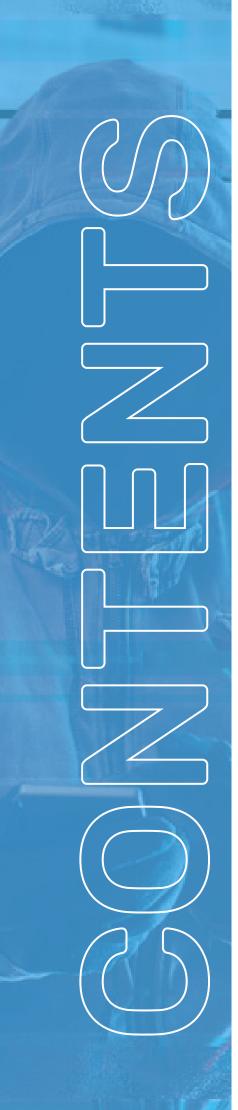


# HOW TO MANIPULATE TWITTER AND INFLUENCE PEOPLE: **PROPAGANDA AND THE PANDORA PAPERS IN KENYA**



Kenya's disinformation industry tried to discredit the Pandora Papers to distort public opinion and protect the country's elite. Investigating this activity showed us that its disinformation problem is much deeper than we thought.





#### foundation.mozilla.org

Executive Summary	4
Introduction	5
The strategy: Anti-Pandora Papers 'astroturfing'	6
The context: Another entry in a flourishing industry	12
The Upshot: Twitter is complicit in harming democracy	14

Acknowledgements

16

#### Executive Summary

In October 2021, the Pandora Papers were released to much ado, revealing several world leaders' secret offshore finances.

Despite outgoing Kenyan President Uhuru Kenyatta's inclusion in the Papers, legacy media in Kenya provided only cursory coverage – creating a vacuum that social media was quick to fill. But as our <u>past research</u> has demonstrated, social media – and especially Twitter – in Kenya is often leveraged to deceive and manipulate around political issues, rather than foster open discussion. This phenomenon is fueled by the country's flourishing disinformation industry: a shadowy and sophisticated network of fake accounts, artificial hashtags, and well-paid disinformation influencers.

Indeed, while public sentiment on Twitter was at first critical of Kenyatta, an alternative sentiment quickly emerged, supporting the president and his offshore accounts. Our research shows that some of the tweets promoting this alternative sentiment were outright false. But other tweets were more nuanced: technically true, but clearly inauthentic and coordinated to feign public support. In short: Kenyan Twitter is awash in Pandora Paper <u>astroturfing</u>.

This research report unpacks how this astroturf campaign unfolded. It includes quantitative analysis using data from Twitter's firehose. It also includes qualitative insight, like interviews with disinformation influencers. This report also grapples with the bigger issues at play: How Kenya is home to a flourishing online disinformation industry. How Twitter is unwilling – or unable – to address this problem. And how all this threatens a democratic society in Kenya.

#### Introduction

On 3rd October 2021, the International Consortium of Investigative Journalists (ICIJ) published the Pandora Papers, a massive investigation into how several public figures use offshore tax havens to hide assets worth billions of dollars. Almost immediately, the story provoked a sharp reaction from the Kenyan public. This is because the family of Kenya's outgoing president, Uhuru Kenyatta, was among those whose offshore financial engagements were exposed.

Kenya has indeed had a long history of political families stashing massive amounts of money abroad. This news, however, came at an especially charged time. Political tensions in Kenya are at a high as the country nears the 2022 elections. The Pandora Papers also showed up at a time when the country's political scene was fresh from a heated debate on wealth declaration – one ignited by the government's public revelation of the wealth belonging to the neo-opposition leader, Kenyatta's deputy, William Ruto.

For many Kenyans, the Pandora Papers were the perfect summation of Uhuru Kenyatta's regime: One that has been plagued by corruption scandals, soaring inequality, debt and skyrocketing commodity prices. Therefore, through social media, they let their frustrations be known loud and clear. Very quickly, the hashtag, #Pandorapapers, went trending on Twitter's Kenyan ecosystem and remained a topic of conversation for the entirety of that week. However, peculiarly absent from the conversation within the first 48 hours of the Pandora revelations, was coverage from local media outlets in Kenya.

While the rest of the world and Kenyans dissected the implications of ICIJ's revelations, Kenya's mainstream media outlets were peculiarly silent. The silence didn't go unnoticed. Alongside #Pandorapapers going trending, the term <u>"Githeri Media"</u> also surfaced on Twitter's trending section. Kenyans started calling out the media demanding that they cover this issue.

The temporary absence of critical mainstream media coverage of the Pandora Papers created an information vacuum that became fertile ground for a disinformation campaign aiming to pacify the scandal following the leaks.With the government and the president under pressure, due to soaring online outrage, a counternarrative operation was mounted – and found a strong ally in Twitter. Twitter's trending algorithm amplified these campaigns to millions of Kenyans who were using the platform to find information about the leaks and engage in discussions about them. As a result, a distorted perspective began to gain momentum – one where Kenyans appeared outraged not by the Pandora Paper's damning findings, but by their implication that Uhuru Kenyatta is guilty of wrongdoing.

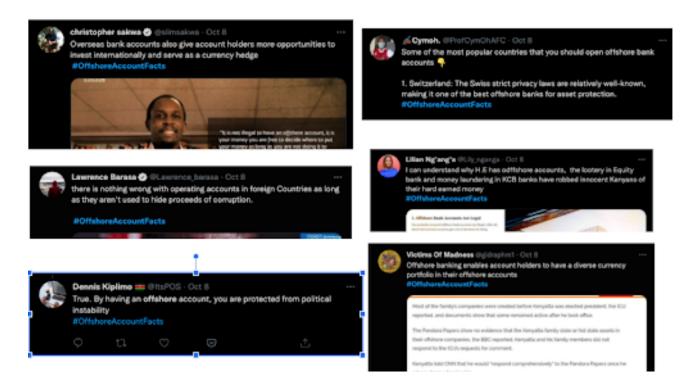
We therefore decided to follow up on our initial investigation into <u>Kenya's disinformation</u> industry by political actors. For this, we used data from Twitter's firehose to dig into the campaigns. We also once again managed to interview influencers who were paid to participate in the campaigns. Examining this activity provided us with further insight into the goals of Kenya's disinformation industry and just how long this industry has existed.

#### <u>The strategy:</u> Anti-Pandora Papers 'astroturfing'

On 6th October, three days after Africa Uncensored (an independent journalism outlet in Nairobi) and ICIJ released the Pandora Papers, the hashtag, #phonyleaks, was promoted by Twitter's algorithms on the platform's trending section. It would go on to feature prominently for 13 hours that day and even persist on the <u>following day</u>.



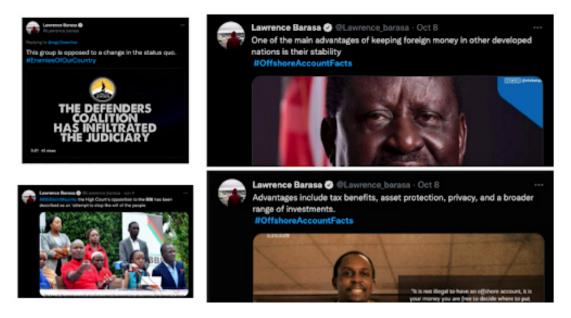
In total, we examined 8,331 tweets produced during the campaign in the week of the Pandora Papers' release (3rd to 10th October 2021) from 1,935 participating accounts. Additionally, on October 8th, another counter-narrative effort got published on Twitter's trending section, <u>#offshoreaccountfacts</u>. We identified 3,870 tweets that were published under it by 833 authors.



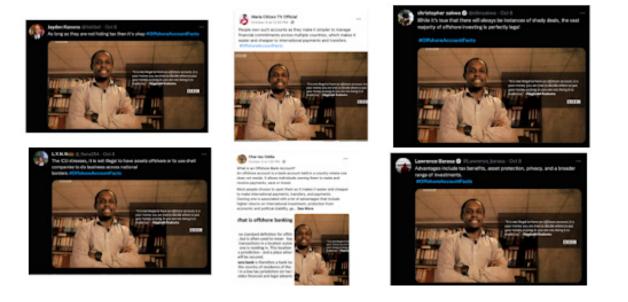
An analysis of the content within the hashtags and the behavior of the accounts involved showed us that these assets that were indeed used for inauthentic behavior i.e.

Several of the accounts that participated tweeted for days on end on nothing but political hashtags. This is because these accounts are mostly used for the aim of getting something to trend, almost every tweet they put out has a hashtag.

Accounts within the campaigns also popped up in the database of inauthentic actors that we had identified after our initial investigation into Pro-BBI campaigns in May 2021. Many of the accounts we found have been tweeting pro-government propaganda. We brought forward these accounts to Twitter earlier this year. Some were taken down and some still stayed up.



There was a heavy presence of the repetition of content within the hashtags. The accounts involved in the campaign repetitively used the same set of media assets in their tweets or <u>repeated</u> the <u>phrases</u> within them. This wasn't just on Twitter alone but happened on Facebook too.



Inauthentic name behavior was also present. Accounts mimicking celebrities were used heavily by the influencers in classic bait and switch fashion. This is part of a formula where individuals use a widely familiar likeness and appealing content to grow audiences for an account. From there they then use it for the purposes of disinformation. It's also known as "Inauthentic Viral Spam." The identities of Diana Marua and Lillian Ng'ang'a, both local celebrities in Kenya, were used to spread disinformation for the #offshoreaccountfacts campaign. We reached out to them and both denied any association with the accounts we identified.

#### Inauthentic Virality Spam



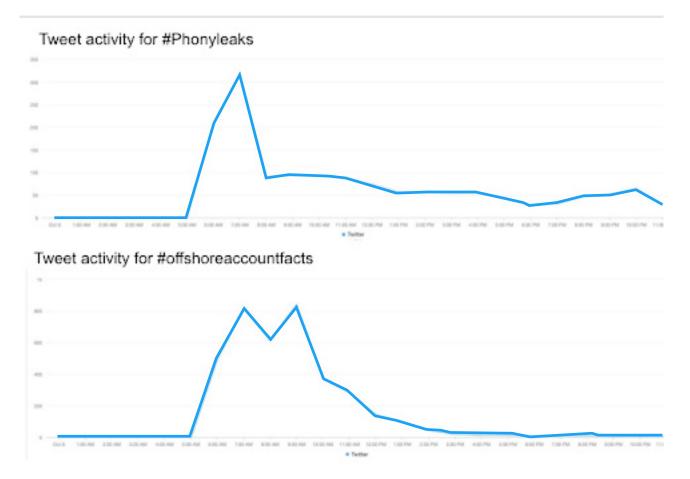
Step 1: set up account using the likeness of a celebrity.



Step 2: post appealing content such as jokes, memes and pictures from the celebrity who's identity you've stolen. With this you'll gather a steady audience.



Step 3: Use the account with the audience you've gathered to spread disinformation. The highest volume of content produced was also in the morning hours. We witnessed this in our research on the BBI propaganda campaigns as well. According to Robin Kiplagat, a disinformation researcher at Code for Africa, the time of the initial content is vital: "Based on previous research we've conducted, our observation is that many of these hashtag campaigns are launched in early morning hours. Mostly between 6-7am. It's easier to go trending at that time." Furthermore, based on the timestamps tweet activity would experience a sharp spike and then fizzle out soon after. For #Phonyleaks however, the influencers claimed they would schedule additional tweets to run later so that the hashtag could remain on the trending section.



By examining the two hashtags and interviewing influencers involved in the campaign, we were able to understand its core message. The focus was on emphasizing that:

-Uhuru Kenyatta was not a direct beneficiary of the accounts and he did not steal any of the money that was highlighted.

-The house the family bought through the offshore activity actually appreciated and therefore it was a good investment.

-The amount that was identified, \$30 million, was very little compared to other bits of wealth that the Kenyattas have in Kenya and therefore it isn't a big deal.

-Uhuru Kenyatta should be praised for the transparency he showed in his response to the Pandora Papers.

-It's not illegal to have any money in offshore accounts as a public official and they actually confer some advantages.

Of key importance here was that many of these narratives were not explicitly lies. This was political astroturfing that used a mix of <u>propaganda</u> and <u>mal-information</u>. It was designed to fabricate consensus – specifically, the consensus that most Kenyans support Uhuru Kenyatta and distrust the Pandora Papers.

However, some of the content within these campaigns was completely fabricated. Reginald Kadzutu, an economist with Amana Capital, found that his likeness had been used to make it appear as if he had done an interview for the BBC. The image attributes a quote about the legalities of offshore accounts to him. Kadzutu and the BBC confirmed that no such interview had happened. The image is fake.



By focussing on these specific points their aim with the campaign was to dispossess the Pandora Papers of their potential as a driver of change. They were meant to allow politicians to distract citizens from what was crucial. There are many people in Kenya who are supporters of Uhuru Kenyatta and did not view the details within the Pandora Papers as a bad thing. This campaign was designed to appeal to them.

Purity Mukami, who was one of the Kenyan reporters on the Pandora Papers who broke the story in Kenya said: "I saw the campaigns against our story on Twitter and it was very clear that they were trying to discredit our work. It however, was very difficult for them to do that because of how well researched our reporting was. The best they could do was try to pacify the story and lessen its impact. If you combine these campaigns alongside the kind of media coverage the Pandora Papers in Kenya got, you can see how Kenyans may have gotten confused as to what the way forward should be. Kenyans were denied the opportunity to discuss what makes holding government to account in this country impossible. That is the goal these disinformation campaigns had and they were successful."

"If you combine these campaigns alongside the kind of media coverage the Pandora Papers in Kenya got, you can see how Kenyans may have gotten confused as to what the way forward should be."

-Purity Mukami, Kenyan journalist covering the Pandora Papers

## <u>The context:</u> Another entry in a flourishing industry.

While the Pandora Papers revelations may be new, the existence of Kenya's disinformation industry is not. Our interviews with disinformation influencers about the Pandora Papers campaign and May's BBI campaign provided a perspective on just how long this activity has been going on. One of the influencers we spoke to, via WhatsApp and on the condition of anonymity, mentioned that they had been working on manufacturing Twitter trends for over five years: "Back then, I was in campus and it was an avenue for me to make money." In that time, he also claimed to have participated in campaigns in 2017 as a keyboard warrior for political parties. At that time, he claims influencers were paid at a rate of around \$25 per hashtag, with an average of 6 hashtags a day. For the past five years, he has essentially been able to set up multiple sock puppet accounts and spread propaganda by manipulating Twitter's trending algorithm. Twitter, meanwhile, has never caught him.

Over the years, disinformation influencers also seem to have a formula for what it takes to get Twitter to publish their content on its trending section. The influencers we spoke

to also gave us some insights as to what their perceived formula for getting the hashtags to trend would be; "On a normal day, 500 tweets will do the magic. Keep in mind, that even a retweet is a tweet. Within that mix you'll also need around a 100 original tweets with photos. The #phonyleaks campaign, because it came on a day where there was a lot going on, needed about 2,000 tweets to trend."

Furthermore, the providers of these services confirmed to us that the service they provide is given to not only to political bidders, but just about anyone. Politicians and their operatives however, get charged much higher than others because of the risk involved. "Twitter and social media are just a game. It's just business," one influencer said. What was even more astounding to us, was that the disinformation industry is complete with success stories of individuals who've "made it big." "I know a guy who started doing this business when he was very broke and now the guy even got married, bought himself a car and his cheeks are very round. There's money in this thing," one influencer told us.

"I know a guy who started doing this business when he was very broke and now the guy even got married, bought himself a car and his cheeks are very round. There's money in this thing."

-Disinformation influencer

What's especially noteworthy is how this space operates more and more like a marketplace. Our interviews with the influencers also revealed that there are multiple influencer groups on WhatsApp which serve this exact function. Based on screenshots that we were able to review, here they exchange information about available political campaigns and recruit each other.

Influencers' own politics rarely motivate or inform their participation. As Kenya's ruling Jubilee Party has fractured, so has influencer groupings within the disinformation apparatus of the party. "Many of the people participating in campaigns are aligning themselves with where the money is." Some influencers provide services across the divide. "I have one set of accounts which I use when the Pro-Uhuru (pro-government) guys come calling and another set of accounts which I use when the P<u>ro-Ruto guys</u> come calling (pro-opposition)."

Influencers also mentioned that because of the tribal nature of Kenyan politics, the name of the pseudo accounts they set up are important. "Name matters because, from there, audiences can tell which tribe you're in and assume which party that account is aligned to. It makes it more believable." One of the influencers also similarly informed us that "There are accounts I can't use for certain jobs, I align myself with people based on the audience they can reach."

"Name matters because, from there, audiences can tell which tribe you're in and assume which party that account is aligned to. It makes it more believable."

-Disinformation influencer

## <u>The</u> Upshot: Twitter is complicit in harming democracy

As we mentioned in the last report, social media platforms, especially Twitter, matter. The public outcry that followed the publication of the Pandora Papers continues to show how important these platforms are in terms of citizen accountability. This is especially true in light of how critics view the framing of the revelations by legacy media. As an attempt to place it as an issue of legality instead of dignity and ethics.

The situation that prevailed in Kenya after the Pandora Papers were published, is one that does not bode well for its civic space. The delayed and watered down reaction of Kenyan media to the international controversy over the leaks, created an information vacuum for propaganda campaigns to avidly scuttle credible reporting by smaller independent outlets. Disinformation campaigns go hand in hand with the prevailing media environment of a society, just as they can do in the aftermath of natural disasters, where media vacuums leave voids that are easily filled with misinformation.

These disinformation campaigns can fuel more than chaos and distrust online – they can also preempt physical violence offline. One Kenyan activist who was targeted by the same accounts involved in the Pandora Papers campaigns campaigns, Boniface Mwangi, had his house bombed just two days after the online attacks. <u>Police are investigating</u> the bombing and haven't yet determined the culprit, although Mwangi speculated about a potential link: "They realised hashtags can't scare me and they're now using explosives," he <u>tweeted</u> on October 20. As journalists and activists in the <u>Philippines</u>, <u>Myanmar</u>, and elsewhere in the world have revealed, disinformation and harassment campaigns online can spiral into violence and tragedy offline.

Meanwhile, once again, Twitter's trending algorithm is a vital ingredient in this equation. We were able to identify a similar campaign running on <u>Facebook</u> but it is likely that they got nowhere near the same viewership on Twitter. The problem is not that there's bad stuff on the internet. The problem is that some features of social networks are designed to make that bad stuff highly visible to many people. Manipulating Twitter's trending algorithm, allows people lacking large organic followings to reach wider audiences with harmful messages. As one of the influencers we spoke to put it, "Twitter is easy."

*"Twitter is easy."* –Disinformation influencer

We reached out to Twitter for comment on the activities we highlighted. After their own internal investigation, they also took action on 230+ accounts operating in the country which they found had engaged in violation of their platform manipulation and spam policy.

Through a spokesperson they said: "Twitter's top priority is keeping people safe, and we remain vigilant about coordinated activity on our service. Using both technology and human review, we proactively and routinely tackle attempts at platform manipulation and mitigate them at scale by actioning millions of accounts each week for violating our policies in this area. We are constantly improving Twitter's auto-detection technology to catch accounts engaging in rule-violating behavior as soon as they pop up on the service."

One of the things that we're fast understanding is that disinformation is an industry. For platforms like Twitter, disinformation clearly won't go away. There is a very strong demand for successfully capturing public attention. Gaming Twitter's algorithm provides a sufficient prize for politicians seeking to harass activists and journalists. This report is not the first to highlight the harms that the platform's features can cause. Fixing the problems with Twitter's ecosystem will take more than just removing the trending feature from the platform. For Kenyans, it means taking a deep look at how their politics has morphed in the digital age and seeing how to develop rules to govern this. For Twitter, it's about being more transparent about their trust and safety efforts in African countries, being consistent in enforcing their Terms of Service and paying more attention to our context. Let's start there first.

## Acknowledgements

This report was written by Mozilla Tech and Society Fellow, **Odanga Madung**. Additional reporting was done by **Brian Obilo** (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.



