

# First Amendment Bootcamp: Content Moderation

*James Grimmelman*

February 13, 2021

# Platforms!

- Social media: Facebook, Twitter, YouTube, TikTok, Reddit, Parler, Discord
- Publishing: Medium, Spotify, Substack, itch.io
- Infrastructure: Cloudflare, AWS, Verizon, GoDaddy, Verisign
- Payments: PayPal, VISA, Venmo, Bitcoin, Coinbase, Patreon, GoFundMe
- App Stores: Apple, Android, PlayStation, Epic
- Communications: Gmail, Cornell, Slack, Signal
- Marketplaces: Redbubble, Amazon, eBay, Poshmark
- Gig economy: Lyft, Grubhub, Doordash
- Search: Google, Bing, DuckDuckGo, Yelp, Redfin, Kayak
- Aggregators: Netflix, Disney+, Google Play, Spotify
- One size does not fit all

# 1A media analogies

- Broadcast: formerly subject to the fairness doctrine, equal time, etc., but now increasingly unregulated (and increasingly irrelevant)
- Newspapers: basically unregulable
- Cable: lightly regulable (e.g. public access)
- Telephone, mail: common carriage
- Broadband Internet: network neutrality???
- Search engines: editor, conduit, advisor?

# Harmful speech challenges online

- Virality
- Anonymity and impunity
- Context collapse
- Spam spam spam spam spam spam spam spam spam spam spam
- New (?) kinds of harms, e.g. nonconsensual pornography

# Section 230

# 47 U.S.C. § 230 (“Section 230”)

(c)(1) No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.

(c)(2) No provider or user of an interactive computer service shall be held liable on account of (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected ...

# Leaving up and taking down

- (c)(1) provides absolute immunity for *leaving up* UGC
- (c)(2)(A) provides qualified immunity for *taking down* UGC
  - Courts have read (c)(2)(A)'s “otherwise objectionable” and “good faith” requirements relatively narrowly, at least at the MTD stage
  - But many of these suits fail due to TOS or no cause of action
  - And courts have read (c)(1) to protect takedown decisions, too (!)

# § 230 exceptions and end-runs

- Online marketplaces, e.g. Airbnb
- Contracts?
- Federal criminal law
- IP: copyright, trademark, right of publicity?
- Communications privacy law
- FOSTA/SESTA



# What is the relationship between 1A and § 230?

- Maybe the 1A leaves an open space that Congress filled with § 230
- Maybe § 230 will have to deal with every issue the 1A raises?
- Maybe § 230 *violates* the 1A? Maybe the 1A *requires* § 230?

# § 230 reform proposals

- Specific carve-outs: FOSTA/SESTA
- Negligence standard
- Require some level of overall effort to address harmful content
- Condition immunity on viewpoint-neutrality

# Content Moderation

# Layers of rules

- TOS: very broad, legally binding
- Community standards: moderately broad, binding only on users
- Published guidance: application of standards to particular cases
- Internal moderator handbooks: detailed guidance on specific cases

# Why is this stuff so hard?

- Context
- Parody
- Criticism
- Users determined to find and test the lines

# Moderation procedure

- Algorithmic flagging + user reporting
- Low-paid contractors in call centers do high-volume review
  - Stressful work dealing with psychically scarring content
- Internal escalation to supervisors, company, Trust & Safety team
- Limited ability to file request for review

# Facebook's "Supreme Court"

- Oversight Board can hear appeals from users whose content was taken down or whose accounts were banned
- Results "binding" on Facebook in specific case
- Trump!
- Jurisprudential issues: judicial appointment, judicial ethics, jurisdiction, mootness, scope of precedent, advisory opinions, ...

Deplatforming



# Deplatforming?

- Cloudflare bans the Daily Stormer
- Twitter bans Trump
- YouTube demonetizes PragerU
- Instagram (mostly) bans female nipples
- Ravelry bans pro-Trump posts
- Parler bans defamation
- Facebook bans spammers

# Some questions to ask

- How far up or down the stack is the platform?
- How good are the alternatives?
- Is money changing hands?
- How far beyond the pale is the speech?
- Who has to see the speech?
- Is there an affected community? Is it on or off the platform?

*The End*