Stanford Law School

Intermediary Liability

Daphne Keller

June 2018

Intermediary Liability

The EU's horizontal regulatory framework for illegal content removal in the digital single market - towards a balanced and predictable overall liability regime for online platforms.

Intermediary Liability Policy Goals

- Reduce *unlawful* online content and activity
- Protect *lawful* online content and activity
- Promote innovation and economic growth

Reducing *Unlawful* Online Content and Activity

- Address harms ranging from movie piracy to child pornography.
- "Gatekeeper" role can make platforms powerful enforcers.
- Platforms benefit from online content, so asked to bear some costs of negative externalities.
- Platforms may have superior technical tools for identifying suspicious content or activity (but inferior tools for legal assessment)

Protecting *Lawful* Online Content and Activity – the Over-Removal Problem

- Human error (Urban et al 2016, my CIS blog post listing other studies)
- Filtering error (CDT, Feamster & Engstrom)
- False accusations are very common.
  - Censorship goals: Ecuador, Retraction Watch
  - Commercial goals: in one study, 55% of takedowns targeted commercial competitors (Urban 2006).
- Platforms are motivated to err on side of removal.
  - Economic risk: liability risk, cost of vetting process
  - Reputational risk: media and political costs

Protecting Lawful Online Content and Activity – Consequences of Over-Removal

Over-removal harms much more than expression. It affects:

- Business and commercial activity
- Privacy, data protection, dragnet surveillance (pervasive private monitoring + police reporting)
- Social, religious, and political participation and assembly (loss of tools like Google Docs, WhatsApp, etc.)
- Discriminatory impact on minority groups (particularly in errors re "terrorist" content)

Protecting Lawful Online Content and Activity – "Terrorist" Content Example

Errors silence specific groups based on language, race, religion, etc.:

- *Justice concerns:* erasing prosecution material uploaded by witnesses and human rights organizations
- *Expression concerns:* Curtails both public political participation and innocent ordinary posts
- *Equality concerns:* disparate impact on racial, religious, and language minorities
- *Security and public order concerns:* Exacerbating social isolation, undermining counter-radicalization efforts

Removal Tools & Commission Proposals

Filters

- Powerful tools, but introduce major new sources of error (context failures) and amplify human error.
- Civil society (CDT), computer scientists (Feamster & Engstrom, Farid) and public examples (Syrian Archive) suggest serious limitations

Human review of filter results

- Documented high rate of over-removal in existing human systems
- Growing evidence of implicit or explicit bias

Tools for Correcting Removal Errors

Counter-notice from affected individual

- Data suggests little use (under 1%)
- Not effective for key categories of public interest material, such as videos from witnesses to human rights abuses

Transparency to broader public

- Transparency about specific content removed (not aggregate data) can crowd-source error correction
- For particularly sensitive or dangerous content, could substitute limited expert review

Can We Remove Bad Content But Not Good?

Optimist answer:

- Human review and counter-notice will meaningfully correct for over-removal by filters
- The platforms will figure it out

Realist answer:

- Every known version of privatized enforcement has highly foreseeable errors of both over- and under-removal, filters will add new over-removal errors
- Lawmakers' choices will determine real outcomes and drive platform behavior

Thank You

https://cyberlaw.stanford.edu/about/people/daphne-keller
@daphnehk