

Prompt: How should social media companies deal with disinformation campaigns?

By: Ali Alawami

Social media platforms (SMP) make their money by serving ads to their users based on the data they have gathered on said users. In order to serve them more ads, SMP platforms aim to engage their users while using the service to keep them site longer. This resulted in many disinformation campaigns being served up more frequently to users and echo chambers forming as they are more likely to react to it. SMP need to realize the drawbacks of such approaches in the longterm and work to edit their engagement based algorithm, break echo chambers, soft banning certain posts, block spam accounts, stop serving disinformation and make it harder for the perpetrators of disinformation to realize it.

Firstly, SMP need to reevaluate their engagement based algorithms. By focusing only on driving engagement up and not on what type of engagement they are driving up users will eventually realize a pattern in the type of content they are being served and associate it with the platform. For example, user X who believes in vaccines is being served up a lot of disinformation regarding the COVID-19 vaccine because they keep trying to correct or discredit the information which the algorithm classified as engagement. Eventually, user X will burn out from this and forever associate the platform with promoting misinformation which is a bad feeling and may result in them abandoning the platform.

Secondly, SMP should work to break disinformation echo chambers when identified. This can be by simply stoping serving disinformation content once the echo chamber forms effectively killing it. Better yet, the SMP can try to reverse some of that damage by serving up more factual information to those who were stuck in the echo chamber. Repeated exposure can be a good thing too.

Thirdly, many disinformation posts are simply copies of other posts shared by users. When a post is identified to be disinformation SMP should work to also soft ban all copies and shares of that post to help stop the spread of that disinformation.

Fourthly, many echo chambers and disinformation campaigns are launched by few individuals utilizing many spam accounts to drum up fake engagement to trick the algorithm into serving that content more and trick potential victims of the campaign into believing more people are on board with that information than is true.

Lastly, many of the bad actors who spread disinformation make new accounts when they have been banned or soft banned. This means that the longer they are unaware of their banned status the more time they will spend using these banned accounts effectively wasting their time and lessening the damage they cause.

By prioritizing positive engagement, breaking echo chambers, better identifying disinformation spam accounts, and wasting bad actors time social media platforms can lessen the effectiveness of disinformation campaigns.