

Banning words on Instagram doesn't help—it makes it worse

March 9 2016



Stevie Chancellor is a doctoral student in Human Centered Computing.

A new Georgia Tech study finds that Instagram's decision to ban certain words commonly used by pro-eating disorder (pro-ED) communities has produced an unintended effect. The use of those terms decreased when

they were censored in 2012. But users adapted by simply making up new, almost identical words, driving up participation and support within pro-ED groups by as much as 30 percent. The Georgia Tech researchers found that these communities are still very active and thriving despite Instagram's efforts to moderate discussion of the dangerous lifestyle.

People in pro-ED communities share content, and provide advice and support for those who choose eating disorders, such as anorexia or bulimia, as acceptable and reasonable ways of living. They use specific hashtags to form very connected groups, often using anonymous names to keep their lifestyle choice a secret from the families and friends.

Instagram banned some of the most common pro-ED tags four years ago. People can still post these censored [terms](#), but the [words](#) no longer show up in search results. Banned examples include "thighgap," "thinspiration" and "secretsociety." Other pro-ED words received advisories. They can be searched, but notifications about graphic content were added, along with public service links for people looking for help.

The Georgia Tech researchers looked at 2.5 million pro-ED posts from 2011 to 2014 to study how the community reacted to Instagram's content moderation.

"People pretty much stopped using the banned terms, but they gamed the system to stay in touch," said Stevie Chancellor, a doctoral student who led the study. "'Thinspiration' was replaced by 'thynspiration' and 'thynspo.' 'Thighgap' became 'thightgap' and 'thygap.'"

The 17 moderated terms morphed into hundreds of similar, new words. Each had an average of 40 variables. Some had more: the researchers found 107 variables of "thighgap."

Instagram's censorship polarized the pro-ED community.

"Likes and comments on these new tags were 15 to 30 percent higher compared to the originals," said Munmun De Choudury, assistant professor in Georgia Tech's School of Interactive Computing. "Before the ban, a person searching for hashtags would only find their intended word. Now a search produces dozens of similar, non-censored pro-ED terms. That means more content to view and engage with."

The team also found that the content on these so-called lexical variants discussed self-harm, isolation and thoughts of suicide more often than the larger community of sufferers of eating disorders.

Instagram has also blacklisted words related to sex, racism and self-harm. What is more effective than banning tags? The Georgia Tech team suggests a few alternatives.

"Allow them to be searchable. But once they're selected, the landing page could include links for help organizations," said Chancellor.

"Maybe the search algorithms could be tweaked. Instead of similar terms being displayed, Instagram could introduce recovery-related terms in the search box."

The study, "[#thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities](#)," was presented at the ACM Conference on Computer-Supported Cooperative Work and Social Computing on March 1 in San Francisco.

Provided by Georgia Institute of Technology

Citation: Banning words on Instagram doesn't help—it makes it worse (2016, March 9) retrieved 1 May 2024 from <https://phys.org/news/2016-03-words-instagram-doesnt-helpit-worse.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.