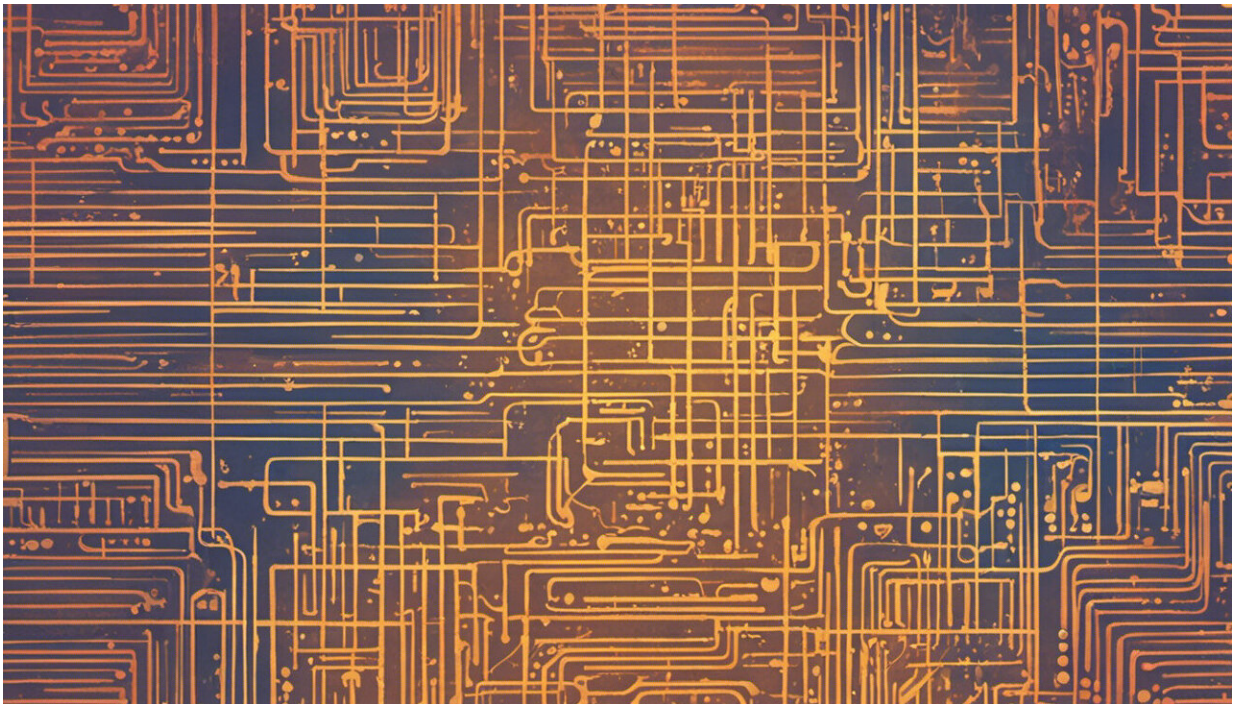


Facebook's moderation rules prove it's OK with being a hostile place for women

May 24 2017, by Amy Binns



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

Every day, thousands of [social media moderators](#) scroll through feeds of threats, pornography, animal and child cruelty, car crashes and bloody beatings in order to decide what is acceptable and what must be removed. And the [leaking of Facebook's training manuals](#) means we now know what standard they are working to, allowing content most users

would find abhorrent.

But removing as little as possible in order to protect free speech, built into the American DNA of most social media sites, perversely means others are likely to be invisibly silenced. By allowing threatening language aimed disproportionately at [women](#), Facebook is effectively accepting that they may be forced out by hostility.

Moderating huge, fast-moving social media sites is a major headache for their owners. After taking an initial light-touch approach to moderation, these sites are clearly [anxious to solve the problem](#) of controversial content. But without their own robust policies, they are at risk of [tempting increasingly impatient governments](#) to bring in regulation.

Facebook's hundred-plus training manuals cover everything from cannibalism to match-fixing and show the company accepts the need for moderation, but its libertarian values mean it will still allow content that no mainstream media outlet would tolerate for a moment.

Facebook classes aspirational threats such as "I hope someone kills you" as acceptable. Under a section on "credible violence", it also permits conditional threats such as "unless you stop bitching I'll have to cut your tongue out" or "little girl needs to keep to herself before daddy breaks her face", to take a couple of examples from its guidelines. It also allows calls for action such as "to snap a bitch's neck, make sure to apply all your pressure to the middle of her throat". It even allows direct threats if no specifics such as timings or the target's whereabouts are included.

These rules are wrapped in caveats and exclusions. For example, celebrities are given more protection, as are vulnerable groups such as the homeless. The sheer volume and complexity of the rule-book means that the moderators themselves are reported to be struggling with the concepts, let alone the massive volume.

But what doesn't seem to be considered is that allowing such violent language in the name of freedom of speech for some creates a hostile environment for others. This slowly leads to the withdrawal or forced exclusion of the targeted groups, particularly outspoken women and minority groups.



Window on a hostile world. Credit: Kaboompics

There's [20 years' worth of research](#), going back to the days of [listserv email mailing lists](#) that shows that [women fall silent](#) when encountering misogynistic abuse online. Of course, this is not limited to the internet. The tedious debate about women speaking in church is still rumbling on,

2,000 years after St Paul [set out his views](#) in his New Testament letter to the Corinthians.

But it is particularly important if this hostile environment is allowed to flourish on the social media sites that pride themselves on being the great debating chambers of our age. If women are forced off these platforms, then it has a significant impact on which issues we consider important, and by extension, democracy itself.

The internet also allows types of abuse that we haven't seen before. Facebook is particularly struggling with [revenge pornography and sextortion](#), reportedly trying to assess nearly 54,000 potential cases in a single month. Research suggests women are more likely to be made afraid by revenge porn.

New issues

But even more frightening issues include "doxxing" and "swatting", threatening real violence for minor grudges. Doxxing is the simple but widespread practice of publishing personal details about a target such as their home address, bank statements, [social security numbers](#) and anything else the uploader can find, with the implicit message, "do with these what you will".

Swatting is the more disturbing practice of calling the police and claiming there is a gunman holding hostages at the target's house. This became a prank in the US gaming community, but has also been used [against British women](#) such as Mumsnet founder Justine Roberts.

The most depressing part of all this is that the people most likely to be targeted, and so likely to disappear from online communities, are those who are most likely to have something to say which challenges stereotypes. A [major Guardian review](#) of comments blocked by

moderators on its site found eight of the ten writers most likely to attract abuse were women, four white and four non-white. The other two were black men.

Facebook seems to have decided that freedom of speech means the lightest possible touch of moderation that doesn't turn off its increasingly mainstream users, or lead to calls for government regulation. But freedom of speech for some comes at the cost of the absence of others.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: Facebook's moderation rules prove it's OK with being a hostile place for women (2017, May 24) retrieved 13 May 2024 from <https://phys.org/news/2017-05-facebook-moderation-hostile-women.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.