The Primary Question of Policy and Research: Why Do We care?

No matter where you go, there you are. But how do you know where that is? And how do I get where I want to go rather than where I am?

"Bad Content" v. "Good Content"

Bad Content: Harassment, threats, hate speech, fraud, disinformation treated as news, terrorism/radicalization/recruitment, specific content based on geographic region (e.g., "blasphemy," "undermining confidence in the government")

Good content: News – including raw data (Ferguson), citizen journalism, representation for marginalized communities, education, positive networking, social organizing, Red Lion/diversity of views and perspectives, social chatter.

What Is NetzDG?

- German law that went into effect January 1, 2018.

- Requires social media platform with 2 million or more users to take down "clearly illegal" within 24 hours of notification, 7 days if more complicated. Complaints can come from "individuals" or "reporting agency." Platform can refer to outside counsel for determination.

- Must publish report twice a year containing: the number of the received complaints, the number and qualifications of employees who are handling the complaints, the network's association memberships, the number of times an external party has been used to decide the illegality of the content, the number of complaints that led to the content being deleted, the time it took to delete the content, and measures that were taken to inform the complainant and the member who posted the deleted content.

- Potential fine of €50 million.

What Do The Transparency Reports Tell Us?

- "We know they are taking it seriously." --Reuben Hoffman, German Lawyer

- "Apparently, this pressure has led these companies to delete large amounts of content that was in fact legal in an effort to ensure that they will not be punished under the Network Enforcement Act. When deleting the content, Facebook and Google cite their community standards. In these standards they stipulate what kind of content users may share on their platforms and reserve the right to also remove content that is protected by communicative freedoms." – Reporters Without Borders

What Were Our Goals Again?

- What were the problems we were actually trying to solve? People running off and joining terrorist groups? People feeling unsafe or assaulted? Reduction in crime? Make the platform safe for ads?

- If policy is trade offs between benefits and costs, need to identify what constitutes a benefit and what constitutes a cost.

- We need to recognize this is extremely complicated, made worse by the frequent failure to articulate "obvious" goals and concerns. Articulation of success by legislators usually idealized/unrealistic.

- Not helped by longstanding total resistance to *any* content moderation policies or *any* regulation by many stakeholders.

My Very Preliminary Plea For Help

- We need to find ways to encourage or require platforms to give researchers access to raw data on how they do content moderation – including the platform's own content moderation through self-detection.

- Researchers need to socialize the complexity of the problem, while not suggesting that complexity should mean paralysis. Knowledge need not be perfect. Problems may not become clear until after policies are implemented and scaled.

- If reporting obligations are going to be required by law, need to think about what we want to track long term.

- We need qualitiative metrics, not just quantitative metrics. And we need to break down problem into specific sub-problems (e.g., impact on time-sensitive speech)