THE CATALYST

Not a friend, but an algorithm

When online discussions of mental health are curtailed, people suffer. Why advocates are calling for social media platforms to be more accountable for their content safety practices

By: Moira Farr | Posted: September 22, 2026

“We built our lives on these platforms. We came out on them, we found each other on them, we organize, we mourn, we store our data … we make them rich, and in return, they shadow ban us…. That’s not a friend, that’s an algorithm.”

David Begor, a U.S-based podcaster and online advocate for 2SLGBTQI+ rights, addresses his audience with a look of hurt seriousness as he says these words. He’s referring to the practice of shadow banning, which is when social media platforms set algorithms to limit certain words from gaining reach online. Words like “queer or trans,” says Begor in his May 1, 2026, podcast, can mean posts aimed at seeking help or helping others go nowhere, while anti-trans content faces no such ban.

Talking mental health: Bad for mental health?

Social media shadow bans have also been applied to words such as depression, self-harm, eating disorders, and suicide. Policies vary across different platforms. One social media company, for instance, favours community moderation of mental health discussions, while another lays out its community standards policies for content related to suicide, self-harm, and eating disorders, including restrictions for users under the age of 18. All policies are no doubt intended to ensure that users do not share information that could harm others, but when they are applied without nuance, or with a political motive (Begor points out that one of the social media companies instituted shadow bans and outright censorship on 2SLGBTQI+ content when the new administration took charge of the U.S. government in 2025), positive discussions among people struggling with mental illness may also get scooped up in the algorithmic net.

For Begor, it meant leaving his accounts on two social platforms behind in protest of practices he and many others believe do more harm than good to equity-deserving communities.

The burdens of being banned

Evidence suggests the psychological impact of being shadow banned is serious. “Non-transparency of the social media platforms causes distress to individuals, and it needs to be addressed urgently,” write communications experts Sobi Thomas and Paul Manalil in their paper entitled “Digital silence: the psychological impact of being shadow banned on mental health and self-perception” published in the journal Frontiers in Psychology in October 2025.

Thomas and Manalil cite several studies that document and measure the effects of shadow bans on social media users. “When a user is shadow-banned, they are systematically excluded from the social world. The shock invisibility disrupts emotional regulatory protocols and can induce

depression symptoms, anxiety, and compulsive checking of content behaviours (Wikman et al., 2022)."

If the content that is shadow banned concerns mental health, these kinds of negative effects may be magnified for users. That may be especially true for young people looking for information and support as they struggle to cope with and understand their own mental health. "Social media policies come from a place of wanting to help, but we don't know what evidence they're based on. We don't have good data from the platforms themselves. We need a careful examination of the positive and negative impacts on users," says Dr. Julie Robillard, associate professor of neurology at the University of British Columbia, scientist in patient experience at BC Children's Hospital and BC Women's Hospital, and leader of the Neuroscience, Engagement and Smart Tech (NEST) lab with Neuroethics Canada. In a 2023 article entitled "Debate: Social media content moderation may do more harm than good for youth mental health" published in the journal *Child and Adolescent Mental Health*, she and her co-authors state, "Current evidence does not offer strong support for the relationship between censoring mental health content and preventing harm." They further suggest that social media moderation policies "can perpetuate negative consequences for mental health by creating isolated and polarized communities where at-risk youth remain exposed to harmful content." Robillard points out that rural youth, for example, may have nowhere to turn for help except social media; if they can't find it, they may feel even more isolated. She and her co-authors write that their article is "a call to action for evidence-based social media policies that engage youth mental health practitioners in their development." Robillard adds that young people themselves should be involved in social media policy making. "They know a lot about the platforms and are well aware of what is helpful and what isn't."

If social media platforms lack sufficient transparency about how they arrive at their moderation policies, especially about sensitive subjects like suicide, other sources of information are known to base their guidelines on evidence gained through research and the input of users. They also continue to gather and publish helpful data*. #ChatSafe: A young person's guide to communicating safely online about self-harm and suicide* is one such resource, aimed at young people age 12 to 25.

A first in the world, the #ChatSafe guidelines were developed and published in 2018 by Australian researchers and young people who wanted to engage in positive mental health discussions and avoid harmful content. They are available in versions designed for users in Canada and several other countries (the Mental Health Commission of Canada supported the project and was involved in adapting the guidelines for a Canadian audience, in both English and French). The guidelines also include resources for parents, educators, and community leaders, as well as links to the research concerning online safety surrounding suicide behind #chatsafe's own practices.

The guidelines give specific advice — but not rigid rules — on a variety of issues, including how to avoid stigmatizing language, myths and facts about suicide, what constitutes unsafe content, and, of course, how and where to get help for yourself or another person who you think needs it. #ChatSafe has its own social media accounts and provides contact information for the help centres of 13 social media platforms.



Elise Carrotte

Because Australia restricts social media for people under the age of 16 (similar legislation is pending in Canada), and at least one social media company restricts discussions of suicide and self-harm among those under age 18, the creators of #ChatSafe find other ways to reach younger people. “They can still use the internet, and we encourage them to find alternative ways to get support,” says Dr. Elise Carrotte, a research fellow in the Suicide Prevention Research Unit at the Centre for Youth Mental Health (Orygen) at the University of Melbourne, Australia. In Canada, these resources include 9-8-8 Suicide Crisis Helpline, Kids Help Phone, Hope for Wellness Helpline (for First Nations, Inuit, and Métis youth), and the LGBT YouthLine.

Recent data collected by researchers at Orygen show that the #ChatSafe project has reached 12 million views, which Dr. Carrotte and her colleagues see as a measure of success. But without transparency on the part of social media platforms, she says, it’s difficult to know if any shadow bans are at play, or how this level of growth compares with that of organizations promoting something other than suicide prevention.

Like Canada’s Dr. Robilliard, Dr. Carotte would like to see social media platforms be more forthcoming about their moderation practices, the extent to which their approach is evidence based, and the data they have gathered, if any, to determine user behaviour in the mental health space. She would also like to see young people included in the platforms’ moderation discussions, as they are with #ChatSafe.

Will social media platforms heed the advice of mental health researchers and suicide prevention experts? It is undeniable that there is growing public discontent with “big tech” and its seeming insensitivity to the human needs of users. One of the social media companies has recently been engaged in court battles and reached settlements (with more to come) with the families of young people whose mental health the families argued had been negatively affected by social media use; in some cases, the young people died by suicide. As a result of these developments, that company may have to provide transparency about its mental health moderation policies in the changes it says it will make. Without greater trust in social media companies’ practices, more users may choose, like David Begor and others who feel marginalized and silenced, to leave the platforms in search of different ways to connect with others.

Resources, sources, and documents

Dave Begor's Podcast – May 1, 2026

TechTimes

Reddit – r/Mental Health

Meta – Community Standards

Frontiers in Psychology

Neuroethics Canada

Debate: Social media content moderation may to more harm than good for youth mental health

#ChatSafe: A young person's guide to communicating safely online about self-harm and suicide

#ChatSafe for parents and carers

The research behind #ChatSafe

Ottawa moves to restrict social media for kids under 16 – CBC

Tech backlash reaches fever pitch – CNBC

Social platform agrees to make changes, pay up to $18B US over social media harms to children – CBC

Author: Moira Farr. An award-winning journalist, author, and instructor, with degrees from Ryerson and the University of Toronto. Her writing has appeared in The Walrus, Canadian Geographic, Chatelaine, The Globe and Mail and more, covering topics like the environment, mental health, and gender issues. When she's not teaching or editing, Moira freelances as a writer, having also served as a faculty editor in the Literary Journalism Program at The Banff Centre for the Arts.

Mental Health Commission of Canada
https://mentalhealthcommission.ca/
350 Albert Street, Suite 1210 Ottawa ON K1R 1A4
Tel: 613.683.3755 | Fax: 613.798.2989
mhccinfo@mentalhealthcommission.ca