



Selected Papers of Internet Research 16:
The 16th Annual Meeting of the
Association of Internet Researchers
Phoenix, AZ, USA / 21-24 October 2015

CYBERBULLYING POLICIES OF SOCIAL MEDIA COMPANIES

Tijana Milosevic
University of Oslo

Scope, Theoretical Framework and Significance of the Research Topic

“Why don’t you kill yourself?” was the question 14-year-old Hannah Smith from the United Kingdom received on social media website Ask.fm, before she hanged herself in her bedroom (Smith-Spark, 2013). The teen had allegedly been exposed to a series of online taunts prior to leaping to her death. Ask.fm allows users to pose questions to other users anonymously. It boasts more than a hundred million users from more than 150 countries (Ask.fm, 2015), around half of which are under 18. In response to Hannah Smith’s suicide, some 15,000 people have signed an online petition requesting the UK government to act (HM Government, 2013). UK Prime Minister called the website “vile,” and asked advertisers to boycott it. The social media website issued a statement of condolences and promised to improve its safety measures and cyberbullying policies. Yet, a year after her suicide, a coroner’s report concluded that the girl had been sending vile messages to herself on Ask.fm and a police investigation could not find any evidence of cyberbullying (Davies, 2014).

In the United States, state laws and proposed federal laws contain anti-cyberbullying provisions which stipulate the role of schools in working with parents and sometimes law enforcement to address cyberbullying; the situation is similar in the European Union. However, these laws do not contain provisions regarding social media companies. When an incident with severe consequences unfolds, the companies become embroiled in controversy, which increases incentives for them to regulate cyberbullying behavior on their platforms. The companies require that users agree to Terms of Service (TOS), contracts which typically contain anti-cyberbullying provisions. They have a number of other corporate documents such as “Statements of Rights and Responsibilities,” “Principles,” or “Community Guidelines,” which stipulate the behavior that is allowed on the platform.

This study addresses a gap in academic research about the role of social media companies in addressing cyberbullying –as one aspect of digital imaginary. It draws on two theoretical frameworks: 1. Privatization of the digital public sphere, a body of works that analyze the increasing role of private companies in regulation of civil liberties (Balkin, 2008; DeNardis, 2014; DeNardis, 2012; Hestres, 2013; Vaidhyanathan, 2011; Zuckerman, 2010); 2. Youth risks and opportunities online, a body of research led by

Suggested Citation (APA): Milosevic, T. (2015, October 21-24). *Cyberbullying Policies Of Social Media Companies*. Paper presented at Internet Research 16: The 16th Annual Meeting of the Association of Internet Researchers. Phoenix, AZ, USA: AoIR. Retrieved from <http://spir.aoir.org>.

Dr. Sonia Livingstone and EU Kids Online Project, which seeks to attenuate moral panics that can emerge around youth's use of technology, and provide evidence-based understanding of children's digital experience (Livingstone, Haddon & Gorzig, 2012).

Cyberbullying or online bullying, typically defined as "repeated harm inflicted," can appear in the form of verbal abuse, harassment, rumors or social exclusion and can result in anxiety, low self-esteem, depression and maladaptive behavior (Hinduja & Patchin, 2009; Subrahmanyam & Smahel, 2011). Cyberbullying is a contested term that is difficult to define and that youth tend not to identify with (boyd & Marwick, 2011; Marwick & boyd, 2014), making it increasingly difficult for companies to regulate incidents on their platforms in ways that users would find helpful. Although there is no single agreed-upon definition of cyberbullying (Mishna, 2012; Vandenbosch & Van Cleemput, 2010), similar to offline bullying, it is defined as "willful and repeated harm inflicted" towards another person which typically involves power imbalance between two people (Hinduja & Patchin, 2009: 9). There is no academic consensus on how many times such harm needs to happen in order for a case to classify as "repeated." Once a case occurs it may be difficult to classify it and determine if it constitutes cyberbullying (Nocentini et al., 2010).

By "cyberbullying policies" I refer to self-regulation mechanisms that social media companies have in place to intervene in existing and prevent future cyberbullying incidents on their platforms. These mechanisms include, but are not limited to, reporting and social reporting tools, blocking, filtering, geofencing as well as any forms of human or automated moderation systems such as supervised machine learning; and social-emotional learning-based educational materials.

Key research questions include:

- What is the role of technological platforms in intervening with and preventing cyberbullying cases?
- What cyberbullying-related provisions are stipulated in companies' corporate documents and how do they differ among the companies? What are the implications of such phrasing?
- What are the tools of enforcing compliance of cyberbullying-related provisions?
- What issues might these companies be encountering and what can be known about effectiveness of these policies?
- How is "effectiveness of cyberbullying policies" constructed in the companies' discourse?
- How do the companies react to cases of perceived ineffectiveness of their tools of enforcement?

This kind of systematic analysis of cyberbullying policies of social media companies has not been conducted thus far and through interviews with key companies, which are difficult to obtain, this study seeks to provide pioneering steps towards a framework for analysis of effectiveness of industry self-regulation.

Brief Overview of Method

The companies included in this sample have been selected on two criteria: the large number of users they had at the time when this study was conducted; and appearance in media reports in relationship to online bullying, especially in reference to self-harm or suicide. Social media companies included in this sample are: Facebook, Instagram (Facebook-owned), Twitter, Ask.fm, YouTube (Google-owned), Yik Yak, Secret app, Google+, Tumblr (Yahoo! owned), Snapchat and Whisper and digital messengers Voxer, WhatsApp (Facebook-owned), and KIK.

Some of these platforms gain and lose user base as well as media attention in the course of a couple of months. As an illustration of such state of affairs: in March 2014, there were hardly any media reports about Secret app, which was only being launched at the time; in August 2014, a Brazilian judge had ordered Google Play and Apple App Store to remove the app due to cyberbullying concerns in media reports (Beasley, 2014). This is why the overview presented here does not purport to be comprehensive, since the landscape changes relatively quickly; rather, it hopes to provide a systematic analysis of relevant social media's cyberbullying policies, which form the basis of self-regulatory efforts in the industry; and which should inform the policies of new apps and companies that are yet to come.

Based on an interpretivist approach to qualitative textual analysis (Daly, 2007) I seek to understand both the policies and mechanisms that exist and the various meanings that they have for participants in this study. To that end, I have conducted: 1. A qualitative textual analysis of these companies' TOS, all corporate documents that are related to online bullying, harassment and abuse; as well as corporate blogs. 2. Twenty-seven in-depth interviews with the companies' representatives; representatives from e-safety NGOs from the United States and the European Union that work with social media companies on their cyberbullying policies; as well as with non-affiliated e-safety consultants.

Brief Overview of Results

The results provide a systematic analysis of tools that social media companies in the sample employ to address cyberbullying cases on their platforms; an analysis of how the companies interpret the relative effectiveness of their tools; as well as an analysis of the implications behind unavailability of some data that would testify to tools' effectiveness. For instance, most companies do not reveal the guidelines that their moderators use to determine if a case constitutes bullying; the statistics on how many bullying reports they receive and the number of moderators they employ, as well as other specifics of their moderation systems. The results also detail the differences in sophistication of tools between different types of companies; use of supervised machine learning, blocking and filtering; a comprehensive overview of social reporting¹ and use of emotional intelligence research in guiding the policy; as well as the logic that guides

¹ Social reporting refers to the reporting system that employs emotional intelligence research to enable conflict resolution between two or more users without reporting a case to the company.

employment of these tools. The discussion focuses on the tension between ensuring child safety and users' civil liberties, primarily freedom of speech and privacy.

References

- Balkin, J. (2008) "The Future of Free Expression in a Digital Age," *Pepperdine Law Review*, Vol. 36, 2008.
- Beasley, M. (2014). Brazilian Judge Orders Apple, Google to Remove Secret from the App Stores, Remotely Delete from Users' Phones. *9to5Mac*. Retrieved from: <http://9to5mac.com/2014/08/19/brazilian-judge-orders-apple-to-remove-secret-from-the-app-store-remotely-delete-from-users-phones/>. Accessed on: February 22, 2015.
- boyd, d. & Marwick, A. (2011). Bullying as True Drama. *The New York Times*. Retrieved from: http://www.nytimes.com/2011/09/23/opinion/why-cyberbullying-rhetoric-misses-the-mark.html?_r=0. Accessed on February 22, 2015.
- Daly, K. J. (2007). *Qualitative methods for family studies & human development*. Los Angeles: Sage Publications
- DeNardis, L.E. (2012). Hidden Levers of Internet Control: An Infrastructure-Based Theory of Internet Governance. *Journal of Information, Communication and Society*. 15 (3).
- DeNardis, L.E. (2014). *The Global War for Internet Governance*. Yale University Press: New Haven, Connecticut.
- Davies, C. (2014). Hannah Smith Wrote 'Vile' Posts to Herself before Suicide, Police Say. *The Guardian*. Retrieved from: <http://www.theguardian.com/uk-news/2014/may/06/hannah-smith-suicide-teenager-cyber-bullying-inquests>. Accessed on February 22, 2015.
- Hestres, L.E. (2013). App Neutrality: Apple's App Store and Freedom of Expression Online. *International Journal of Communication*. (7): 1265-1280.
- Hinduja, S., & Patchin, J. W. (2009). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Thousand Oaks, California: Sage.
- HM Government (2013). Government to Take Safeguarding Position Against Sites like Ask.fm (2013). Retrieved from: <http://epetitions.direct.gov.uk/petitions/48886>. Accessed on February 22, 2015.
- Livingstone, S., Haddon, L & Gorzig, A. (Eds). 2012. *Children, Risk and Safety on the Internet: Research and Policy Challenges in Comparative Perspective*. The Policy Press: Bristol and Chicago.
- Marwick, A. & boyd, d. (2014). "It's just drama": Teen perspectives on conflict in a networked era." *Journal of Youth Studies* 17 (9): 1187-1204 (p. 1188).

- Nocentini, A., Calmaestra, J., Shultze-Krumbholz, Scheithauer, H., Ortega, R. & Menesini, E. (2010). "Cyberbullying: Labels, Behaviors and Definition in Three European Countries." *Australian Journal of Guidance and Counseling* 20 (2): 129-142.
- Smith-Park, L. (2013). Hannah Smith Suicide Fuels Calls for Action on Cyberbullying. *CNN.com*. Retrieved from: <http://www.cnn.com/2013/08/07/world/europe/uk-social-media-bullying/>. Accessed on February 22, 2015
- Subrahmanyam, K., & Smahel, D . (2011). Chapter 10: The darker side of the internet: Violence, cyber-bullying and victimization. In *Digital Youth: Advancing responsible adolescent development* (pp. 179-199). New York: Springer
- Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior* 26: 277-287
- Vaidhyanathan, S. (2011) *The Googlization of Everything (and Why We Should Worry)*. University of California Press 2011.
- Zuckerman, E. (2010) "Intermediary Censorship," in *Access Controlled: The Shaping of Power, Rights and Rule in Cyberspace*. MIT Press: Cambridge.