Ahmad, Sana

Research Report — Published Version

COVID-19 and the Future of Content Moderation

Coronavirus and its Societal Impact - Highlights from WZB Research

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Ahmad, Sana (2020) : COVID-19 and the Future of Content Moderation, Coronavirus and its Societal Impact - Highlights from WZB Research, WZB Berlin Social Science Center, Berlin

This Version is available at:
https://hdl.handle.net/10419/223146

COVID-19 and the Future of Content Moderation

By Sana Ahmad

The current measures to combat the COVID-19 pandemic have highlighted the extent to which our societies depend on professionals whose work is often poorly paid and undervalued. Among these professionals is the group of content moderators. Invisible to most people, content moderators work for social media platforms such as Facebook. Their job is to remove undesirable or illegal postings like hate speech, depictions of violence, or pornography. Content moderation is a very stressful activity: moderation decisions have a huge impact but must be made within a short timespan often a few seconds per image, film or text, requiring adept knowledge of client policies, high concentration, and the ability to distance oneself from psychologically distressing content. Content moderation is mostly performed by companies in India and the Philippines who are contracted by major social media platforms. The workforce in India and the Philippines mainly consists of fresh graduates, with middle- class backgrounds. They receive low wages and are provided with meagre upskilling opportunities.

Although companies like Facebook have sought to automate content moderation by Artificial Intelligence (AI) for a long time, human content moderators remain central. As humans, they draw from cultural knowledge and cognitive abilities, both of which are hard to automate. Content moderation moves along a fine line between the justified removal of either illegal or inhumane content and censorship (which inevitably creates conflicts, for example when it leads to the removal of artistic content and satire).

The coronavirus crisis represents a major disruption for the functioning of content moderation. Like everyone else, content moderators are affected by quarantine measures, and since, large platforms have been increasingly relying on AI. The shift to AI content moderation has long been promised by Facebook CEO Mark Zuckerberg. Facebook has increased investments in machine learning capabilities. These developments are confirmed by Tejas*, a former content moderator interviewed by the author of this article. “Every decision we make is stored. They already have enough data to automate.”

However, the deficits of existing AI solutions are becoming apparent. Since March17th, 2020, a large number of user-generated content, including news articles and information on the COVID-19 pandemic was deleted from Facebook for allegedly violating its community spam rules. While recording an ‘unprecedented usage’ across its platforms, Facebook attributed the issue to a ‘mundane bug in the platform’s automated spam filter’. This automated filter was installed to replace the content moderators who cannot work at their offices any more. Experts estimate that some 15,000 content moderators contracted at 20 global sites have been asked to stay at home.

“We are asked to not work from home as it is highly confidential content. They (supplier firm) said that the work should not be seen by another person. I think this is why they never issued us office laptops to work from home.”

- Sunil*, one of the 800 moderators employed at the MNC subsidiary in Hyderabad.

We can therefore expect that the current surge in the use of automated filters will be followed by the increased use of AI solutions for content moderation. On the one hand, this is a good thing. If Facebook replaces its large pool of cheap labor with AI, it might actually do a favor to workers, protecting them from the extremely stressful work of removing the global 'social media's waste'. On the other hand, this development comes with several drawbacks. Workers will find themselves in difficulties. Prevailing unemployment and a looming economic crisis in India make for a grim situation. Indian moderators are highly concerned. Secondly, the deficiencies of AI solutions are quickly becoming apparent. Facebook's current fight against the COVID-19 'infodemic' is a case in point. Moderation errors and mistakenly deleted content occur frequently. The scale of content and the cost of developing AI is challenging, even for Facebook. Moreover, it is doubtful whether decisions on the removal of content on social media should really be left to software, and whether this does not represent a major threat to the functioning of the public sphere.

*Names have been changed

--

15 April 2020

Sana Ahmad is a Research Fellow of the Research Group Globalization, Work, and Production.