

Down-ranking polarizing content lowers emotional temperature on social media – new research

Tiziano Piccardi, Assistant Professor of Computer Science, Johns Hopkins University

Published: December 4, 2025 8:18am EDT



Social media posts that stoke division don't have to top your feed.

Gama5/iStock via Getty Images

Reducing the visibility of polarizing content in social media feeds can measurably lower partisan animosity. To come up with this finding, my colleagues and I developed a method that let us alter the ranking of people's feeds, previously something only the social media companies could do.

Reranking social media feeds to reduce exposure to posts expressing anti-democratic attitudes and partisan animosity affected people's emotions and their views of people with opposing political views.

I'm a computer scientist who studies social computing, artificial intelligence and the web. Because only social media platforms can modify their algorithms, we developed and released an open-source web tool that allowed us to rerank the feeds of consenting participants on X, formerly Twitter, in real time.

Drawing on social science theory, we used a large language model to identify posts likely to polarize people, such as those advocating political violence or calling for the imprisonment of members of the opposing party. These posts were not removed; they were simply ranked lower, requiring users to scroll further to see them. This reduced the number of those posts users saw.

We ran this experiment for 10 days in the weeks before the 2024 U.S. presidential election. We found that reducing exposure to polarizing content measurably improved participants' feelings toward people from the opposing party and reduced their negative emotions while scrolling their feed. Importantly, these effects were similar across political affiliations, suggesting that the intervention benefits users regardless of their political party.

Why it matters

A common misconception is that people must choose between two extremes: engagement-based algorithms or purely chronological feeds. In reality, there is a wide spectrum of intermediate approaches depending on what they are optimized to do.

Feed algorithms are typically optimized to capture your attention, and as a result, they have a significant impact on your attitudes, moods and perceptions of others. For this reason, there is an urgent need for frameworks that enable independent researchers to test new approaches under realistic conditions.

Our work offers a path forward, showing how researchers can study and prototype alternative algorithms at scale, and it demonstrates that, thanks to large language models, platforms finally have the technical means to detect polarizing content that can affect their users' democratic attitudes.

What other research is being done in this field

Testing the impact of alternative feed algorithms on live platforms is difficult, and such studies have only recently increased in number.

For instance, a recent collaboration between academics and Meta found that changing the algorithmic feed to a chronological one was not sufficient to show an impact on polarization. A related effort, the Prosocial Ranking Challenge led by researchers at the University of California, Berkeley, explores ranking alternatives across multiple platforms to promote beneficial social outcomes.

At the same time, the progress in large language model development enables richer ways to model how people think, feel and interact with others. We are seeing growing interest in giving users more control, allowing people to decide what principles should guide what they see in their feeds – for example the Alexandria library of pluralistic values and the Bonsai feed reranking system. Social media platforms, including Bluesky and X, are heading this way, as well.

What's next

This study represents our first step toward designing algorithms that are aware of their potential social impact. Many questions remain open.

We plan to investigate the long-term effects of these interventions and test new ranking objectives to address other risks to online well-being, such as mental health and life satisfaction. Future work will explore how to balance multiple goals, such as cultural context, personal values and user control, to create online spaces that better support healthy social and civic interaction.

The Research Brief is a short take on interesting academic work.

This research was partially supported by a Hoffman-Yee grant from the Stanford Institute for Human-Centered Artificial Intelligence.

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