Social media platforms remove fewer than 1 in 10 posts reported for coronavirus misinformation

Facebook, Instagram and Twitter are failing to take action against more than 9 in 10 posts containing misinformation about coronavirus which are reported to the platforms, new research finds.

649 posts that contravened their stated standards were reported to the social media platforms, but action was taken on only 61 (9.4%) of the posts. Of those, 6.3% of posts were removed, 2% of posts were from accounts which were subsequently taken down, and 1% were flagged as false information but remain online. 90.6% of misinformation posts reported to the platforms were left untouched.

Examples of the posts which were not removed include conspiracy theories and false claims that:

- Viruses cannot be transmitted by air
- Covid-19 is caused by vaccines
- 5G mobile technology poisons cells and causes the illness known as Covid-19
- Coronavirus is a hoax, in order to cover for the erection of 5G towers
- Covid-19 is a “false flag” in order to force compulsory vaccinations
- Colloidal silver cures and prevents Covid-19
- Seven children died in a vaccination trial in Africa
- Taking vitamins C, D and zinc, and drinking hot water, cures Covid-19
- Covid-19 can be cured by drinking aspirin dissolved in hot water with honey
- Wearing a facemask causes cancer
- The pandemic is fake and will be used to microchip and enslave the public

The findings are in stark contrast to public statements made by each social media company, claiming to be taking strong action to stamp out coronavirus misinformation. Facebook says it removes false information and conspiracy theories that could be harmful, including promotion of false cures or 5G conspiracy theories. Twitter has promised to remove tweets that deny the effectiveness of measures to prevent the spread of Covid-19, promote false cures for the disease or deny scientific facts about it. Instagram also claims it removes harmful misinformation related to Covid-19.

The research is published by the Center for Countering Digital Hate (CCDH) and Restless Development and was conducted by Youth Against Misinformation, a group of young activists coordinated and trained by Restless Development and the CCDH. The volunteers found and reported the misinformation posts between 20 April & 26 May.

Imran Ahmed, CEO of the Center for Countering Digital Hate, said:

“In this pandemic, we all have to do our bit to keep our communities safe. Social media giants are shirking their responsibility to stop dangerous misinformation spreading. Their systems for reporting misinformation and dealing with it are simply not fit for purpose.
"Social media giants have claimed many times that they are taking Covid-related misinformation seriously, but this new research shows that even when they are handed the posts promoting misinformation, they fail to take action. What is a real kick in the teeth is that while they were shedding content moderators and complaint reviewers, Facebook nevertheless found $400 million dollars to buy a meme library.

“If social media giants continue to publish misinformation on their websites, then politicians need to hold them to account by imposing financial sanctions for the costs to the NHS, fire service, police and all of society that misinformation causes, and legislate for deeper, faster regulation.”

Rosanne Palmer-White, Director of Restless Development UK, said:

“Youth Against Misinformation is youth power in action. Young people are doing their bit to stop the spread of misinformation about COVID-19 online, but social media companies are letting them and all of us down. This research shows they are failing to keep their promises despite the efforts of young people to help them take action on the infodemic. Actions speak louder than words; we urge these companies to listen to young people who are their allies in finding dangerous misinformation, removing it swiftly and stopping it doing harm.”

Sara, a Youth Against Misinformation activist who helped find and report the posts, said:

“It’s discouraging to see such few pieces of misinformation removed, but it is particularly frustrating that there often doesn’t seem to be any rhyme or reason as to which pieces are removed.

“However false something is, there is clearly no guarantee these platforms will take the responsibility to stop it spreading seriously.”

ENDS

Notes

1. The research can be found in the new report published by the Center for Countering Digital Hate and Restless Development, titled ‘Will To Act,’ found here: https://bit.ly/WillToAct

2. Please find attached screenshots of a selection of posts that were reported but not removed by Facebook, Instagram and Twitter.

3. Youth Against Misinformation, formed of 27 young volunteers from the UK, Ireland and Romania, is coordinated and supported by Restless Development and the Center for Countering Digital Hate. It has two aims: to monitor and root out dangerous COVID-19 misinformation online; and build the skill-set and agency of young people to understand how misinformation is spread online, identify fake news and develop strategies to tackle it. Volunteers receive comprehensive training on how to identify,
report and respond to harmful misinformation online, following five principles of CCDH's Don't Spread the Virus campaign.
https://restlessdevelopment.org/young-people-join-forces-to-tackle-infodemic

4. “Things like saying that something is a proven cure for the virus when in fact it isn’t: we will take that down. Another example which I know has been very prevalent in the UK has been 5G misinformation, which has led to some physical damage of 5G infrastructure, so we believe that is leading to imminent risk of physical harm, we take down that content.”
Mark Zuckerberg, BBC News, 21 May 2020

“We will also start to remove content with false claims or conspiracy theories that have been flagged by leading global health organizations and local health authorities that could cause harm to people who believe them. We are doing this as an extension of our existing policies to remove content that could cause physical harm. We’re focusing on claims that are designed to discourage treatment or taking appropriate precautions. This includes claims related to false cures or prevention methods — like drinking bleach cures the coronavirus — or claims that create confusion about health resources that are available.”
Facebook, 30 January 2020
https://about.fb.com/news/2020/05/coronavirus/

Twitter has promised to remove tweets that deny the effectiveness of measures to prevent the spread of COVID-19, promote false cures for the disease or deny scientific facts about it. Full list at the link below:
Twitter, 27 March 2020

“We also remove false claims or conspiracy theories that have been flagged by leading global health organizations and local health authorities as having the potential to cause harm to people who believe them.”
Instagram, 24 March 2020
https://about.instagram.com/blog/announcements/coronavirus-keeping-people-safe-informed-and-supported-on-instagram

5. Facebook put most of its content moderators on leave in March, limiting its ability to remove harmful content.

6. Facebook spent $400 million purchasing GIPHY, an online image library, last month.
https://www.axios.com/scoop-facebook-to-buy-giphy-for-400-million-4a75a359-833b-484d-b15b-87e94d3de017.html
7. Executives from the major tech companies will be appearing before the Digital, Media, Culture and Sport Sub-committee on Online Harms and Disinformation on Thursday 4 June, to face questions from Members of Parliament about actions they have taken to tackle the infodemic of misinformation about Covid-19. [Link to the committee's website](https://committees.parliament.uk/committee/438/digital-culture-media-and-sport-subcommittee-on-online-harms-and-disinformation/news/146692/social-media-companies-questioned-on-tackling-misinformation/).