

How Social Media Algorithms Manipulate Users and Push Harmful Content

Marlen Vining

February 16, 2026

Abstract

Social media platforms primarily generate revenue from the collection of user data and targeted advertising. This is why it is in their best interest to keep users engaged as long as possible. To do this, algorithms are designed to serve content to users that will keep them using the platform. Many of these algorithms purposefully serve content to users that is designed to trigger strong emotional responses. This is done to extend the time users spend on the platform and create a dependence. This often has a direct negative impact on the mental health and well being of users. The goal of this paper is to understand the tactics used by these platforms and uncover the harm it causes.

How Social Media Algorithms Work

Social media platforms use algorithms to show users content that will keep them engaged with the platform. This is done by analyzing the habits of users to create a profile of what content is most effective at keeping them on the site. "The reason why social media platforms use algorithms is to more organically filter through the large amount of content that is available on each platform. Algorithms do the work of delivering content that is potentially more "interesting" for a user to the detriment of posts which are deemed irrelevant or low-quality - either in general, or to a specific user."² These algorithms are typically designed to prioritize engagement over truth and what is healthy for a user to see.

Why Algorithms Push Harmful Content

Social media sites have a financial incentive to keep users on the platform as long as possible. These sites are also incentivized to push content that increases the effectiveness of advertising on the platform. "Algorithms designed to enhance user engagement have significantly shaped how content is prioritized, recommended, and disseminated. While these algorithms aim to create a personalized experience, they often operate without sufficient safeguards against the spread of harmful content, including racial and ethnic misinformation."⁴ Content that produces a strong emotional response in the user, as well as content that builds a dependence on the platform keeps users engaged longer and creates a pattern of repeated use. Because of this, algorithms are often designed to push these kinds of contents. While this may be an unintentional result, it can still cause real harm.

The Impact On Users

The way social media algorithms work can lead to real world harm if vulnerable users are shown the wrong kind of content. In a study done by Meta to measure the impacts of certain content on teens, "the 223 teens who often felt bad about their bodies after viewing Instagram, "eating disorder adjacent content" made up 10.5% of what they saw on the platform. Among the other teens in the study, such content accounted for only 3.3% of what they saw."³ These algorithms create feedback loops that can trap users in a cycle of negative content. This can have a serious detrimental effect on the user's mental health. Because algorithms are designed to recommend content based on user behavior, the more users interact with these kinds of posts, the more they will see. When a French court investigated the platform Tik Tok over concerns that it was having a negative psychological effect on youth, it was noted that, "insufficient moderation of TikTok, its ease of access by minors and its sophisticated algorithm, which could push vulnerable individuals toward suicide by quickly trapping them in a loop of dedicated content."¹

Conclusion

This is a complex issue that cannot be solved easily. On one hand, social media platforms have an ethical responsibility to not harm their users. On the other hand, with millions of accounts posting daily, it is impossible for moderation to be perfect. For many of these companies, profit is the primary concern and user safety is often overlooked. Because of this, the best way to avoid harmful content is to understand the types of content that may be harmful to you and do your best to avoid them. Information presented in social media posts should not be trusted until verified by a reputable source.

References

- [1] DE CLERCQ, G. French court probes tiktok for algorithms pushing to suicide. *Reuters* (Nov. 2025). Accessed 2025-02-16.
- [2] GOLINO, M. A. Algorithms in social media platforms. *Institute for Internet and the Just Society* (2021).
- [3] HORWITZ, J. Instagram shows more "eating disorder adjacent" content to vulnerable teens, internal meta research shows. *Reuters* (Oct. 2025). Accessed 2025-02-16.
- [4] LUCAS, W. The influence of social media algorithms on racial and ethnic misinformation: Patterns and impacts. *ResearchGate* (Dec. 2024).