



July 27, 2023

Social Media Algorithms: Content Recommendation, Moderation, and Congressional Considerations

Social media plays an integral role in modern life for many. It facilitates the spread of information and serves as a key source of news, entertainment, and financial opportunity. In 2022, over 70% of Americans received some of their news from social media, according to the Pew Research Center. Recently, social media companies have faced criticism for potentially enabling the spread of harmful content, suppressing certain viewpoints, contributing to social polarization and radicalization, collecting and monetizing personal data, and adversely affecting children. As part of broader discussions around social media, some stakeholders and policymakers have taken interest in legislative proposals to regulate or address “social media algorithms.”

This In Focus provides a high-level overview of content recommendation and moderation algorithms employed by social media platforms. It examines issues that arise from the use of social media algorithms and discusses considerations for Congress.

Overview of Social Media Algorithms

Social media companies use a number of algorithms and artificial intelligence (AI) systems to recommend or moderate content on their platforms and perform a variety of other functions. Algorithmic *recommendation systems* sort, curate, and disseminate content deemed relevant to specific users. Algorithmic *content moderation systems* are often used, along with human moderators, to identify and restrict illegal material and content that violates a company’s policies and terms of use and service. Social media companies may also use algorithms for other purposes, such as targeting and delivering digital advertising or providing in-app search functions.

Due to definitional ambiguity, algorithms are often conflated with a variety of different technologies and applications. For example, “algorithms” are often colloquially used to refer to “artificial intelligence,” but the two terms are not synonymous. Certain algorithms may fall under the broad category of AI (such as machine learning algorithms), while others do not use the predictive or data-mining techniques that are characteristic of AI. Due to this definitional ambiguity, some scholars and policymakers have opted to instead use the language of “automated decision-making systems” or “automated systems” in broader policy discussions—such as in the White House’s Blueprint for an AI Bill of Rights. Additionally, some scholars and policymakers instead focus on specific outcomes and impacts regardless of what technology or technologies are implicated—such as discriminatory or disparate impacts. Congress may consider what language and terms are best suited for legislation targeting social media, or other technologies.

Definitions

- **Algorithm:** A specific process or sequence of computational steps followed by a computer in performing a task or problem-solving operation. Algorithms vary in complexity depending on the task and context.
- **Recommendation Systems:** Systems that use algorithms to personalize the sorting, ranking, and displaying of content for a user based on their previous engagements and other collected data. Also called recommendation engines or recommendation algorithms.
- **Moderation Systems:** Systems that use algorithms to identify, filter, and flag undesirable or illegal content for removal, demonetization, downranking, or other forms of moderation.

Section 230 and Algorithms

Section 230 of the Communications Act of 1934 (47 U.S.C. §230), enacted as part of the Communications Decency Act of 1996, broadly protects providers of interactive computer services from liability for information provided by a third party and content moderation decisions. There has been debate whether Section 230 liability protections should extend to the use of recommendation algorithms. So far, courts have held that recommendation algorithms are protected under Section 230.

In 2023, the Supreme Court declined to weigh in on the Section 230 issue in two cases: *Twitter v. Taamneh*, No. 21-1496, and *Gonzalez v. Google LLC*, No. 21-1333. Both cases considered whether social media companies could be liable for recommending terrorist content. The Court did not rule on whether Section 230 granted immunity to the companies’ recommendation algorithms. Instead, it concluded that the companies—regardless of Section 230—were not liable under the relevant antiterrorism federal statute because their conduct did not amount to aiding and abetting an act of international terrorism. For more information on Section 230, see CRS Report R46751, *Section 230: An Overview*, by Valerie C. Brannon and Eric N. Holmes.

Issues and Concerns

Algorithms are a key component of social media platforms. They help sort, moderate, and disseminate massive volumes of user-generated content to individuals. This in turn facilitates targeted digital advertising, a major source of revenue for social media platforms. However, policymakers, stakeholders, and researchers have raised concerns about their use.

Algorithmic Amplification of Harmful Content

Many social media platforms use algorithms to recommend content to maximize user engagement—measured through “likes,” time spent on the platform, reposts, and other metrics. However, there is debate whether these systems thereby increase the spread of, or amplify, harmful content, often called *algorithmic amplification*. There is concern that social media algorithms may amplify harmful content, create echo-chambers (or filter bubbles) that may contribute to user radicalization and polarization, or drive social media addiction in children.

Algorithmic amplification on multiple social media platforms has been a recent topic of inquiry and research. While amplification effects are difficult to measure by outside researchers without access to social media platforms’ proprietary data, some recently released company research supports some of these concerns. In 2021, a former Facebook employee leaked company documentation that revealed it weighted interactive elements known as “reactions” more than “likes” in content ranking. By prioritizing content that received emotional reactions (such as an angry reaction) rather than likes, critics believe the company amplified divisive or sensational content. Other Facebook documents reviewed by the *Wall Street Journal* in 2020 found that 64% of users who joined extremist groups on Facebook’s platform did so “due to [Facebook’s] recommendation tools.” According to the Mozilla Foundation’s “YouTube Regrets” report, 12% of content recommended by YouTube’s algorithms violates the company’s community standards.

Some experts also allege that hostile foreign actors can manipulate social media recommendation algorithms or skirt automated moderation systems to conduct influence operations and spread propaganda. For example, in 2021, the *New York Times* found that Chinese information campaigns utilized bot-like accounts to manufacture virality (the rapid spread of online content between users) through liking and reposting government and state media posts. The *Times* found that, “The contrived flurry of traffic can make the posts more likely to be shown by recommendation algorithms on many social media sites and search engines.” Recently, the popular video-sharing app TikTok has faced criticism for possibly amplifying propaganda and censoring content critical of the government of the People’s Republic of China.

Removal of Content

Some critics are concerned that social media platforms remove lawful speech and suppress certain viewpoints through moderation policies and automated moderation systems. Some groups and online communities contend their posts have been disproportionately flagged, removed, or downranked, meaning the platform adjusted their algorithms to make the content less visible or prominent. Companies use various automated demotion or reduction practices, which are difficult to measure through external research without access to social media companies’ proprietary systems, documentation, research, or internal policies that guide moderation decisions.

Other critics have expressed the concern that social media companies do not remove enough harmful content, and that automated moderation systems fail to adequately filter certain types of harmful content. For example, research has found that content moderation systems may be less effective addressing non-English content. This has led to claims that non-English misinformation and hate speech may be under-moderated and therefore more prevalent in certain online language communities. This may be due to a confluence of factors. Automated moderation systems may lack the necessary training data for a particular language; companies may not employ enough people fluent in a particular language to address language nuances or evolution; or companies may not provide sufficient resources for moderation in specific countries and languages. Failures in automated moderation may coincide with failures in company practices and policies—leading to inaccurate or undesirable moderation outcomes.

Congressional Considerations

Some Members of Congress have introduced legislation in the 117th and 118th Congresses to ban or significantly limit the use of recommendation algorithms. Some recently proposed bills would ban their use for children, restrict the use of personal data in recommendation algorithms, or require companies to provide disclosures or offer alternative versions of the platforms without algorithmic recommendations. If the 118th Congress considers restricting recommendation algorithms in certain contexts or for certain users, it may consider how to target interventions given the ubiquity of recommendation algorithms on other online platforms, such as search, marketplaces, and video and music streaming services.

The 118th Congress may also consider legislative approaches to increase the transparency of social media algorithms modeled on provisions in previously introduced bills. S. 5339 in the 117th Congress, for example, would have created disclosure requirements for social media platforms. Other recent bills would require third-party risk and impact assessments and audits of social media algorithms that could be submitted to agencies, such as the Federal Trade Commission, for review and investigation.

Congress could consider amending Section 230 to address perceived risks associated with algorithms. For example, bills—such as H.R. 2154 in the 117th Congress—would have removed Section 230’s protections for recommendation algorithms in certain lawsuits involving terrorism or civil rights. Amending Section 230 could have unintended consequences for existing online ecosystems, potentially by incentivizing platforms to over- or under-moderate content in an attempt to avoid legal jeopardy or by providing incumbent platforms that have the resources to fight legal challenges, an advantage over new market entrants that do not.

Kristen E. Busch, Analyst in Science and Technology Policy

IFI2462

Disclaimer

This document was prepared by the Congressional Research Service (CRS). CRS serves as nonpartisan shared staff to congressional committees and Members of Congress. It operates solely at the behest of and under the direction of Congress. Information in a CRS Report should not be relied upon for purposes other than public understanding of information that has been provided by CRS to Members of Congress in connection with CRS's institutional role. CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.