

Task 1

I watched the video.

Task 2

Answer a)

Since freedom of expression protects an individual's right to express themselves freely, it could support a free-for-all approach to moderation.

Answer b)

A key problem with PleaseLikeMe's approach is that it doesn't take into account the content of the post in determining whether or not a post should be removed. Having more likes than dislikes doesn't mean that the content of the post is not harmful or offensive. Moreover, a group of individuals with certain biases can unfairly target certain posts that have content that needs to be addressed in the public, but that group of individuals doesn't want it to be.

Answer c)

One major challenge is that different countries have different laws and cultural norms regarding free speech and acceptable behavior. What may be acceptable in one country but may be offensive or illegal in another country. This creates a challenge for social media platforms in determining a universal set of rules and guidelines that can be applied globally.

Task 3

a) Auto-Detection: The key difficulty with auto-detection is that it may result in over-censorship or under-censorship, it can be challenging for algorithms to accurately detect and filter out problematic content. Algorithms can detect words such as b*tch, s*x, or rape in the tweets but deleting tweets that are having these words, doesn't make any sense and more than this there can be certain tweets that are not using these words but are offensive and harmful for the public.

b) Pre-Moderation: It is time-consuming and expensive to manually review each post before it becomes visible to others, which is the primary difficulty with pre-moderation. It is time-consuming and expensive to manually review each post before it becomes visible to others, which is the primary difficulty with pre-moderation. It is also possible that pre-moderation could not keep up with the volume of tweets on Twitter, which can cause users to feel frustrated or delayed in posting.

Task 4)

Best approach is to use Auto-Detection algorithms when content is posted because Auto-Detection algorithms can work quickly and cheaply, but they may fail to detect some problematic posts or mistakenly flag innocent posts as offensive. So, it is better to have a Pre-Moderation for the tweets that are filtered by these Algorithms and are reported by humans. Mastodon, which is decentralized, allows for more community-specific rules and moderation control. It could, however, make things worse by making it harder to enforce platform-wide rules and control harmful content.