Securing Kenya's Electoral Integrity in the Digital Age: Censorship

By Abdulmalik Sugow and Isaac Rutenberg

On 21 October 2021, when faced with protests against the King, the southern African kingdom of Eswatini directed mobile operators based in the country to suspend access to Facebook. This is just one example—out of hundreds—where governments have turned to internet shutdowns as a means of suppressing opposition and stifling organising. At the same time, social media platforms have been accused of suppressing political speech when purportedly enforcing their terms of service. In our first article in this series, we highlighted some of the ways in which Kenyans have used social media platforms for civic participation and grassroots organising. Through these platforms, people across the world have enjoyed the ability to seek information, engage in debate, and drive movements.

The increasing centrality of these platforms to democratic processes such as elections means that the harm posed when access is disrupted is often immeasurable, though several researchers have attempted to quantify this harm qualitatively through citizens' experiences, and quantitatively through economic impact. The Centre for Intellectual Property and Information Technology Law for example, has framed these harms as including increased citizenry backlash, economic losses, and eroded international reputation. In this article, we discuss the intentional suppression of political speech online by social media platforms and by governments, detailing the ways in which these manifest, and the dangers they pose to electoral integrity.

Suppression by platforms

Almost exactly a year prior to the Kingdom of Eswatini's blocking of Facebook, there were widespread protests in Nigeria over the Special-Anti Robbery Squad (SARS). Fed up with the corruption and brutality of SARS officers, citizens took to the streets. During these protests, it was reported that the army used live ammunition, resulting in the death of some of the protestors. When these reports were shared, Instagram's algorithms [flagged them as 'potentially false'](), further inflaming sentiments around the entire ordeal and leading Instagram to [issue an apology]().

This occurrence is but one example of the challenges facing content moderation by these platforms in Africa, and by extension the Global South. These platforms are empowered to enforce their Terms of Service (ToS) which, often, contain guidelines that restrict the spread of false information and hate speech. When detected, such content is either taken down entirely or downranked. However, enforcing these restrictions is often difficult. For one, due to the sheer amount of content shared, these platforms often use artificial intelligence to flag violations of their ToS.

These technologies are often trained on datasets that are not representative of the lived experience of Africans and as such, are biased from the outset. These biases manifest in erroneous actions such as Instagram's response to the #EndSARS protests. Platforms such as Facebook are aware of these shortcomings; recently leaked internal documents revealed that Facebook's artificial intelligence moderation tools were [flagging cockfights as car crashes, and mass shootings as either paintball games or a carwash.]() While these platforms also have human reviewers to verify some of the actions taken by the moderation tools, these humans are almost always blind to the specific contexts and nuances of the societies to which the content they moderate relates.

Alive to these concerns, Facebook pledged to hire [100 moderators and trusted flaggers to cover every African market.]() It is not clear how many have been hired so far. It also remains unclear whether this would have a noteworthy impact as these "African markets" can further be segmented into thousands of language groups, all making use of these social media platforms, complicating moderation efforts. Erroneous flags that result in either the taking down or downranking of content have undermined efforts to raise awareness around injustice or to organise movements, as was the case in [Tunisia when Facebook took down accounts belonging to 60 activists](). These decisions are often consequential from a political perspective and take place in relative opacity.

> These technologies are often trained on datasets that are not representative of the lived experience of Africans and as such, are biased from the outset.

Aside from errors in content moderation, social media platforms sometimes intentionally engage in suppression of speech at the direction of governments. Through [transparency reports](), Facebook and Twitter disclose instances where they have been requested by governments to either take down or restrict access to content based on such content violating local laws. With the enactment of laws criminalising the spread of subjectively defined false information (laws which Matt Bailey calls "[fake censorship]()"), these platforms are effectively co-opted into censoring political speech which governments deem unfavourable.

During the 2017 elections, Facebook reported that it [restricted access to 13 items that allegedly violated hate speech and election laws in Kenya](). The specific content is not available on the transparency report, but this highlights the government's ability to invoke local law to require these platforms to take down content. On the face of it, such an arrangement is understandable as governments are better placed to assess compliance with their laws. However, it becomes an issue

where laws were enacted with a view to shrinking the civic space and suppressing any activism in the face of authoritarian tendencies.

In [our previous article](), we highlighted the subjective and selective application of Kenya's Computer Misuse and Cybercrimes Act. Applied to this context, if the government were to request a takedown based on this law, Facebook or Twitter would simply comply, furthering the potential for suppression. While these platforms make it clear that they screen these requests prior to complying, the growing use of local laws which impose liability on platforms for failure to comply would incentivise them to err on the side of caution and blindly comply with requests. Turkey for example recently enacted a law that requires platforms to [respond to complaints within 48 hours or face fines of up to US$700,000](). That being said, some platforms such as [Twitter have made it clear]() that where the complained of content does not violate its ToS, it would only restrict access in the jurisdiction with the law in place.

There is a discernible trend towards bringing these platforms under the control of governments. India recently [withdrew safe harbour protections]() from Twitter due to its taking down of content associated with the ruling party. It also came under fire for allowing content critical of the government; the [police raided Twitter's offices in Delhi](). India enacted new guidelines requiring platforms to appoint local representatives to handle complaints. Following these rules, the Indian government's takedown requests have increased, often targeting content that is [critical of the government or the ruling party](). The co-opting of these platforms by governments or public authorities may sometimes put lives at risk. Recently, the Facebook Oversight Board recommended an independent investigation into [Facebook's suppression of pro-Palestine content at the request of the Israeli government](). According to activists, this [suppression sometimes puts lives in danger]() as they were unable to share information regarding the state of security at the time.

Internet shutdowns

In some instances, when unable to lawfully secure the takedown of content on platforms, governments resort to internet shutdowns to quell opposition. In the past decade, it is estimated that several governments have either [wholly or partially shut down the internet a total of 850 times, with 90 per cent of these instances taking place in the last 5 years](). These shutdowns occur on a spectrum, from blocking specific websites such as social media platforms, to completely shutting down access to the internet for entire regions. A [report by Jigsaw and Access Now]() documented some of the most recent internet shutdowns such as by Uganda earlier this year, and [Tanzania late last year]().

These shutdowns are [easier to implement]() where there are few internet service providers operating in a country, and where the government maintains significant control over them. These shutdowns have untold political and economic consequences (due to the informal sector's use of social networks such as WhatsApp to trade). In Myanmar for example, it is estimated that [approximately 2.5 per cent of the country's Gross Domestic Product was lost]() due to a partial internet shutdown – the military junta [blocked access to Facebook during the day, and wholly shut down the internet every night for 72 consecutive nights](). It must however be noted that using economic statistics as a measure of impact is not highly accurate in several African countries due to [informal and unreported economic activity]().

> Erroneous flags that result in either the taking down or downranking of content have undermined efforts to raise awareness around injustice or to organise movements.

While these shutdowns are often linked to temporal events such as elections, they may sometimes persist indefinitely. In June 2021, Nigeria indefinitely banned Twitter after it enforced its ToS against President Buhari and deleted one of his tweets for violating its policies. Since the ban, researchers have estimated that Nigeria has [lost US$366 million](#) due to a decrease in economic activity which ordinarily took place through the platform. To lift the ban, Nigeria is requiring Twitter to comply with a [raft of measures](#) such as registration in Nigeria, payment of local taxes, and appointing a representative. In the past two years alone, internet shutdowns have been reported in [Algeria, Burundi, Eswatini, Ethiopia, Guinea, Mali, Sudan, Togo, Tanzania, Uganda, Zambia, and Zimbabwe](#). Bearing in mind the persisting pandemic and the centrality of digital interactions during this time, these shutdowns have set a dangerous precedent, both for public health and for political speech.

Safeguarding civic space online

Earlier this year, Kenya's Cabinet Secretary for Interior and Coordination of National Government publicly assured Kenyans that the government would [not shut down the internet.](#) A few months later, he cast doubts over the strength of this assurance by stating that the government would not hesitate to [shut down mainstream media involved in disseminating harmful content by invoking the Public Order Act](#). For this reason, such assurances ought not to detract from ongoing efforts to remain vigilant of suppression of content online, whether by platforms or governments. Organisations such as Access Now and the Open Observatory of Network Interference (OONI) have been tracking internet shutdown trends across the world and raising awareness – [Access Now through its #KeepItOn program](#) and [OONI through its publicly accessible probe and shutdown reports](#). These efforts ought to be amplified by civil society in Kenya and could be plugged into by regulators such as the Communications Authority (CA) to demonstrate the goodwill in the Cabinet Secretary's commitment not to shut down the internet. It would also be prudent for the CA to commit to transparency in any takedown requests they make to social media platforms. On their part, these platforms should also work with the electoral body—the IEBC—and the CA to make transparent the moderation tools they intend to deploy in Kenya during the elections. Such a collaboration ought to also involve civil society so as to boost accountability.

–

This is the fourth of a five-part op-ed series that seeks to explore the use of personal data in campaigns, the spread of misinformation and disinformation, social media censorship, and incitement to violence and hate speech, and the practical measures various stakeholders can adopt to safeguard Kenya's electoral integrity in the digital age ahead of the 2022 elections. This op-ed series is in partnership with [Kofi Annan Foundation](#) and is made possible through the support of the [United Nations Democracy Fund](#).

Published by the good folks at [The Elephant](#).

The Elephant is a platform for engaging citizens to reflect, re-member and re-envision their society by interrogating the past, the present, to fashion a future.

Follow us on [Twitter](#).