



## Alexandra Baddour

Alexandra Baddour is an Integrated Marketing Communication major at Slippery Rock University, with a minor in Graphic Design. She has taken an interest in how technology affects communication and how written communication has evolved since the introduction of the Internet.

# Hate Speech & Censorship

Politics in the United States are highly polarized, especially since the beginning of the presidential campaigns for the 2016 presidential election. People have taken to the internet with their opinions, starting and inflaming arguments on social media sites. With those arguments have come slurs and antagonizing comments. Add to that the rise of white supremacist groups and people on all sides 'doxxing' each other, the internet can seem like a toxic and dangerous place

A catch-all term for these slurs and antagonizing comments is 'hate speech.' Many people think speech like this should be restricted while others say restricting such speech would be a violation of the First Amendment. There is a lot to unpack in this debate. What does the First Amendment say and does it protect hate speech? What does 'freedom of speech' mean? What constitutes 'hate speech'? Can such speech be restricted? Should it be restricted?

The First Amendment, in its entirety, reads, "Congress shall make no law respecting an establishment of religion, or protecting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the government for a redress of grievances" ("The 1st Amendment"). That is a lot to squeeze into just one amendment. It guarantees several rights in very broad strokes. Concerning speech, it simply says that the United States government cannot impose on citizens' freedom of speech. An important detail to note here is that the First Amendment protects citizens from the government. It does not protect citizens from each other or from corporate entities. While this is certainly an important fact on its own, it becomes even more relevant when discussing if and how hate speech should be censored. After all, according to the broad wording of the First Amendment, there is no reason hate speech should not be protected.

Freedom of speech, or free speech, simply means that you are legally allowed to say things, whether through spoken word, physical text, or digital text. Freedom of speech is generally understood to be what is called a 'negative right.' This means that "it is exercisable only against the state and acts as a shield against interference" (Stone). Put more simply, the government cannot limit its citizens' speech. Smith points out a common misunderstanding, noting that "people often mistake freedom for license – for the prerogative to do as one pleases, subject to no boundaries whatsoever" (Smith). This is an incorrect understanding because there are many scenarios where speech is understood to be limited through rules or social norms.

'Hate speech' is considered free speech. Leets categorizes hate speech under the broader category of 'harmful speech,' which she references Leets and Giles defining as "utterances that are intended to cause damage, and/or irrespective of intent, that their receivers perceive to result in damage" (Leets). Leets then explains that, within the legal community, 'hate speech' is understood to mean "speech that denigrates persons on the basis of their race or ethnic origin, religion, gender, age, physical condition,

disability, sexual orientation, and so forth" (Leets). What "and so forth" means is not defined by Leets, but many people argue that rude comments made about a person's political leanings can be categorized as hate speech, and a few people consider anything said against themselves to be hate speech. For this paper, 'hate speech' will be used as it is defined in legal contexts as described above.

Leets also points out that people often do not understand the harm that can come from being a target of hate speech. "Most would agree that hate speech is ugly and regrettable, but not all understand how it is injurious, yet alone traumatic" (Leets). Leets says that being a target of hate speech "strips people of their dignity," "can elicit anxiety and distress," and that, while reactions are not always traumatic, the victims' response over time follows a pattern similar to trauma response (Leets). Words, both verbal and written, can have a notable negative impact on people's lives. The answer to whether or not hate speech, or any other speech for that matter, can be restricted is an easy "yes." There are legal ways to restrict speech. The key is in how speech is restricted - or rather, by whom. Facebook, Twitter, Instagram, and other social media sites can restrict speech. The government, however, cannot. For the most part, it is just that simple.

The reason that social media sites, and most other companies, are legally allowed to restrict speech is that when you sign up for an account or download an app, you are asked to read and agree to a contract, though most people do not really read it. That contract is usually called something along the lines of "Terms of Use," "Terms of Service," "Terms and Conditions," or "User Agreement." They are long legal documents about what you can and cannot do with the site or app, and what steps will be taken if you do something you are not supposed to, among other things. For example, Twitter's Terms of Service simply says that if you use Twitter, you should be prepared to see "offensive, harmful, inaccurate or otherwise inappropriate" content ("Twitter Terms of Service"). You are responsible for your own content and your own experience. In May 2018, Twitter made changes to its Terms of Service, but those changes mostly pertain to data collection.

On the other hand, Facebook's Terms of Service asks users to help keep Facebook safe by not posting anything intimidating, harassing, threatening, or anything that may be hate speech, though it does not define hate speech. It goes on to add that you cannot use Facebook for "anything unlawful, misleading, malicious, or discriminatory" ("Terms of Service"). Facebook also says, "We can remove any content or information you post on Facebook if we believe that it violates this Statement or our policies" ("Terms of Service"). Facebook updated their Terms of Service in May 2018 in response to the Cambridge Analytica issue and resulting pushback from users. In the update, Facebook takes a stand against "misuse" and "harmful conduct," saying it is using technology to look out for such content and vowing to take "appropriate action," whether that means blocking content, banning users, or even contacting law enforcement ("Terms of Service"). What counts as "harmful content" remains vague and undefined, allowing Facebook to arbitrarily decide when and if to enforce this.

Instagram falls in the middle, saying that you cannot “bully, abuse, harass, or threaten” others, and declaring that each user is responsible for the content they produce (“Instagram Help Center”). As for what might happen if you go against these guidelines, Instagram simply says that they might do something about it, but that they do not have to, saying “We may, but have no obligation to, remove, edit, block, and/or monitor Content or accounts containing Content that we determine in our sole discretion violates these Terms of Use” (“Instagram Help Center”). There are two big problems with these Terms of Use. The first is that what is considered “harmful” content is only vaguely defined. The legal definition of hate speech is a good place to start. Adding something about physical threats wouldn't be amiss either. Internet threats are worth mentioning, and are an important and related issue, but doing any justice on the topic would require a paper of its own. The second is that screening for such content effectively is difficult, though Artificial Intelligence is making it easier. It would be incredibly difficult for a team of people to scan all the content posted on a social media site. Artificial Intelligence, or AI, could scan through content much faster and flag any content that may go against the Terms of Use then send flagged content to be evaluated by people who make the final decision about what action to take, if any.

The issue then would be programming the AI. What should it flag? Giving it a list of slurs would be a good place to start, but what about threats? What about cyberbullying? What about when people put spaces or symbols in slurs to try to trick the sensors? What about discriminatory comments that get creative and avoid using slurs altogether? What about the way language changes and groups reclaim slurs? No AI is ever going to be perfect, but with constant upkeep and regular human supervision, it can be a big step in the right direction. That being said, it should absolutely be humans, not AI, who make the final call on what action needs to be taken. There are AI that are mastering natural language, such as IBM's Watson, but they take a long time to train and are not as well equipped as humans to handle language nuances and evolving internet grammar rules.

There has been talk of national laws restricting hate speech, too. People against such laws say they would infringe on the First Amendment right to free speech and worry that they might open the door for the kind of censorship laws that could mean the end of free speech. People supporting such laws point out that the First Amendment was written well over one hundred years ago – society has changed, and now the internet allows us to communicate in ways people could not have imagined when the Amendment was written. Many find themselves somewhere between these two extremes. Neither side is incorrect, and it is likely to be a long time before any laws restricting hate speech can be seriously considered.

Even before we get to that point though, it is important to start thinking about how such a law might work. A law restricting free speech would likely face problems similar to those faced by social media sites' attempts to screen content. That is, how will hate speech be defined and how can such a law be enforced? The legal definition near the start of this paper is still a good starting point.

It covers various groups that are discriminated against. It does not give a list of slurs to be restricted, allowing the law to remain flexible even as language evolves, which is important.

The question of enforcement is much more difficult. The spoken word is hard to censor. Unless it was recorded, it is difficult to prove anything was said. Censoring printed speech can run dangerously close to infringing on the freedom of the press as well as freedom of speech. Then there is the matter of who is responsible for the hate speech in printed forms – the writer, the editor, the publisher, or someone else entirely? Decisions will have to be made about whether hate speech is allowed to be used in educational texts or historical fiction.

Censoring the internet is going to be extremely difficult too. In most places in the United States, the internet is provided by a company, not the local government. Either those companies, or individual websites, will have to oversee censoring. But what can, or should, be done about hate speech hosted by servers in other countries? What about people in other countries that post hate speech on social media sites hosted in the United States? Lines will need to be drawn on how the law applies on the internet, and that will likely impact how long it takes for a law to be decided on. Alternately, the law could be made without much thought of how it will apply on the internet, and the United States court systems will have to figure it out as they go. If the United States government does pass a law allowing censorship of hate speech, there is the potential for the government "to overstep their powers, selectively enforce laws, or otherwise misuse power in a way that undermines or frustrates any legitimate role for hate speech laws" (Stone). The First Amendment specifically mentions freedom of speech and freedom of the press so that the government cannot overstep their power and censor citizens. The government who puts a law allowing censorship on the books may not overstep the boundaries of the law, but there is no guarantee that governments in the future will do the same. Is disapproval and condemnation enough? Or should steps be taken to restrict hate speech?

There is no easy solution to the debate between restricting hate speech and saving free speech. Restricting hate speech could save a lot of people from harassment and self-hatred, but there are too many issues that will need to be worked out before such a law is useful or even enforceable. It also needs to be considered that restricting a person's hateful speech is not going to change their hateful attitudes or behaviors. But does that mean people should have to deal with hate speech aimed at them? A thoughtful debate of the issue goes around in circles. Smith says, "we should defend the racist's freedom to speak, for instance, on my view, but we should also condemn vehemently and articulately the depravity of his message" (Smith). Smith echoes a sentiment most often attributed to Evelyn Beatrice Hall - "I disapprove of what you say, but I will defend to the death your right to say it" (Smith).

## Works Cited

- "Instagram Help Center." Privacy Policy | Instagram Help Center, Instagram, [help.instagram.com/478745558852511](https://help.instagram.com/478745558852511).
- Leets, Laura. "Experiencing Hate Speech: Perceptions and Responses to Anti-Semitism and Antigay Speech." *Journal of Social Issues*, vol. 58, no. 2, 2002, pp. 341–361., doi:10.1111/1540-4560.00264.
- Smith, Tara. "The Free Speech Vernacular: Conceptual Confusions in the Way We Speak About Speech." SSRN, 8 May 2018, [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3166234](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3166234).
- Stone, Adrienne. "VIEWPOINT DISCRIMINATION, HATE SPEECH LAWS, AND THE DOUBLE-SIDED NATURE OF FREEDOM OF SPEECH." The Free Library, 2017, [www.thefreelibrary.com/VIEWPOINT+DISCRIMINATION%2c+HATE+SPEECH+LAWS%2c+AND+THE+DOUBLE-SIDED...-a0514177529](http://www.thefreelibrary.com/VIEWPOINT+DISCRIMINATION%2c+HATE+SPEECH+LAWS%2c+AND+THE+DOUBLE-SIDED...-a0514177529).
- "Terms of Service." Facebook, Facebook, [www.facebook.com/terms.php](https://www.facebook.com/terms.php).
- "The 1st Amendment of the U.S. Constitution." National Constitution Center – The 1st Amendment of the U.S. Constitution, [constitutioncenter.org/interactive-constitution/amendments/amendment-i](https://constitutioncenter.org/interactive-constitution/amendments/amendment-i).
- "Twitter Terms of Service." Twitter, Twitter, [twitter.com/en/tos](https://twitter.com/en/tos)