Prompt: Why does Miguel, interviewed for "The Trauma Floor" have concerns about accuracy as the metric used to judge content moderators?
By: Ali Alawami

Facebook has user rules and guidelines dictating what they allow on their platform and outsource the enforcement of these rules to third party companies. These moderation companies are subject to Facebook's "accuracy" policy which states that moderators must have a 95% accuracy rate on the 50 to 60 audited posts out of every 1500 posts the moderator reviewed. Miguel pointed many problems with such a cut and dry policy and how the internal systems of moderation reacted to such a policy.

To start with the guidelines are not super clear nor are they very nuanced. The moderators' official source of platform policies are the public community guidelines and a "known questions" document with some clarifications on some policies. However, those documents do not cover every possible scenario of what counts or does not count as a violation. This resulted in many instances where rules are just made up by the team manager. One given example was how erotic asphyxiation was not explicitly banned leading some managers to deem dipictions of fingers pressing on the skin around the throat to count as a violation. To top that out, Facebook's own internal tool meant to inform moderators of new changes or clarifications of some guidelines presents updates based on engagement rather than by recency resulting in many employees reading outdated or completely missing important updates. "It was horrible — one of the worst things I had to personally deal with, to do my job properly" said Diana, an ex moderator. These breakdowns in informing and clarifying guidelines make it near impossible for moderators to achieve the accuracy demanded by them, adding a considerable amount of stress to their already stressful jobs.

These work conditions have also resulted in pitting moderators and quality assurance (QA), those who audit the moderators, against each other. When a post is audited it is up to the

QA whether or not the moderator made the right judgment or not. Meaning that employees with low accuracy scores and ex employees focus all their outrage at QA employees. This led to many QA employees quitting out of fear for their mental and physical safety from current or ex moderators. “Part of the reason I left was how unsafe I felt in my own skin,” said Rand, a past QA employee.

Overall, using a strict accuracy based system to judge employees’ job performance without having clear and well communicated guidelines robs employees of the tools they need to achieve the accuracy thresholds demanded by them. Such breakdown serves only to increase stress and uncertainty moderators face and puts QA employees at the forefront of all the bitter feelings of the moderators who failed to reach the accuracy threshold.