Areo

Will Coronavirus Prompt Social Media Giants to Address Disinformation?

May 25, 2020

Nayef Al-Rodhan

In what now seems like a lifetime ago, before the World Health Organization had characterized the spread of Covid-19 as a global pandemic, Tedros Adhanom Ghebreyesus, the WHO's director general, stated at February's Munich Security Conference that, alongside the escalating epidemic, the global population was fighting a parallel "infodemic," referring to the highly contagious spread of false information.

Since then, it has become abundantly clear that, like the pandemic, the infodemic has increased in potency in an exponential manner, in large part due to the confusion caused by the ongoing pandemic.

The Reuters Institute recently published a report identifying the main types and sources of misinformation surrounding Covid-19. Despite acknowledging that social media platforms have, for the most part, made efforts to react to posts containing false information, the report highlights the fact that 59% of Twitter posts that the institute's fact-checkers had rated as false remained live. On YouTube, 27% remained unflagged, and, on Facebook, 24% of content rated as false was overlooked.

The coronavirus has exposed the limits of the ability of social media companies and governments to deal with the colossal volume of disinformation that is broadcast to users. Drinking lemon tea and gargling salt water are just two of the false remedies circulating in Facebook groups and WhatsApp forwards, while Twitter has enabled even public figures to promote conspiracy theories, linking 5G radiation to the spread of the virus.

The persistence of disinformation continues to be relentless, despite newly enforced measures by both private entities and public institutions. The added difficulty posed by closed groups on Facebook and encryption on WhatsApp has allowed false advice to manifest in close-knit circles. Misinformation hijacks emotional ties between family and friends, allowing such messages to gather self-reinforcing momentum.

In these precarious times, we are reminded of the significant role emotions play in our thinking and decision-making. In times of increased vulnerability, we are more likely to buy into questionable information, and seek validation by sharing this information with others.

The pandemic has invigorated what was already an urgent conversation for social media platforms, much maligned for their inaction in response to misleading content during political campaigns, as they look into more extreme measures to fight the slew of harmful and often deliberately orchestrated disinformation campaigns. The graduation of this problem from the realm of politics to that of public health has radically increased their responsibility to act quickly and effectively.

Some progress is being made. Among other new policies, Facebook has promised to notify users who have interacted with content from unreliable sources and direct them to the WHO.

Twitter's definition of *harm* now includes content that opposes public health guidelines. However, the platform appears unwilling to abandon its self-appointed role as the "free speech wing of the free

speech party." Abiding by its longstanding principles of neutrality, it relies primarily on users *reporting* harmful content and on *requiring* users to remove tweets.

Media platforms are right to defend free speech, but with freedom comes responsibility. Self-regulated tech giants might well prioritise *verified* information on their users' timelines and increase their cooperation with fact-checkers, but they still appear reluctant to prevent the spread of disproved information by directly removing it from their platforms. Facebook's focus is on "reducing" the likelihood of fake news going viral and Twitter caveats its guidelines by permitting false information from world leaders "in the interest of the public."

This sends out a confusing message to users. A platform that bases its values on free speech should, of course, encourage healthy debate, but there is nothing debatable about unsubstantiated cures or the disregarding of official health advice, as Twitter itself points out.

For example, the approach taken to Brazilian president Jair Bolsonaro's promotion of hydroxychloroquine as a *cure* has been distinctly different to the treatment of US president Donald Trump's engagement with accounts encouraging civilians to abstain from social distancing. Both presidents have violated Twitter's own guidelines, yet Twitter has chosen to act against Bolsonaro's account, but not against Trump's. What we require is consistency.

The difficulty, of course, lies in the matter of public interest. If social media platforms are to remain platforms, rather than media outlets in their own right, they are not in a position nor do they have a responsibility to censor publicly available information, especially that emanating from (supposedly) authoritative sources.

What we urgently require is a discussion of how such information is scrutinised and validated, how we might better identify and classify the risk of potential harm from the dissemination of controversial content in purely neutral terms.

As the Reuters Institute report shows, most engagements with Covid-19-related fake news involve reconfigured information: content that reworks the truth by including some false information. This allows potentially dangerous information to remain unidentified in its diluted form, and reinforces the need for social media companies to take a more rigorous approach to employing flagging mechanisms that alert users to the informants' sources.

The evolution of the social transmission of disinformation has run parallel to that of this unprecedented virus. Neither fight is going to be an easy battle.

The world has become smaller, as we all focus on one story, across endless platforms, with more time available to dedicate to the news, all against the backdrop of increased emotional vulnerability. How social media companies and governments act now will be decisive in how they are viewed in the post-pandemic world.

The focus now should be on improved fact-checking technology, greater dialogue with the international scientific community and the securitisation of fake news. Companies have a duty to filter reliable information. As the Reuters report shows, 39% of false claims have focused on public health institutions. People are desperate for official guidance, but struggle to extract the facts.

The pandemic has illustrated the essential role that governments and legislators play in times of crisis. They have deployed unprecedented efforts to mitigate the sanitary, social, and economic impacts of Covid-19.

However, the infodemic must be taken more seriously by governments and social media firms worldwide. And, while the Network Enforcement Act in Germany and the Avia law in France have exemplified the difficulty of legislating social media content, they have also highlighted the need for states to tackle an issue that threatens democratic legitimacy. If technology platforms cannot effectively determine whether or not content is harmful, or can only do so in a pointlessly ambiguous manner, then the law must step in.

If we cannot rely on the platforms to keep a tighter rein on fact-checking and mine information in order to save lives, as opposed to the separate mission of enabling public discourse, we will continue to suffer an infodemic as well as a pandemic—and the consequences of the former could well be longer lasting.