Social Media Regulation
How should we decide what speech is acceptable/unacceptable on social media?
About this issue guide

This guide was collaboratively produced by

- Dr. Ryan Kennedy, University of Houston Machine-Assisted Human Decision-making (MAHD) Lab Director
- Dr. Windy Y. Lawrence & Dr. John Rountree, University of Houston-Downtown Center for Public Deliberation (UHD CPD)
- UHD CPD Election Public Deliberation Team

Last updated 2023.
Introduction

Social media has a profound influence on public life. Indeed, social media certainly has had positive benefits for society, like uniting lost family members, helping people keep in touch with people who live far away, helping find organ donors, and encouraging people to acquire new skills. Additionally, some have also noted the positive benefits it can have on children and teens. For instance, the Houston Chronicle concluded, “Studies have shown that social media may support the well-being of LGBTQ youth by enabling them to connect with peers. And seven out of 10 adolescent girls of color reported that they encountered positive or identity-affirming content on social media, the advisory says.”

However, while we certainly can observe ways that our social media can be used for good, increasingly, many are alarmed by awareness of the increasing number of indicators of its harm.

- The usage of social media seems to be correlated with online harassment and increased division between people. A 2021 study by the United Nations found that online hate speech is increasing around the world.ii Pew Research Center survey, published in 2021, reported growing numbers of Americans saying they have experienced online harassment or abuse or have experienced it in increased severity. The same survey said half of those reporting online harassment felt it was because of their political views. Significant numbers also cited their sex or their racial and ethnic backgrounds as reasons why they believed they experienced this online harassment. Notably, the survey showed “fully 79% say social media companies are doing an only fair or poor job at addressing online
harassment or bullying on their platforms.”

- Also in 2021, *Nature* journal published its own survey reporting that “more than two-thirds of researchers reported negative experiences as a result of their media appearances or their social media comments, and 22% had received threats of physical or sexual violence.” Similarly, celebrities and political leaders from both sides of the political aisle have seen their social media profiles flooded with severe abuse and death threats.

- Indeed, we are witnessing increasing concerns with social media usage and its potential correlation with hate crimes. On October 27, 2018, a man entered the Tree of Life Synagogue in Pittsburgh during Shabbat services. He opened fire, killing 11 people and wounding six more, including several Holocaust survivors. The attack took place in the midst of rising anti-Semitism online. The attacker had participated in this activity, spending time and posting on extremist sites. One post just hours before the attack verbally attacked a Jewish group and concluded, “I’m going in.”

- In another story, social media usage was tied to severe vandalism. October 2021, a Tennessee school district reported 100 acts of vandalism and theft in just a few weeks. Bathrooms were damaged, halls were flooded, and dozens of students had to pay fines and restitution. The district blamed social media challenges, including one on TikTok that was called “devious licks.” TikTok said such a challenge would violate their policies—but they could not find the challenge on their platform.

- The relationship between social media use and mental health is also a top concern. Studies indicate there are links between overusing social media and depression, loneliness and a host of other mental problems. In Britain, teenagers now spend about an average of 18 hours a week on their phones, much of it using social media.

- U.S. Surgeon General Vivek Murthy, in May 2023, issued a public health advisory that says social media could harm the mental health and well-being of children and teenagers. According to a *Houston Chronicle* article by Evan MacDonald, “While the advisory acknowledges those risks are not fully understood, it said there are ‘ample indicators’ that social media can have a ‘profound risk of harm’ for them.”

These examples illustrate just part of an issue that many Americans, and those who represent us, are concerned about: social media content that may cause harm to individuals, communities, and potentially to our democracy.
Digging Into the Issue: Regulating Social Media

Critics of social media claim that content on these sites can cause harm in varied ways. They can be a tool for hate-based harassment. They can contribute to terror attacks and cause discord among people. They can threaten community health with false information. They can further polarization and hatred between various groups. They can influence young people to develop negative views of themselves and encourage addictive use, which distracts people from engaging in other activities outside of social media. These critics argue that more should be done—by the social media platforms or by the government—to keep such content off social media sites.

But the United States has a long history of protecting speech that is widely regarded as offensive. The First Amendment to the U.S. Constitution says “Congress shall make no law... abridging the freedom of speech, or of the press.” When the Constitution and the Bill of Rights were written and adopted, the United States was still emerging from the war for independence from Great Britain. Americans rebelled against Britain’s tyrannical rule and endless abuses, like religious persecution and taxation without representation. This history has contributed to a strong tradition of defending freedom of speech from government overreach.

In contrast, many European democracies take a different approach to free speech. Following the Holocaust, many European nations prioritized preventing something similar from ever happening again. Thus, many European nations prohibit hate speech. For example, in Germany, it is a crime to deny the Holocaust. It is also a crime to incite hatred against groups or to insult, slur, or defame them in a way that violates their dignity. A law passed in 2017 holds social media sites responsible if they do not remove hate speech within a week of its being posted.
As a result of U.S. experience with the power of hate speech and its symbols, some U.S. cities and states have tried to ban certain speech acts as hate crimes. The Supreme Court has not looked positively on such laws. The Supreme Court has ruled that laws cannot ban certain speech activities simply because they express despicable ideas. Indeed, the Supreme Court has shown great reluctance to limit speech based on content. Justice Anthony Kennedy said, “The history of the law of free expression is one of vindication in cases involving speech that many citizens may find shabby, offensive, or even ugly. It follows that all content-based restriction on speech must give us more than a moment’s pause.”

Despite the strong protection for freedom of speech, the Supreme Court has allowed the government to limit or punish certain kinds of speech. Speech that can be banned includes obscenity, slander/libel, and incitement to imminent lawless action. In addition, the Court has allowed certain limitations on the time, place, and manner, in which speech takes place.

How does all of this apply to social media? That question remains mostly open. Social media platforms are different from traditional publishers. One big difference is that the sites generally do not create the content or interact with the authors of that content. Rather, platforms simply provide a mechanism for getting the information and opinions out. Thus, some people argue that the platforms should not be responsible for the content on their sites.

Back in 1996—before social media had really come into play—Congress passed the Communications Decency Act. Section 230 of that law protects Internet companies from being held liable for what users share. But there’s an exception: the companies are responsible for copyright claims and child sex trafficking on their sites. Section 230 also allows companies to take down speech if they choose to without being sued for infringing on a user’s rights.

Section 230 has been the subject of much discussion in Congress. Some policymakers want to stop social media sites from banning politicians from their platforms. Others want to remove the protection from liability now afforded the companies. Of course, the companies have their own policies regarding acceptable content. Instagram summarizes its guidelines as follows (the full version is much longer):

“We want Instagram to continue to be an authentic and safe place for inspiration and expression. Help us foster this community. Post only your own photos and videos and always follow the law. Respect everyone on Instagram, don’t spam people or post nudity.”

Companies use artificial intelligence, thousands of employees, and outside companies to monitor and remove offensive content and fake accounts. In India, where social media companies are required to report on content they have removed, Instagram reported deleting almost 13,000,000 posts in one month. The categories of posts removed are shown in Table 1.
This data suggests that content moderation does happen. Yet, when whistleblower Frances Haugen testified before Congress, she revealed that Instagram knew that research showed its app made body image issues worse for teens, especially girls. The company’s leadership chose not to do anything about that problem.\textsuperscript{x}\textsuperscript{i}

While not illegal, inaction by social media companies, in the face of evidence of harm, strengthens some peoples’ beliefs that something needs to change. While some lobby for tighter government restrictions, others see the infringement of government on social media as a violation of our first amendment. Clearly, finding common ground will require careful deliberation.

\textit{How does this issue affect you, your friends and family, and your community?}

How should we decide what speech is acceptable on social media? Many different policy options have been proposed, but today, we might consider \textbf{three different options for how to address this issue}:

- Allow social media platforms to decide individually how they will regulate content. The government should not be involved in regulation.
- Develop standards and allow our government to make platforms legally liable for harmful content.
- Allow the government to hold a relatively small role, focusing on giving users more control over what they see. For example, the government could require that platforms make multiple algorithms available.
Option 1: Social Media Companies Should Self-Regulate

This option essentially means affirming the current situation. Social media companies would moderate the content on their sites, following their own standards.

As an alternative, the companies might agree to create an independent group to develop a set of industry standards. The group could include representatives of social media platforms, users, and perhaps even government officials. This approach has worked with other industries. It could create predictability and consistency across platforms.

Below are positions people might take on this option, presented as people who support this option might state them.
### Pros

The First Amendment applies to government action, not to actions by private companies. Thus, this approach is, without question, constitutional.

Companies say they are doing their best now. But many people are unhappy with hate speech, lies, and violence-inducing posts. Without some additional motivation, it’s unclear the companies will step up their game.

The government should only regulate business when there is a compelling reason to do so. The companies have the resources to make this approach work, so government action seems unnecessary. The companies already police child pornography, nudity, recruiting posts for terrorist organizations, and copyright violations. They have recently begun adding advisories for false information. The tools already being used would work similarly with new categories of posts to be removed.

Former President Donald Trump was banned from Twitter in January 2021 until different leadership was assumed. The platform cited his role in the January 6 riot as a key to their decision. But many people feel that his posts were very similar to posts he’d been making for years. Thus, some felt the decision was arbitrary. Belief that the companies’ decisions are arbitrary will undercut public support for this approach.

Companies want to know where the line between acceptable and unacceptable falls. They do not want to have to draw the lines themselves and having an independent group create standards would help them. At the same time, the government would not be making judgments about the content of speech. That should eliminate any constitutional issue.

Communities of color, women, LGBTQ+ communities, and religious minorities feel that they are at risk for over-enforcement of platforms’ standards. On the other end of the spectrum, conservative users also think they are discriminated against. Some members of these communities also think that self-regulation is a way for companies to pretend they are doing something rather than actually addressing the problems. More transparency will be necessary to make all users feel confident about self-regulation.

With this approach, if users are unhappy with the way social media platforms are moderating content, they know who to blame and how to respond. They can switch to another platform. There is no shortage of social media options out there. That’s motivation for companies that want to maximize profits.

We need more hard data on the effects of online hate speech and false information. The government should fund research on this topic before a final policy decision is made.

### Cons

Below are several policy actions that would be consistent with this option and their drawbacks.
<table>
<thead>
<tr>
<th>Action</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengthen current regulations to require large social media companies to remove illegal and pornographic content within four days, while letting sites regulate all other content as they see fit.</td>
<td>Content that qualifies as illegal only covers a small fraction of problematic content. This would do nothing to address issues of misinformation, harassment, or offensive content.</td>
</tr>
<tr>
<td>Form a social media association that would help standardize content regulation across platforms and allow sharing of information about best practices among social media companies.</td>
<td>Some social media sites may decide not to join the association or not to abide by their voluntary standards. Indeed, some sites (e.g. 4chan) might use their violation of these standards as a marketing tool to get more users.</td>
</tr>
<tr>
<td>Social media sites should form independent oversight boards to evaluate when a public official or celebrity has violated the community standards of the site. The board would make determinations about whether the user should be barred from the site and for how long.</td>
<td>The decisions of these boards can be highly controversial, as we have seen with the removal of former President Trump from Twitter. There are also questions about how independent these boards can really be, given the large amounts of money at stake from removing popular personalities from social media.</td>
</tr>
</tbody>
</table>
Option 2: Social Media Companies Should Be Legally Liable for Harmful Content

This option would give government a fairly large role in content regulation. Congress would have to revise Section 230 of the Communications Decency Act to permit holding companies liable for content posted on their sites. It would also set the standards for unacceptable content, although the companies would be responsible for identifying and removing the content.

Government would need to provide some monitoring of companies’ compliance with the standards. One option for doing so would be to create an agency that would develop the standards and monitor compliance, or to task an existing agency, like the Federal Communications Commission, with this work.
Below are **positions** people might take on this option, presented as people who support this option might state them.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>This approach will provide the social media companies with a standard they will need to meet. That is something they’ve been asking for. However, by making the companies responsible for enforcing the standards, it may avoid a First Amendment challenge. The First Amendment applies to government limitations on free speech, not limits placed on speech by private companies.</td>
<td>Since the government would be creating the standards, this approach could be challenged on First Amendment grounds. The Supreme Court has been very protective of free speech rights. It uses the highest level of scrutiny when looking at cases involving content-based suppression of speech. It has protected hate speech and false information in the past. There is a good chance this approach could be struck down by the Supreme Court.</td>
</tr>
<tr>
<td>Making the companies liable for content posted on their sites would provide a strong motivation for the companies to do a better job with content moderation. If they do a good job, this option shouldn’t be any more expensive than giving them the power to self-regulate.</td>
<td>Letting the government decide what speech is acceptable is getting on a “slippery slope”—the first step in allowing greater censorship. This flies in the face of our nation’s history and values. The nation’s philosophy has been that the solution to offensive or inaccurate speech is more speech, not censorship.</td>
</tr>
<tr>
<td>This approach could be adapted to be less coercive. For example, the government could say that, if a company meets the standards established, it would be protected from liability. Only if a company failed to meet the standards would it be liable for hate speech or dangerous false information.</td>
<td>Laws banning hate speech in Europe have not eliminated hatred or hate crimes. We need to look for more innovative ideas to get people talking to each other. We need to be creative and to practice direct, representative democracy online.</td>
</tr>
<tr>
<td>Our country is facing some serious problems. People are severely divided along partisan lines. Support for our democratic values seems to be wavering. Hate crimes have risen. It may be time to rethink the meaning of the First Amendment when it comes to hate speech and other harmful speech. We shouldn’t let our history trap us in an environment of hate.</td>
<td>More than half of Americans oppose making it possible to sue social media companies for what users post. Among all adults, 56% oppose making social media companies liable. Among social media users, the percentage is a bit higher—59%. While Republicans are more likely to oppose letting people sue the companies, Democrats are also opposed (60% vs. 52%).</td>
</tr>
</tbody>
</table>

---

\[xii\] We need to look for more innovative ideas to get people talking to each other. We need to be creative and to practice direct, representative democracy online.
Below are several **policy actions** that would be consistent with this option and their **drawbacks**.

<table>
<thead>
<tr>
<th>Action</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Require social media sites to adhere strictly to their own terms of service and community standards, with fines for violating these standards enforced by the Federal Communications Commission (FCC).</td>
<td>Social media sites may have incentive to weaken their standards to avoid regulation. The FCC (or other government institutions) may also have difficulty monitoring compliance. Requiring the FCC to monitor would be costly and increase the size of government which some people believe is already too large.</td>
</tr>
<tr>
<td>Allow government litigation and civil lawsuits against companies, CEOs, and other executives for violations of a social media site’s content standards.</td>
<td>This could result in too many frivolous lawsuits, hurting company profits. It may also incentivize companies to be overly aggressive in taking down content that could be controversial.</td>
</tr>
<tr>
<td>Create estimates of the amount of revenue generated by harmful content and misinformation. Tax or fine companies according to these estimates to remove the incentive to allow harmful content.</td>
<td>While some methods have been developed for estimating these revenues, they are likely to be disputed and companies may choose to fight the estimates rather than remove content that generates profitable engagement, no matter how harmful.</td>
</tr>
</tbody>
</table>
Option 3: Social Media Companies Should Be Required to Provide Alternative Algorithms

This option, rather than deleting content from social media platforms, would give users more control over what they see. After all, some people believe that the way we combat bad speech is with more speech, not through censoring speech. Algorithms control what appears in a user’s social media feed. Social media companies design algorithms to figure out how to increase their users’ engagement as much as possible and to manipulate user behavior for the purposes of selling ads. Social media users, in essence, are the product social media companies sell. Social media companies have an unprecedented amount of information about its users. Shoshana Zuboff, PhD, explains that social media companies can directly manipulate social media behavior and emotions without its users even being aware. Tens of millions of users are not only increasingly becoming more addicted, these algorithms, designed to manipulate, are increasing user emotionality and decreasing users’ ability to be rational about current issues.

While we know the impact of algorithms, how those algorithms work exactly is not made public. This option would require companies to be transparent about how their algorithms work. It would also require them to provide alternatives so users can decide how they want their content to be determined. Making algorithms more transparent might help users be more aware of how they are manipulated to form silos and like-minded groups and to become less likely to seek diverse perspectives.
Below are **positions** people might take on this option, presented as people who support this option might state them.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users of social media have not had much control over the content they see. That content is determined by algorithms. The algorithms are designed to keep users scrolling. Often, the algorithms push users to ever-more-extreme content. Users deserve to know how these algorithms work and to choose another way to determine the content they see.</td>
<td>Algorithms are the intellectual property of the companies that develop them. Transparency would violate their property rights. Offering additional algorithms would be an additional expense as well. And there’s no evidence that the companies would be able to develop algorithms that would work the way users want.</td>
</tr>
<tr>
<td>Algorithms are biased in a variety of ways. In 2020, for example, a Twitter algorithm that generated photo previews was cropping out African American faces. Such biases can be especially harmful to teens. Making algorithms more transparent and offering options can address this problem.</td>
<td>This approach assumes that users will opt for strong, positive content. But letting users control the content they see could result in users never seeing any posts they disagree with. This could increase the “echo chamber” effect that observers of social media have already cited as a problem.</td>
</tr>
<tr>
<td>Since this approach does not involve censoring content, it is clearly constitutional. Groups whose members have been banned from social media platforms feel that there is already too much interference with content. This approach would help protect users from unwanted speech without limiting freedom of expression for alternative viewpoints.</td>
<td>We can respect companies’ intellectual property while giving users more control by supporting education for social media users. There are many ways to work around the algorithms used by social media platforms. The government could fund programs to educate users about how algorithms work and how users can work around them. In addition, the materials could teach people how to evaluate online information and even to engage with people different from themselves.</td>
</tr>
<tr>
<td>Although this approach does not involve censorship, it would require government monitoring. Some observers have suggested that an oversight committee be set up. This could be a new branch of the FCC made up of coders. The group could monitor new media using algorithms of their own. With the right expertise, a group like this could make social media safer for users.</td>
<td>An alternative would be based on a more deliberative approach. Every system that uses algorithms should have a designated person or group to maintain and update the algorithm. In addition, that person or group should develop a structure for deliberating with the community of users to determine how to improve the algorithm and insure it is not harming people.</td>
</tr>
</tbody>
</table>

---

xvii Such biases can be especially harmful to teens. Making algorithms more transparent and offering options can address this problem.

xvi This approach assumes that users will opt for strong, positive content. But letting users control the content they see could result in users never seeing any posts they disagree with. This could increase the “echo chamber” effect that observers of social media have already cited as a problem.

xviii An alternative would be based on a more deliberative approach. Every system that uses algorithms should have a designated person or group to maintain and update the algorithm. In addition, that person or group should develop a structure for deliberating with the community of users to determine how to improve the algorithm and insure it is not harming people.
Below are several policy actions that would be consistent with this option and their drawbacks.

<table>
<thead>
<tr>
<th>Action</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Require social media companies to disclose their algorithms for displaying content. This would allow for open discussion and development.</td>
<td>Making the workings of these algorithms public will encourage people to modify their content to avoid detection and increase engagement.</td>
</tr>
<tr>
<td>Require that social media companies allow users to personalize how the site’s algorithms display content, including allowing options that remove the role of the algorithm altogether.</td>
<td>Most people are not aware of how these algorithms work (or even of their existence) and are unlikely to take advantage of this personalization even if it is available.</td>
</tr>
<tr>
<td>Create an oversight committee in the Federal Communications Commission (FCC) made up of coders who could independently monitor new media algorithms and determine when they are likely to promote harmful material.</td>
<td>While not technically censorship, having the government evaluate algorithms might allow them to have undue influence on how social media sites display content. A better approach might be to let the free market resolve the control of algorithms and to encourage individuals to seek out or encourage the development of social media sites who take a more ethical approach to algorithm design.</td>
</tr>
</tbody>
</table>
Image sources


Glossary

**Advisory:** an official announcement, typically a warning of some type.

**Algorithm:** a process or set of rules followed in solving a problem, especially when used by a computer.

**Anecdote:** a short story about a real life incident.

**Anti-Semitism:** hatred or prejudice against Jewish people.

**Arbitrary:** based on random choice rather than reason or logic.

**Artificial intelligence:** computer systems able to perform tasks that normally require human intelligence.

**Compelling:** powerful, convincing.

**Compliance:** meeting rules or standards.

**Content moderation:** monitoring of user-generated content based on a platform’s rules/guidelines.

**Democratic values:** beliefs and principles underlying our government and its founding documents (e.g., justice, liberty, equality).

**Double standard:** unfair, differing application of a rule to different groups.

**Echo chamber:** an environment in which a person only talks with people who agree with them.

**Harassment:** repeated, unwanted conduct intended to demean, humiliate, or threaten a person.

**Hate crime:** a crime, usually involving violence, motivated by bias.
Hate speech: abusive or threatening speech expressing prejudice against a group.

Imminent: happening soon.

Incitement: provoking or urging someone to commit a crime.

Infringing: limiting, especially someone’s rights.

Intellectual property: creations of the human mind that do not have a physical presence.

Legally liable: responsible according to the law.

Libel: a published false statement that damages a person’s reputation.

Normalized: made to seem ordinary, standard.

Obscenity: indecent or extremely offensive speech.

Partisan: unreasonably biased in favor of a cause, often the cause of a political party.

Property rights: the legal right of a person to hold ownership of belongings.

Regulate: control or supervise.

Scrutiny: critical examination.

Shabbat: the Jewish holy day or Sabbath.

Slander: making a false spoken statement that damages a person’s reputation.

Slippery slope: an action that will lead to a bad outcome.

Social media: the means by which people are able to create and share information and ideas virtually.

Social media platform: internet sites that host social media activities.

Standards: a norm or rule to be met.

Transparency: the condition of being transparent.

Transparent: open to public scrutiny.

---

iii (https://www.pewresearch.org/short-reads/2021/01/13/qa-what-weve-learned-about-online-harassment/#:~:text=Fully%2079%25%20of%20Americans%20think,17%20percentage%20points%20since%202017


v Campbell Robertson, Christopher Mele, and Sabrina Tavernise, “11 Killed in Synagogue Massacre; Suspect Charged with 25 Counts,” The New York Times (October 27, 2018),


vii Social media apps are 'deliberately' addictive to users - BBC News


xiv Social media apps are 'deliberately' addictive to users - BBC News


xix Ibid.