political hashtags, including the ones involving BBI, between May and June. That translates to at least one disinformation campaign every two days. Many Kenyans that frequent Twitter therefore, will see these manipulated trending topics on close to a daily basis. There is little evidence that such kinds of operations actually sway people's opinions. However, they do have an effect on how Twitter users interact with their information environment. The goal of such operations is to overwhelm. It is to create an environment where nobody knows what is true or false anymore. Their job is to exhaust critical thinking and annihilate truth.

All this is leading to self censorship by some of the activists who use the platform. Many feel it's pointless to use a platform that can't deliver any meaningful engagement. One activist we spoke to said she significantly reduced her Twitter activity thanks to all the trolling she experienced.

"What was once a place where one could have some semblance of a healthy discussion on topics has now been completely poisoned. It's impossible to have healthy debate around many issues. Dissenting voices will often find that an entire army of bots will sit in your mentions should you voice your opinions."
—Kenyan activist

4. Where does this leave Twitter?

The platform is failing Kenyans — and Africans more broadly

To many Kenyans, Twitter matters. The platform is a very important avenue of expression and information for them. It is also an important avenue of citizen accountability and #KOT (Kenyans on Twitter) is one of Africa's loudest and most lively internet communities. Whereas Twitter was once considered a partner to Kenyan citizens by enabling freedom of expression, some of its features are being exploited for authoritarian purposes. Political actors are using it to try to control political narratives by poisoning the information environment and harassing dissenting voices. The hashtag campaigns we have highlighted are clear examples of this.

Of key contention here is that malicious actors can use Twitter to create sock puppet accounts en-masse; get some of those accounts verified or use someone else's verified account; create malicious content; generate fake engagement; and eventually have Twitter boost their content for them through its trending algorithm to millions of Kenyans. If they are successful with all this, they'll find normal people participating in their campaigns, too.

Twitter's trending algorithm specifically has been a key target for the malicious actors, as it often serves as an on-ramp for users who are trying to find information on the platform. Sources we spoke to mentioned that it is the primary KPI by which most of the campaigns they run are judged. They admitted that without it, their jobs basically wouldn't exist.

"The main goal is to go trending on Twitter. I'm not sure what our jobs would look like without that target."
—Disinformation influencer

When asked about this, Twitter carried out an investigation of its own and were able to take action on just over 100 that they found to be in violation of their Platform Manipulation and Spam policy.

"After investigation, our teams took action on just over 100 Twitter accounts operating in Kenya which we found had engaged in violations of our Platform Manipulation and Spam Policy. While we weren't able to independently confirm the tweet-for-pay activity described in your report, we could confirm the presence of at least one network of coordinated accounts — which appeared to link back to an earlier set of enforcements against similar activity, carried out by our teams in 2020."

This is not the first time that such kind of engagement on the platform is coming under scrutiny. In Nigeria, Buzzfeed news reported on how accounts were being used to drum up support for a Colombian drug dealer. Additionally, in 2018, a report by Portland Africa showed that bots were becoming increasingly prevalent in trying to sway opinion in African elections. In 10 elections across nine African countries between 2017 and 2018, bots were seen as important vehicles for spreading misinformation and furthering negative narratives about major issues. Kenya itself underwent an extensive disinformation campaign at the hands of Cambridge Analytica. Cambridge Analytica may be no more, but Kenya's ruling class seems to have learned something from them. Their campaign in 2017 contributed to a poisonous electoral environment. It was an election season that saw sections of Kenyans cheering as police brutalized protestors during post election skirmishes in the country.

Another example is in Uganda. Both Twitter and Facebook suspended a network of inauthentic accounts for engaging in a coordinated campaign to promote Yoweri Museveni in the elections. The ensuing fight over the deletion of the accounts would lead to both platforms' suspension in the country. Twitter was reinstated a few months later. Facebook to this day is still banned in Uganda.

Twitter clearly has had this problem for a long time, but won't fix the issues that lead to it arising. All the evidence we've pointed to in this investigation is not a new phenomenon to executives at Twitter. The trending algorithm in particular is a big part of how disinformation campaigns and attacks are run. It's considered to be a big part of how Twitter works but it is doing more harm than good. It selects and highlights content without paying attention towards the potential the content might cause. Many of these campaigns would be significantly hampered if it didn't exist.

One solution could be that Twitter's moderation team should pay close attention to its trending section and edit it out whenever a trending topic is seen to majorly harbor malicious content. On the other hand, activists such as Sleeping Giants have repeatedly called for Twitter to "untrend" itself. This could either be by removing the feature completely or by disabling the feature during critical times, such as elections. From the information we've gathered speaking to some of these disinformation purveyors, it would greatly disrupt their operations.

Arguably, Twitter does have an incentive to fix this. It sells ads for "promoted trends" and "promoted tweets" within the feeds of hashtags on its trending topics section to businesses.

This puts it squarely in this mess, as it means they put ads from brands in the middle of inauthentic and dangerous content. Twitter profits from this harmful activity yet advertisers are very sensitive to brand safe environments. Ad Dynamo, the agency that sells Twitter Ads in Kenya, currently offers promoted trends for $3,500 per day within the country.

The overall message this sends is that it's okay to sow hate on the platform, so long as its owners can place ads next to the content and make money from it.

Twitter's current whack-a-mole approach to solving the problem isn't working. It clearly shows that Twitter isn't as committed as it should be towards solving this problem. Sustainable solutions that take into account the mechanics and incentives that drive these campaigns must be sought.

As Kenya heads to its elections in 2022, the activity we have highlighted in the pro BBI campaigns is likely to be heightened in terms of frequency and their use of violent, targeted rhetoric. Politicians' playbook in the next elections will be no different from what they've done in the past. They will seek to divide Kenyans along tribal lines. Demand for these services will therefore increase and many political parties will seek them out as part of their campaign strategies. Tensions in Kenya are indeed rising and Twitter needs to pay attention to how their platform is serving political motives in the country. If they don't however, it's quite possible Twitter could have blood on its hands for what they allowed to fester within their platform.

Acknowledgements

This report was written by Mozilla Tech and Society Fellow Odanga Madung with Brian Obilo as the contributing author (Also a Mozilla Tech and Society fellow). The views of this paper do not reflect the views of Mozilla or the tech and Society Fellowship program.

We also want to thank Kevin Zawaki and Solana Larsen for their contributions in editing this report. The views in this paper do not necessarily reflect theirs, nor that of their employers.

Annex

Summary of All Hashtags that were measured can be found here
All the memes that have been used can be found here

Thank You

foundation.mozilla.org