

Digital Platforms in the Age of Content Moderation
Harvard Law School
Fall 2020

Prof. [Jonathan Zittrain](#)
TA: [Will Marks](#)

Tuesdays 8:00-10:00 PM EST
[Zoom link to be distributed]

COURSE OVERVIEW

The conventional wisdom around the responsibilities of platforms like Facebook, Google, and Amazon has evolved greatly. 1995-2010 saw the era of “rights,” in which the prevailing view was one of intermediaries’ abstention from shaping the flow of users’ contributions.

From then until now there’s been a new framework, “public health,” emphasizing the harms that can result from unchecked online activity, just as AI techniques promise new ways to scale platforms’ monitoring and intervention in the activities they facilitate.

What’s next? We’ll explore.

Please contact Will Marks <wmarks@law.harvard.edu> with any questions relating to course content or logistics.

COURSE MATERIALS

We have worked to ensure that the readings for this course are manageable. As discussion will be tightly focused around assigned readings, we require that students prepare carefully and completely for each session. Given our discussion-based seminar format, failure to complete the readings will be both obvious and an impediment to the flow of class.

Most course readings are available on this course’s [H2O playlist](#). Links to readings are also provided in the reading list that follows. Additional materials may be circulated via email in advance of class.

Day One (9/08): The Past, or The Era of Rights

The era of rights, from roughly 1995-2010, emphasized freedom of speech and platforms' abstention from moderating users' content. In this era John Perry Barlow articulated a vision of the internet free from government repression and wrote that "[w]e are creating a world where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity." Section 230, enacted in the same year, was similarly intended to encourage the unregulated development of free speech on the internet: by saying that online services weren't liable for third party content, it removed the incentive for service providers to over-regulate online speech. This environment allowed for the rise of platforms.

- ❖ ["Three Eras of Digital Governance"](#) by Jonathan Zittrain (2019).
- ❖ ["The Crypto Anarchist Manifesto"](#) by Timothy May (1988).
- ❖ ["A Declaration of the Independence of Cyberspace"](#) by John Perry Barlow (1996).
- ❖ ["Code V.2"](#) by Lawrence Lessig, Basic Books (2006).
 - Read Chapter 1: "Code is Law."
- ❖ ["Some Peer-to-Peer, Democratically, and Voluntarily Produced Thoughts"](#) by Ann Bartow, *University of New Hampshire Scholars' Repository* (2007).
- ❖ ["Facebook is Running Anti-Vax Ads, Despite Its Ban on Vaccine Misinformation"](#) by Caroline Haskins, *Buzzfeed News* (2020).
- ❖ *Algorithms of Oppression* by Safiya Noble, New York University Press (2018).
 - Read Chapter 1: "A Society, Searching."

Day Two (9/22): The Present, or the Era of Public Health

In the era of public health, we have become cognizant of the tangible effects of platforms' policy decisions. By the 2010s, users' expectations had changed. In the wake of growing awareness of online harassment and hate speech, sites like Twitter with blanket refusals to police users' content went from internet heroes to internet villains. Platforms responded by increasing their moderation out of a fear of losing users, and states have stepped in.

- ❖ [47. U.S. Code § 230](#) (1996).
- ❖ <https://cda.wtf/> (2020).
- ❖ "The Offensive Internet: Speech, Privacy and Reputation" by Martha Nussbaum and Saul Levmore, Harvard University Press (2010).
 - Read Introduction.
- ❖ ["The New Governors: The People, Rules, and Processes Governing Online Speech"](#) by Kate Klonick, *Harvard Law Review* (2018).
 - Read Section II.
- ❖ ["Herrick v. Grindr: Why Section 230 of the Communication Decency Act Must be Fixed"](#) by Carrie Goldberg, *Lawfare* (2019).
- ❖ ["Towards a Clearer Conversation about Platform Liability"](#) by Daphne Keller, Knight First Amendment Institute (2018).
- ❖ ["Online Networks of Eating-Disorder Websites"](#) by Antonio Casilli et al., *Perspectives in Public Health* (2013).

Day Three (10/06): The Future, or (maybe) the Era of Legitimacy

In this coming era of legitimacy we might focus less on theories of rights or studies of concrete harms, and instead emphasize *process* – the creation of frameworks for adjudicating disputes which can achieve buy-in from all relevant participants. Facebook has already taken a step in this direction with the external review board, a sort of “Supreme Court” which might externalize some of the most important content policy decisions that the platform makes. According to Kadri and Klonick, Facebook already relies on many of the tools used by courts to resolve tensions between regulating harmful speech and preserving free expression. What can we learn from Facebook’s earlier efforts to create a governance system based around user votes? Do these developments bring us closer to early internet proponents’ vision of cyberspace institutions that run by the rule of the people? And what are the promises and pitfalls of Facebook’s other proposed solution to its moderation woes: using algorithms to clean up content?

- ❖ [“Mark Zuckerberg Wants to Democratize Facebook – Here’s What Happened When He Tried”](#) by Adi Robertson, *The Verge* (2018).
- ❖ [“Google’s Comment-Ranking System Will Be a Hit with the Alt-Right”](#) by Violet Blue, *Engadget* (2017).
- ❖ [“Just in CASE: Legislation Proposing Small-Claims Court for Copyright Disputes Headed to Congress”](#) by Lauren Heperi, *BYU Copyright* (2019).
- ❖ [“A Jury of Random People Can Do Wonders for Facebook”](#) by Jonathan Zittrain, *The Atlantic* (2019).
- ❖ [“Twitter’s Least-Bad Option for Dealing with Donald Trump”](#) by Jonathan Zittrain, *The Atlantic* (2020).
- ❖ [“The Technology 202: Lawmakers Unveil Bipartisan Proposal to Make Social Media Moderation More Transparent”](#) by Cat Zakrzewski, *The Washington Post* (2020).
- ❖ “The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression” by Kate Klonick, *Yale Law Journal* (2020).
 - Read Abstract, Introduction, Sections I & III, Conclusion.

Day Four (10/20): Global Platforms, Local Law

Thus far we have considered the ways that America has approached the regulation of digital platforms. Cyberspace was supposed to be independent and international. We now know it's not. Is it helpful to ask how many internets there are, and how many there should be?

- ❖ "Who Controls the Internet: Illusions of a Borderless World" by Jack Goldsmith and Tim Wu (2006).
 - Read Pages 1-22.
- ❖ "[There May Soon Be Three Internets. America's Won't Necessarily Be the Best.](#)" by the Editorial Board, *The New York Times* (2018).
- ❖ "[Google's Gatekeepers](#)" Jeffrey Rosen, *The New York Times Magazine* (2008).
- ❖ "[Facebook Can Be Forced to Delete Content Worldwide, E.U.'s Top Court Rules](#)" by Adam Satariano, *The New York Times* (2019).
- ❖ "[Glawischign-Piesczek vs. Facebook Ireland Limited](#)" (2019).
 - Paragraphs 76-103; 109
- ❖ "[Canada's Supreme Court Rules Google Must Block Certain Search Results Worldwide](#)" Dani Deahl, *The Verge* (2017).
- ❖ "[US v Equustek](#)" (2017).
- ❖ "[Trump Targets WeChat and TikTok, in Sharp Escalation with China](#)" by Ana Swanson, Mike Isaac and Paul Mozur, *The New York Times* (2020).

Day Five (11/10): Deplatforming

What are the limits to private deplatforming? Would deplatforming become more acceptable if we could offer due process for users who risk being deplatformed, perhaps along the lines of Facebook's external review board or the EFF's [Manila Principles](#)? When should we see the rights of users as substantive, and when procedural?

- ❖ "[Terminating Service for 8Chan](#)" by Matthew Prince, *The Cloudflare Blog* (2019).
- ❖ "[The Lawless Way to Disable 8Chan](#)" by Evelyn Douek, *The Atlantic* (2019).
- ❖ "[AirBnB Cancels Accounts Linked to White Nationalist Rally in Charlottesville](#)" by Jonah Bromwich, *The New York Times* (2017).
- ❖ "[Uber, Lyft ban right-wing activist after anti-Islamic tweets](#)" by Zac Estrada, *The Verge* (2017).
- ❖ "[Facebook Bars Alex Jones, Louis Farrakhan and Others From Its Services](#)" by Mike Isaac and Kevin Roose, *The New York Times* (2019).
- ❖ "[What is the Streisand Effect?](#)" by T.C., *The Economist* (2013).
- ❖ "[William F. Buckley Program at Yale Hosts its Second Annual 'Disinvitation Dinner'](#)" by Nicholas Quinn Rosenkranz, *The Washington Post* (2016).

Day Six (11/17): Federation vs. Top-down Governance

Different platforms moderate through different mechanisms. Platforms like Facebook, Youtube, and Twitter moderate content internally, at least for their feeds. Others, like Reddit and Nextdoor, allow for varying levels of “community” self-governance. What are the limitations to top-down governance? When should platforms limit the rights of groups to self govern? By what criteria should companies monitor their feeds, groups, and messages?

- ❖ [Community Standards](#), Facebook (2020).
- ❖ [The Twitter Rules](#), Twitter (2020).
- ❖ [Youtube's Community Guidelines](#), Google (2020).
- ❖ [Reddit Content Policy](#), Reddit (2020).
- ❖ [Nextdoor Community Guidelines](#), Nextdoor (2020).
- ❖ [“Facebook Groups are Falling Apart over Black Lives Matter Posts”](#) by Nick Statt, *The Verge* (2017).
- ❖ [“WhatsApp Says its Forwarding Limits Have Cut the Spread of Viral Messages by 70 Percent”](#) by Manish Singh, *TechCrunch* (2020).
- ❖ [“Reddit Bans r/The_Donald and r/ChapoTrapHouse as Part of a Major Expansion of its Rules”](#) by Casey Newton, *The Verge* (2020).
- ❖ [“It's 'Our Fault': Nextdoor CEO Takes Blame For Deleting Of Black Lives Matter Posts”](#) by Bobby Allyn, *NPR* (2020).
- ❖ [“Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance”](#) by Robert Gorwa, Reuben Binns, and Christian Katzenbach, *Big Data and Society* (2020).
- ❖ [“The Trauma Floor”](#) by Casey Newton, *The Verge* (2019).