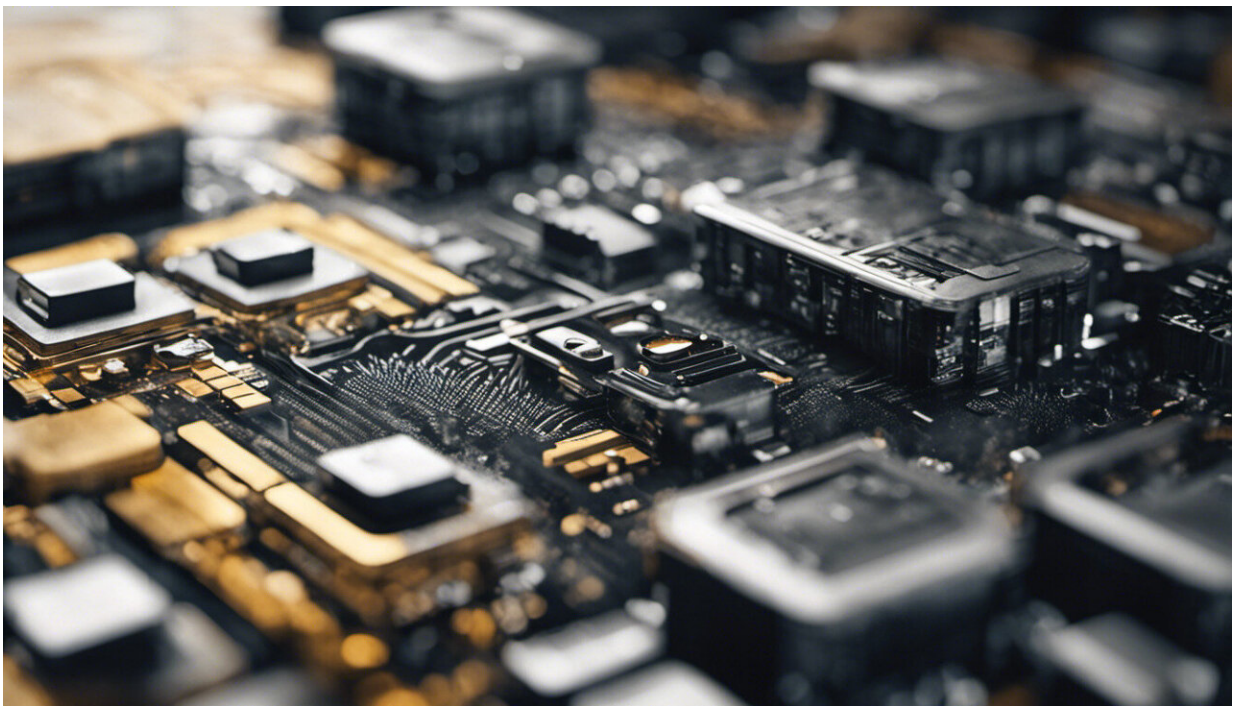# X removing the 'block' feature: Why businesses are better than people at moderating negative comments

August 30 2023, by Denitsa Dineva



Credit: AI-generated image ([disclaimer](#))

In a [recent post](#), the owner of X, (formerly Twitter), Elon Musk, announced his plans for the social media platform to [remove its blocking feature](#), except for in direct messages.

Users are concerned that this change in the platform's content moderation will lead to a rise in hostile and abusive content, leaving those on the platform unable to protect themselves from its consequences.

It is not only social media users who rely on X's blocking feature to control the content they see and interact with. Companies and brands with official social media accounts also depend on built-in moderation features. This ensures their fans and followers engage in positive and civil interactions.

Businesses need to be able to encourage constructive discussions on their social media accounts. This helps them build relationships with customers, increase word-of-mouth referrals and improve sales. Hostile online content directed at a company is not helpful to these business goals.

With the "block" feature significantly limited, companies and individual users seeking to control the spread of hostile online content will be forced to resort to other self-moderation. While individual users are sometimes used by companies to help moderate content, our recent research shows that official company accounts are much better placed to de-escalate hostile content.

## The importance of blocking

Research shows that, when social media users are exposed to offensive or abusive content online, it can lead to an array of negative consequences. They may experience mental distress and anxiety similar to that resulting from harassment that happens in person.

According to the same work, when presented with hostile or offensive content on social media, users are also likely to experience negative

[emotions](#) and refrain from interacting with others. For businesses, this can lead to [negative attitudes towards the company](#), that could also spread, and loss of trust in the brand.

Mute, report and block are built in features on most social media platforms. These enable users to restrict content they are exposed to, as well as who can interact with their profile. These features allow users to enjoy the benefits of social media such as [following trends, staying informed and interacting with others](#). They also allow users to avoid being targeted by offensive or unwanted content.

Mute and report are two features that can still be used to moderate hostile content. But these only partly address the issue, since they do not stop harassers from interacting with social media users or stalking their profiles. Blocking is arguably the most effective platform moderation feature. It gives users full control over who and what content they interact with on social media.

Blocking is not only a desirable moderation feature; it is a responsible business practice requirement. To prevent abusive and offensive content, both the App Store and Google Play Store policies state that the "[ability to block abusive users from the service](#)" and providing "[an in-app system for blocking UGC (user generated content) and users](#)" are necessary conditions for all of the applications they list.

In the UK, the current version of the [Online Safety Bill](#) will require [social media platforms](#) to offer adults appropriate tools to stop offensive or abusive content from reaching them. This is typically enabled by the blocking feature.

## Content moderation going forward

Reporting and hiding hostile content could be viable moderation options

for X going forward. It is, however, unlikely that these will be sufficient on their own. Removing the "block" feature could mean both companies and individual users have to take on increased responsibilities for content moderation.

Some research has already demonstrated that official business accounts employ diverse moderation communications—beyond just censorship—in the presence of hostile content. And this can be suitable for improving business followers' attitudes and its image.

Another way forward would be to rely on individual social media users for moderation, particularly prominent accounts that distinguish their status from others through digital badges such as the "blue check" on X and "top fan" on Facebook.

This is because research shows that digital badge accounts actively and positively participate in discussions and user-generated content on social media. As a result, these accounts could act as informal and occasional moderators.

Our research looked at whether official business accounts or prominent individual accounts were best at moderating hostile content on social media. In our first experiment, we presented participants with two scenarios. In one, an official business account moderated hostile content. In another, a digital badge user account intervened in a hostile interaction.

We found that companies are rated as the more credible moderators. They were best suited to de-escalate hostile content without the need to hide or remove it, or block the accounts involved.

On social media, it is common for moderation interventions to receive reactions and responses from observing users who support or disagree.

The presence of reactions likely influences how the moderator is perceived. To this end, in our second experiment, we studied the appropriateness of the moderator depending on whether the account received positive or negative reactions from other users who had observed the interaction.

Participants were given four scenarios: two in which the moderation intervention by the official business account received positive or negative emojis and two where a digital badge user account moderated the hostile interaction and this received either positive or negative emojis.

Our findings again confirmed that company accounts were seen as most appropriate for hostile content moderation by other social media users. This is the case even when the moderation receives negative reactions from the business account followers. Digital badge user accounts, in contrast, are only seen as credible when their moderation receives positive reactions from those following the company's social media account.

Whether or not the "block" feature on X is removed, moderating offensive and abusive content should not be left to businesses and individual accounts alone.

Content moderation should be a collective effort between platforms, businesses and individual users. Social media companies have the responsibility to equip their users with design features and tools that allow them to enjoy their platforms.

This article is republished from The Conversation under a Creative Commons license. Read the original article.