Week 10: Regulating Online Speech
March 11, 2020
Is Censorship Ethical?
* Censorship and Internet (Warf 2010)

Censorship and Internet authorities that invoke diverse strategies of suppression of various groups and individuals for a broad array of reasons and motivations. Adding to this complexity is the rapidity with which the Internet has grown and changed technologically; often government censors have difficulty keeping up-to-date with changing technologies (e.g., text messaging) or slang terms used to communicate hidden meanings.

The degree and type of Internet censorship obviously varies widely and reflects how democratic and open to criticism different political systems are. Reporters Without Borders, an NGO headquartered in Paris and one of the world's preeminent judges of censorship, ranks governments across the planet in terms of the severity of their Internet censorship (Fig. 2; see also Quirk 2006). Their index of Internet censorship is generated from surveys of 50 questions sent to legal experts, reporters, and scholars in each country. Thus, countries in northern Europe, the US and Canada, Australia and New Zealand, and Japan exhibit minimal or no censorship (scores less than 10). Conversely, a rogue's list of the world's worst offenders, including China, Vietnam, Burma/Myanmar, Iran, and Turkmenistan, exhibit the planet's most severe and extensive restrictions (scores greater than 80). In North Korea, Internet access is illegal, although the government uses it to send messages to the outside world (Hachigian 2002). In between these extremes lies a vast array of states with modest to moderate forms of Internet censorship that reflect their diverse systems of governance, the presence or absence of civil liberties, and the ability of various groups to resist limitations on their ability or right to use the Internet in whatever manner they so prefer.

Using the categories of Fig. 2, Table 1 summarizes the distribution of the world's population and Internet users according to the level of severity of censorship. Thus, only 13% of the world's people, but a third of Internet users, live in countries with minimal censorship; conversely, roughly one-quarter of the world's people and Internet users live under governments that engage in very heavy censorship (the vast bulk of whom are located in China).

Internet penetration rates—the proportion of the population with regular access to cyberspace at home, school, or work—also shape the contours of censorship geography (Fig. 3). Rates vary from as low as 0.2% (Myanmar) to 100% (Falkland Islands).

Fig. 2 Reporters Without Borders Internet Censorship Ranking 2009. Source: data drawn from http://www.rsf.org/en-classement1003-2009.html
Freedom of Expression
First Amendment

- It constrains what the US government can do
  - Congress shall make no law respecting an establishment of religion, or prohibiting 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 primary purpose is political – allow an open discussion of public issues; allow minority perspectives to be represented
  - But extends to other domains

- Not an absolute right
  - The private right must be balance against the public good
First Amendment

• Private groups can censor all they want and in some cases may seem to violate the first amendment
  ▪ E.g., Facebook’s Terms of Service disallow the sharing of explicit or violent content
  ▪ In general written with "offensive"/controversial speech in mind

• Covers spoken & written words, pictures, art, and other forms of expression of ideas (like wearing an arm band)
First Amendment

• Existing restrictions on speech – balancing private right and public good
  ▪ Advocating illegal acts can be acceptable (e.g., civil disobedience)
  ▪ Libel (making false and damaging statements) is illegal
  ▪ Making specific threats is illegal
  ▪ Inciting violence can be illegal
  ▪ Perjury
  ▪ Obscenity
  ▪ False advertising
Discussion point 1: In the US, television commercials for cigarettes are banned. Should there be a ban on commercials for violent video games too? (Describe in the light of “private right versus public good”)
Platform enforced content regulations
Online Hate/Abusive Speech
Free Speech and Hate Speech

- Social media and the Internet have opened up for many new arenas for exchanging opinions.
- Freedom of speech is an absolute value in any democracy, both for the public and for the media.
- At the same time, opinions and debates challenge us as hate speech are spread widely and frequently on new platforms for publishing.
- Hate speech may cause fear and can be the reason why people withdraw from the public debate
  - Women and minority groups most affected
Reddit introduced a new anti-harassment policy

Reddit banned a handful of its worst communities, saying: "Our goal is to enable as many people as possible to have authentic conversations and share ideas and content on an open platform"

Prominent is the removal of r/fatpeoplehate, a forum devoted to posting (among other things) pictures of overweight people for ridicule.

They said “They are banning behaviors or ideas”
- Banning policy was based on specific instances of harassment rather than general offensiveness
Reddit Banned FatPeopleHate

- Deciding which subreddits engage in harassment is going to be a difficult problem, especially because it's hard to even agree on what constitutes harassment — and that assumes everyone involved is acting in good faith, which almost certainly won't be the case.
Reddit Limits Noxious Content by Giving Trolls Fewer Places to Gather

The Shift
By KEVIN ROBSE  SEPTE 25, 2017

Alexis Ohanian, a founder of the website, which banned several forums in 2015 as part of a broad crackdown on poisonous behavior. David Paul Morris/Bloomberg

There are — and always have been, and probably always will be — trolls, scoundrels and reprobates on the internet.

It is a problem that has vexed multibillion-dollar corporations and the smartest computer programmers in the world. Facebook, Twitter and YouTube have all declared war on abuse and harassment, spent years training sophisticated algorithms and hired vast armies of moderators to root out hateful content.

And yet, the trolls persist.

But what if a better way of combating online toxicity were right under our
Pro Self-Harm Content
Why Eating Disorders Are So Hard For Instagram And Tumblr To Combat

Over the last four years, the social media platforms have done a lot to curb content that promotes self-injury. But they'll never fully succeed. Is it worth trying?

Posted on April 14, 2016, at 2:01 p.m.

Stephanie M. Lee
BuzzFeed News Reporter

#anorexia

5,770,983 posts
Interventions

Everything okay?
If you or someone you know is struggling with thoughts of suicide, the Lifeline is here to help: call 1–800–273–8255

If you are experiencing any other type of crisis, consider chatting confidentially with a volunteer trained in crisis intervention at www.imalive.org, or anonymously with a trained active listener from 7 Cups of Tea.

And, if you could use some inspiration and comfort in your dashboard, you should consider following the Lifeline on Tumblr.

Go back

View search results
But deviant behavior subverts attempts to intervene

| anorexic, anorexie, anoressia, anorexi, anorexia, anorexique, anorexia, anorectic, anorexia, anoretic |
| eatingdisorders, eatingdisorder, eatingdisoder, eatingdis, eatingdisorter, eatingdisoreder, eatingdisorde, eatingdisorderrr, eatingdisordered, eating_disorder |
| thighgaps, thygap, thighgapp, thigh_gap, thightgap, thyghgap, thighgappp, thegap, thigap, thighgapss |
| thinspooooo, thynspo, thyapoo, thynsppo, thinspoo, thinspoo, thynsppo, thynsppoo, thinnsp, thinspoooooo |

Increasing and more complex lexical variations have emerged since Instagram enforced moderation of pro eating disorder content in 2012

Automatically detect lexically variant tags that emerged out of moderated tags using edit distance computation + regular expression matching + semantic annotations via crowdsourcing (Amazon’s mechanical turk)

“thigap” > “th*g*p*”
Variants/Tag Chain

#thighgap

Root

#thighgapp

#thighgappp

#thyghgapp

#thygap
**But** deviant behavior subverts attempts to intervene

Increasing and more complex lexical variations have emerged since Instagram enforced moderation of pro eating disorder content in 2012.
Moderation was followed by increased social engagement

(Chancellor, Pater, Clear, Gilbert, De Choudhury, CSCW 2016)
Reddit ran wild with Boston bombing conspiracy theories in 2013 and is now an epicenter for coronavirus misinformation. The site is doing almost nothing to change that.
Spam
• What is spam?

• With ease of internet access, businesses looked for ways to capitalize on market opportunities associated with Internet communications – easier/cheaper to send emails than physical mails
  ▪ How to find email addresses though?
  ▪ Crawling the web; scrape address books with viruses; listen to chatroom conversations; sneaky way to sign up; dictionary attacks on ISPs

• This entrepreneurial behavior has given rise to a new set of legal and ethical problems
Cyber Promotions versus American Online

The scenario:

- AOL blocks Cyber Promotions email
- Cyber promotions changes “from” address
- Cyber promotions sues AOL for violating first amendment rights
  - Loses
- AOL defends self by claiming violation of property rights; carrying spam is a cost to AOL
Discussion point 2: Why is “cold calling” considered to be an acceptable sales practice, but spamming isn’t?