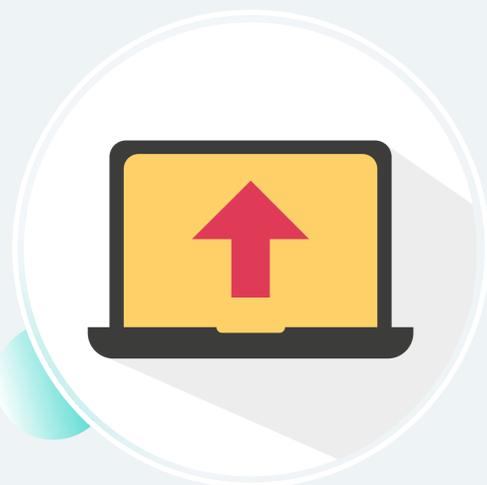


# How researchers used Prolific to analyze social media content moderation



In a collaboration between the Technical University of Munich and the University of Oxford, these researchers aimed to answer the question, when is content on social media perceived as offensive enough to warrant moderation?



## The task

Although social media platforms do currently impose limits on what can be posted, we still know very little about user opinions on offensive language or topics. In these studies, the researchers wanted to measure how users perceive and react to offensive speech online, and whether they would call for any moderation.

The moderation tactics ranged from mild (suppressing the visibility of the post) to extreme (suspension of the account in question). The researchers wanted to track what kind of content crossed the boundary of what is deemed acceptable, analyzing if there are differences in opinions between different groups in society.

Particularly with the recent news of Elon Musk's takeover of Twitter, content and account moderation/restriction has been a hot topic. This means that this first-of-its-kind research has never been more relevant and important.



## The challenge

One of the first challenges that this research encountered was the decision to focus exclusively on US participants. This meant that their research facilitator would have to segment a large pool of participants this way, while also ensuring that no one could take their studies more than once.

A secondary challenge was that, as this is first-of-its-kind research, the resulting data had to be of very high quality. But the research facilitator had to be cost-efficient and easy to use, as this research would involve multiple randomized experiments.



## The solution

Prolific's large data pool and easy segmentation proved a perfect solution for both of these challenges. Researchers were able to easily pick only US participants while also further segmenting them by any other factors needed.

Attention checks ensured high data quality, and researchers praised Prolific's participant pool for fully understanding and engaging with the research.

**"We had some researchers in our team who had experience with Prolific [...] so we knew that data quality was high"**

Franziska Pradel – Postdoctoral Researcher at the Technical University of Munich

Also, Prolific proved to be extremely fast and easy to integrate with their chosen survey creator tool (Qualtrics). This meant that as soon as funding was available, the research could be launched instantly.



## The results

Overall, the demand for content moderation was relatively low, even for more severe instances of offensive language or content. Seeing offensive speech did seem to trigger people, but the post was almost always allowed to stay up rather than being removed from the platform. Instead, participants seemed more comfortable with milder forms of moderation, such as warning flags or suppressing the post.

This was even the case when a minority group was threatened. Only about 50% of participants called for a post to be taken down or the posting user to be banned when the LGBTQIA+ group was targeted in a threatening manner.

In terms of next steps, the researchers are planning to reprise the study, expanding it to analyze factors such:

- Political ideology
- Attitudes towards freedom of speech
- Different countries

This will allow for more global results, analyzing whether opinions differ as a result of location. Findings are available at <https://osf.io/y4xft/> and described in a [Washington Post article](#).