

Technology V12/S5: Moderating Facebook

Content moderators are people that analyze online forums and social media sites to filter out harmful content. Harmful content can be defined as anything published online that could cause distress to the viewer. Artificial intelligence is used to initially search for and remove content before even being reported. This technology will send the images to a human reviewer, and a decision will be made from there.



Facebook was launched in 2004 and has since become one of the most widely used social media platforms. By 2009, Facebook realized that some form of content moderation needed to exist on its site. Today, Facebook employs about 15,000 moderators around the world. These employees spend six hours per shift viewing content that has been reported for graphic imagery and hate speech.

There are criticisms over how Facebook treats the emotional safety of its moderators. Facebook moderators sign disclosure agreements prior to starting the job. Despite the legal constraints, former employees have discussed their experiences working as content moderators with different reporting outlets. The experiences described contain descriptions of animal abuse, child abuse, and violence.

In May 2020, Facebook settled a lawsuit with former employees that claimed they developed mental health problems like Post Traumatic Stress Disorder (PTSD). Many of these moderators say they were not properly prepared for the graphic content they would be working with. Proper mental health support for this level of trauma was not in place. One former employee said that stress-reducing activities included activities like “finger painting”.