Social Media Moderation

IAB Virtual Retreat 2021

Goals of Discussion

- Is there one or more IAB discussions worth having?
  - If so, where and how do we have them?
- Scope control
  - Problem space is wide ranging and highly controversial
  - Many aspects outside of IAB purview
  - Tread carefully -- This subject has a high chance of going sideways.
- Think in terms of practical effects
- Non-goals
  - Solving all the problems today
  - Moral judgements (Look at practical effects)
  - Diving down rat holes (or rabbit holes)

What is Moderation

- Multiple forms of moderation
  - Content Blocking
  - Reducing audience
  - Demonetizing
  - Suspensions and bans
  - Labeling/Fact Checking
- Things that get moderated
  - Illegal content
  - Offensive/Objectionable content
  - Misinformation
  - Unwanted Content

Competing Principles

- Property rights of private parties
  - May establish their own community norms
- Freedom of speech when the forum is controlled by private parties
  - Consolidation issue?
- Internet amplification of harmful content
  - With a fuzzy definition of “harmful”
- Polarization of public discourse
  - Ideology bubbles with lowered tolerance for dissent
- Government policy
  - Across multiple jurisdictions with different local norms

Who are we talking about?

- Mainly the actions of private organizations
  - … but sometimes ordered or ”encouraged” by governments
- Social Media platforms are generally allowed to establish community norms
  - … and will tend to optimize for the most eyeballs

US Legal Landscape – Section 230

- Section 230 of the Communications Decency Act (DCA)
  - “26 Words that Created the Internet” – Jeff Kosseff
  - Intermediary Liability Protections – “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.”
  - “Good Samaritan” provision – Civil Liability protection for actions to restrict availability of objectionable material.
    - Not limited to restriction of illegal or constitutionally unprotected material
- Different jurisdictions are, well, different.

Concerns

- Transparency
  - Users complain that they are not told why and lack paths for appeal
  - Even platform owners don't always know why content is restricted
    - Opaque algorithms/machine learning
- Bias
  - Concerns that platforms are biased against certain opinions or political opinions
- Over-moderation
  - Restriction of more content than intended
    - Anti-nudity rules may impact medical discussions, art
    - Anti-violence rules may impact discussions on reducing violence
    - Facebook – Wala Wala onions are "overtly sexual"
    - Twitter – Whatever you do, don't talk about "Memphis"

More Concerns

- Concentration of power –
  - Power over speech concentrated in a few companies
  - Intersects with concentration/consolidation issues
- Echo chambers
  - Are people protected from diverse ideas?
  - Polarization of opinions
  - May be an issue with the very idea social networks (e.g. subscribing to affinity groups or following specific people)
- Blocking at lower layers
  - Web hosting service blocking entire sites
  - Blocking domain names or IP address ranges
  - (See over-moderation)

Yet More Concerns

- Privacy Impacts
  - Content filtering may discourage e2e encryption
  - If you can filter, you can monitor
  - Endpoint-based filtering not that great either?
- Calls for more regulation
  - Lots of proposals to reform section 230 (from both parties)
    - Make it harder to restrict legal or constitutionally protected material
    - Make it easier to force companies to restrict certain content
  - Even some platforms claim to want regulation
  - High chance of governments screwing this up
    - Lack of tech understanding
    - Inconsistent requirements across jurisdictions
- Slippery slope arguments
  - Do technical filtering capabilities make government censorship and surveillance easier?

Where do we go from here?

- Already a lot of academic, legal, and advocacy literature in this space.
- Can the IAB help on any part of this?
  - Not clear if there is IETF work to do right now. IRTF?
- Do we have any consensus positions?
- Can we enable conversations?

Next Steps?

- What comes next?
  - Workshop on moderation algorithms, techniques, transparency and/or privacy considerations?
  - IAB document on harms of over moderation or blocking at lower layers?
  - Should we opinions on Section 230 reform? (Maybe one for ISOC)
  - Educate the people working on government policy

- (Just making stuff up here…)