Technological University Dublin

ARROW@TU Dublin

2019

Narratives of Industry Responses to Cyberbullying: Perspectives on Self-regulation From and About the Industry

Tijana Milosevic
University of Oslo

Brian O'Neill
Technological University Dublin, brian.oneill@tudublin.ie

Elisabeth Staksrud
University of Oslo

Follow this and additional works at: https://arrow.tudublin.ie/cserbk

Part of the Communication Technology and New Media Commons

2019

Narratives of Industry Responses to Cyberbullying: Perspectives on Self-regulation from and About the Industry

Tijana Milosevic

Brian O'Neill

Elisabeth Staksrud

Narratives of/in industry responses to cyberbullying

Tijana Milosevic

University of Oslo, Department of Media and Communication

Brian O'Neill

Dublin Institute of Technology

Elisabeth Staksrud

University of Oslo, Department of Media and Communication

Abstract

In this chapter, we provide an overview of narratives about online intermediaries' responses to cyberbullying from the perspectives of policy makers; the companies; as well as children and parents. Relevant self-regulatory and self-organizational efforts are discussed as well as the rationales for their adoption; and how the effectiveness of these efforts is seen from the perspectives of various stakeholders. We draw attention to the relative paucity of data on effectiveness of companies' mechanisms, particularly from the perspective of children.

Introduction

A key theme for policy development in the area of cyber abuse and bullying on online platforms has been to focus on the important role that industry plays in managing and mitigating online harm and abuse. In response to increased concern about online harms, politicians frequently argue that despite claims to being a "mere conduit" for social media interactions, industry bears a primary responsibility to address problems that arise on their platforms and that this, if not done voluntarily, could and should be enforced through regulation. This "threat" of legisla-

tion – should industry not cooperate – makes for a popular narrative in bringing about more decisive action to counteract online harm. Yet, this position is at odds with what is said *about* industry by policy makers in actual policy discourse. It also contrasts significantly with how such action is viewed *within* industry deliberations. These two competing perspectives – *of* and *in* industry – form the principal subject of this chapter and constitute two sides of an ongoing dialogue and debate about how the industry position should be best addressed within public policy. In the following, we present an overview of the principal arguments on both sides and argue that a further equally important perspective – that of children themselves – has received less than adequate attention.

Liability and Responsibility

This chapter begins within an overview of narratives of industry responses to cyberbullying from the perspectives of the distinct stakeholders: policy-makers and regulators, including e-safety experts and non-governmental organizations; from the perspective of the industry; and from children and parents. Arguably, industry encompasses various actors, such as social media companies, digital messengers and gaming apps, but also companies offering filtering solutions and software, which are designed to assist parents and caregivers in preventing their children's exposure to cyberbullying, as well as the internet and mobile phone service providers (ISPs). Our focus in this chapter, however, is less on ISPs or other supporting services including vendors of parental controls or filtering solutions. Instead, we concentrate on those providers of platforms that can act as direct venues for cyberbullying incidents, such as social media and messenger apps. In so doing, we survey the available literature on industry self-organization and self-regulation (see Marsden, 2012; Latzer et al., 2013), industry reports, as well as relevant media coverage related to the theme of industry accountability.[1]

The topic of industry involvement in addressing cyberbullying tends to revolve around questions of *responsibility*. In as much as offline bullying was predominantly confined to school environments and interpersonal relations between children, responsibility for addressing the incidents and preventing future ones was seen to be in the hands of schools, parents or caregivers and ultimately, regulators (Hinduja & Patchin, 2009; Sacco et al., 2012). Yet, in online environments, the question of responsibility and adequate policy outcomes becomes significantly more complex (O'Neill & Staksrud, 2012; O'Neill, 2013), especially when bullying happens on platforms that some legal systems internationally exempt from liability for illegal content. Social media platforms and messenger apps fall under the

---

[1]A number of observations and citations summarized in this chapter draw from the forthcoming book *Protecting Children Online? Cyberbullying Policies of Social Media Companies* (Milosevic, forthcoming from MIT Press).

wider label of "online intermediaries" or "third-party platforms that mediate between digital content and the humans who contribute and access this content" (DeNardis, 2014, p. 154). In as much as they are, by definition, not actively involved with the content that users choose to share on their platforms, intermediaries as common carriers or 'mere conduits' are exempt from liability for illegal content that they might host.

In the United States, for instance, Section 230 of the Communications Decency Act (CDA) exempts "interactive computer services" from liability for information "provided by another content provider" and hence "where an entity has provided a forum for online speech, that entity shall not be held liable for tortuous speech of others who may use the forum for harmful purposes" (Lipton, 2011, p. 1132; see also Communications Decency Act, 47 U.S.C., Section 230; Lipton, 2013; Citron, 2014). In the European Union, the Safe Harbor provisions of the Electronic Commerce Directive 2000/31/EC (or eCommerce Directive) provide similar protections (Gasser & Schulz, 2015; Wauters, Lievens & Walcke, 2016). Furthermore, the eCommerce Directive only contains provisions for content that is illegal and it does not specifically regulate content that can be harmful to minors, and for such content, it only encourages "the creation of codes of conduct for the protection of minors and human dignity," and these are not mandatory (Wauters, Lievens & Walcke, 2016, p. 242). Nonetheless, despite the limited liability, companies engage in a number of self-regulatory initiatives or otherwise develop self-organizational structures, and policies against various forms of abuse and harassment, including cyberbullying (Marsden, 2012; de Haan et al., 2013). These are the various forms of blocking and reporting (flagging) tools, as well as, in case of some companies, the mechanisms for proactive handling of abusive content or educational materials that companies tend to develop in collaboration with non-governmental organizations (e.g. Facebook Help Center, 2016; Facebook Reporting Guide, n.d.; Twitter, 2017). The central question is whether the industry is able to provide evidence of effectiveness of these mechanisms. For instance, if the platforms are not able to take down content that is determined to be cyberbullying quickly enough, thus leaving children without adequate remedies (Bazelon, 2013; Van Royen, Poels & Vandebosch, 2016), then such a state of affairs may raise concerns about whether industry self-regulation is sufficiently effective and whether command-and-control regulation is needed.

Some authors suggest that companies' liability exemptions on the grounds of being intermediaries may be re-evaluated in light of the fact that they do have a stake in online content given that they not only provide platforms for it but also capitalize on sharing, actively stirring and monetizing user interactions (van Dijck, 2013; Lipton, 2011, 2013). What can also put pressure on companies as well as policy-makers are high profile cyberbullying incidents, those that are, in some way, said to have played a role in suicides, especially if the victims are children. Such cases can seize public attention and contribute to the overall narrative that companies need to do something to address the problem (Bazelon, 2013; Milosevic, forthcoming).

Other regulatory models with respect to company responsibility and cyberbullying have emerged in the meantime. In 2016, Australia introduced a version of a social media regulator in the form of The Office of the Child Safety Commissioner, under its "Enhancing Online Safety Act for Children" (Parliament of Australia, n.d.; Office of the Children's E-Safety Commissioner, n.d.). Similar provision is under consideration in other countries such as Ireland (Law Reform Commission, 2016). Caregivers and children can report an online bullying case to the Safety Commissioner and the Commissioner can make a request to the social media platform to take the content down or else risk paying a fine for every day that the content stays there. This piece of legislation has not been evaluated as of yet, and it may be difficult at this stage to discuss its implications.

Part I Narratives of industry from the perspective of policy-makers and regulators

There are many ways to define industry self-regulation but it can be characterized as "the creation, implementation and enforcement of rules by a group of actors, industry in particular, with minimal or no intervention by the state" (Lievens, 2016, p. 77, cf. Lievens, 2010). In the policy-making community of the European Union, by way of example, a rationale for referring self-regulation rather than traditional regulation is that the industry knows its technology best while the same is thought to be true of parents with respect to their children (Staksrud, 2013 cf. Staksrud & Livingstone, 2009). Hence, allowing the industry a degree of latitude in creating mechanisms for addressing bullying on their platforms, as well as in providing adequate information about their services and advice to parents, who can then make choices regarding their children's internet use, was considered to be a preferred option to cumbersome regulation that runs the danger of stifling technological innovation. This thinking is further informed by the fear that fast paced technological innovation will also get in the way of effective regulatory enforcement (Bangemann Report, 1994; McLaughlin, 2013 cf. Svantesson, 2005; de Haan et al., 2013). One of the key arguments for self-regulation, therefore, is that when industry actors are in agreement about a policy goal, in this case, effective cyberbullying intervention, they would achieve this objective faster and in a more cost-effective manner than traditional regulation (Milosevic, forthcoming cf. Report of the Round Table on Advertising, 2006). Finally, there is a perception that traditional regulation can suffer from a knowledge gap, as it may not allow for a sufficient degree of feedback from third party experts, such as non-governmental organizations or scholars, whereas self-regulatory fora are designed with such input in mind (de Hann et al., 2013).

Another difficulty cited among the narratives about preferring not to regulate the industry via means of traditional regulation has been described as "international forum shopping." (Lievens, 2010; Newman & Bach, 2004). If companies do not

have offices and employees in a given location, the law may not apply to them. Consider that the previously cited Australian Bill only applies to companies that are legally established in the country (Vaas, 2014). Hence, if companies deem that a piece of legislation may bode negatively for their business activity, they may prefer to anchor their businesses in another jurisdiction and avoid it altogether.

Therefore, rather than regulating industry actions via command-and-control regulation, the European Commission has facilitated fora for Internet and Communication Technology (ICT) companies to participate in self-regulatory (and co-regulatory)[2] initiatives, such as the Safer Social Networking Principles, the CEO Coalition, the ICT Coalition and more recently, the Alliance to better protect minors online (see European Commission, 2009). The participating companies would then voluntarily commit to abiding by the agreed-upon principles of child protection. Among these, of the most immediate relevance for cyberbullying are the creation of reporting tools and instituting adequate means of content review (e.g. see ICT Coalition, n.d., Principles, p. 2). The companies are also asked to submit reports on how they are implementing these measures and the Commission would sometimes evaluate their efforts independently, which can (less frequently) involve testing of whether and how quickly companies respond to user reports (Staksrud & Lobe, 2010; Donoso, 2011); or which, (more frequently) involves an examination of companies' self-reports with regards to their efforts to develop anti-bullying policies and enforcement mechanisms (O'Neill, 2014).

The commitment to self-regulation continues, but the European Commission at one point raised concerns about "inconsistency" and "ineffectiveness" of self-regulation (McLaughlin, 2013, p. 87) and some authors raised concerns as to whether self-regulation is "failing" young users (Lievens, 2016, title; Staksrud, 2013). This commitment may also occasionally be put in question in the aftermath of negative publicity associated with high profile instances of bullying, the ensuing public outcry and accompanying media coverage of which may put pressure on politicians and policy-makers to take action against the companies concerned (Milosevic, forthcoming). Nonetheless, an awareness of complexities of designing effective traditional regulation encourages policy makers to act with caution.

There is, furthermore, a perception and another narrative that can sometimes be heard in the policy making community (including e-safety experts and NGOs): that the smaller and newer companies, those that enlist large numbers of users before they have acquired the necessary e-safety expertise, are problematic, whereas the older and larger companies that are also better established on the market and that regularly participate in self-regulatory initiatives (e.g. Facebook, Google, Twitter, YouTube) are less of a problem (Milosevic, forthcoming). Some experts therefore propose that guidelines for the industry best practices regarding anti-bullying enforcement mechanisms, produced as a result of multi-stakeholder meetings, could be circulated to the venture capital firms that fund the start-ups. The

---

[2] We cannot afford to address the differences in self and co-regulation here, but for more information please see McLaughlin, 2013; Marsden, 2012; and Lievens, 2010.

goal of such action would be to encourage them to make it a requirement for start-ups to have minimum safety standards in order to obtain the funding (Milosevic, forthcoming). This illustrates the narrative that policy makers try to convey to the companies and that the more established companies appear to understand --- ensuring child safety on companies' platforms is good for companies' business (Lievens, 2016; Magid, 2014).

While companies admit that there is, arguably, a limited amount of evidence on how well their tools are working for children (as we explain in Part III), some e-safety experts ask if industry self-regulatory effectiveness shouldn't be regarded as a process rather than merely as an outcome (Milosevic, forthcoming). From the perspective of some policy-makers, it is the learning that happens in multi-stakeholder, multi-cultural, self-regulatory processes that constitutes the key benefits of this type of a system. As recounted by an e-safety expert participating in the UK Council for Child Internet Safety meetings (see Milosevic, forthcoming; UKCCIS, n.d.), e-safety experts and NGOs may come into these multi-stakeholder venues with pre-conceived notions of standards of effectiveness that they would like to see implemented on the platforms. However, after hearing the companies' arguments and views, they may come to understand that some of their initial requirements may have been unrealistic. Likewise, participating companies are able to hear and understand the perspectives of independent experts and the parties are sometimes able to meet half way. According to the expert, it is this dialog and learning process that takes place in such meetings where each side re-examines its position in light of safety standards but also technological and financial feasibility, which can be as important as the outcome effectiveness.

Part II Narratives in industry

Industry, particularly those more established companies that have previously participated in a number of self-regulatory initiatives (e.g. Facebook, Twitter, Google-owned YouTube), increasingly appear to appreciate that ensuring that their platforms *are* safe but also that they are *perceived* as safe, both among policy makers and in the wider public (particularly among parents, caregivers and educators), is good for their business (Magid, 2014; Lievens, 2016; Milosevic, forthcoming). Some of these companies narrate that they do not see their anti-bullying strategies as corporate social responsibility or as a burden on their company's business model, but rather, as the daily work that they do. Safety is an important aspect of how they discursively position their corporate ethos (Milosevic, forthcoming). Most companies have provisions in their corporate documents (either in their Terms of Service or TOS, or "Community Guidelines/Standards/Principles" see e.g. Facebook, 2016) that abuse, harassment, or cyberbullying are not allowed on their platforms. Due to large amounts of content that are being shared on social media, companies can emphasize how difficult it is for them to proactively monitor content on their platforms for bullying. Consider, for instance, that 300 hours

of video was at one point uploaded onto YouTube *per minute* (Dormehl, 2015). This is why providing users with tools to block the abusive ones or report abusive content to the company is considered as a viable option. Once the content has been reported, the company will look into the case and decide if it violates its policy, and thus take it down or address the abusive user in another way in a process known as moderation (Crawford & Gillespie, 2016). Moderation can be done by humans, and some industry actors perceive it to be particularly important to employ human moderators because determining whether a case constitutes cyberbullying can be context-dependent and difficult to handle in an automated way – necessitating a human eye. However, companies typically do not disclose the extent to which their moderation is automated (Milosevic, 2016; Milosevic, forthcoming).

Only if the content is determined to be cyberbullying, as defined by the company, and thus violating the company policy, will the companies take it down (see e.g. Matias et al., 2015). However, from companies' experience, it can be difficult to determine whether a piece of reported content actually constitutes a policy violation. Consider if a photo is reported to the company with no abusive words underneath and without any additional contextual information that would suggest the case constitutes a violation. If the photo itself has no insignia of abuse, the moderation process may not be able to determine it to be a violation. Nonetheless, the user might find the photo abusive for some reason and cannot get the poster to take it down. This is why some of the established companies see users as better positioned to address some of the inter-personal problems on their platforms and they seek to develop tools that would actively empower users in solving these problems on their own. Consider for instance, Facebook's social reporting, which was developed upon extensive research collaboration with scientists from Yale University and The University of California, Berkeley, as well as through Facebook in-house research (Yale Center for Emotional Intelligence, 2013; Greater Good Science Center, 2015). It allows users to leverage pre-made messages, which would ask in a polite way the user who posted offensive content to take it down. Users could also rely on pre-made messages to reach out to a third party (friend or trusted adult) to let them know they're being bullied and that they need help.

In some of the companies' narratives, tools like these were considered as more advanced than simple content takedowns, as content takedowns may not always solve the problem, which may persist offline or on other platforms (Compassion Research Day, 2013; Milosevic, 2016, Milosevic, forthcoming). Tools like these, on the other hand, are thought to be able to help users address the heart of the conflict. Other platforms developed more automated tools with the logic of user empowerment by allowing a greater degree of moderation to users, rather than company moderators. One example is Yik Yak's "up-voting and downvoting" (the logic similar to like and dislike) whereby if a piece of content was "downvoted" by the community of users five times –it would then be taken down automatically.

Companies are increasingly looking into proactive ways to handle abuse on their platforms (without having to rely on user reporting first) and these include

various applications of algorithmic learning (e.g. supervised machine learning, natural language processing) or artificial intelligence-like tools (Greenberg, 2016). Some of these platforms report to hold freedom of speech as one of their primary values and hence see themselves as being in the difficult position of having to balance user safety and protection from cyberbullying on the one hand, against the need to protect free speech on the other (Gillespie, 2010). They may, therefore, report to be reluctant to use specific forms of proactive moderation tools, such as in the application of supervised machine learning that would allow them to crawl their platforms for language-based indicators of bullying. For instance, if the word "bitch" was to be such an indicator, the company might receive too many false alerts, given that the word "bitch" is frequently used by young people in a friendly rather than an offensive way (i.e. to mean "friend" or "mate") (Milosevic, 2016). Companies, nonetheless, may not specify whether and what types of proactive content crawling they might be using in order to identify potentially cyberbullying content before it is reported to their moderation system and would typically not specify in their policies as to which proactive enforcement mechanisms they use and how they are using them. There is a significant lack of transparency as to how bullying moderation is executed and companies' operational policies are typically not shared with the public. A recent leak of Facebook's operational policies, caused significant controversy and public debate with respect to child protection (Hopkins, 2017). The company's operational anti-bullying policies were said not to be either strict enough; or that they were so specific to the point of being rigid, thus allowing bullying to slip through the cracks of the company's moderation system. For instance, among the items that caused criticism was that the company previously allowed "the 'sharing of footage of physical bullying of children under seven" provided that this was done without a caption (Hopkins, 2017, para 7).

In line with some of the recommendations from self-regulatory initiatives, a number of companies are developing educational and awareness-raising advice for children and caregivers, typically housed in companies' Safety and Help Centers, specially designated sections of companies' websites that would contain instructions for using companies' anti-bullying tools (see e.g. Snapchat, n.d.; Yale Center for Emotional Intelligence, 2013). In doing so, companies may partner with non-governmental organizations (NGOs) who provide expert, third party advice (O'Neill, 2014). Companies tend to cite these activities as evidence of their continuous efforts to address cyberbullying in an effective way. Some of this content emphasizes "digital citizenship" and describes users as important actors in ensuring that the platform is safe (Jones & Mitchell, 2015). Young users are encouraged to be good digital citizens who do not engage in bullying and who also help victims when they find themselves in bystander roles. Most importantly, such narratives position responsibility for safety on the platform as not only pertaining to the platform itself and its owners, but as *shared* between the platform and its community of users (and hence the corporate documents' characterization as "community" guidelines/standards/rules); as well as with NGOs, caregivers and regulators (Milosevic, forthcoming).

Part III Narratives from children and parents

Despite the extensive commentary within policy circles and in wider public domain regarding narratives of industry-responses to cyberbullying, when it comes to the perspective of children and parents, there is limited evidence available. Only a handful of studies have inquired whether young users are aware of companies' anti-bullying tools or whether they find them helpful. Among those, several have raised concerns about companies' responsiveness to children's complaints and found that teens were largely unaware of social media companies' efforts to stop bullying on their platforms, while those that had been aware of them, doubted their effectiveness (Van Royen et al., 2016; Schneider, Smith & O'Donell, 2013, p. 22). One of these studies in the United States, found that parents, too, were to a great extent unaware of social media companies' efforts against cyberbullying, and that, while they accepted responsibility for their children's safety on the sites, they would often lack the capability to ensure such safety. These parents would like companies to provide more mechanisms for restricting and monitoring various behaviors of their children on these sites, but they also expressed a desire for companies to provide digital citizenship advice (Schneider, Smith & O'Donell, 2013, p. 18). An unpublished, exploratory study in Norway conducted by the first author of this chapter, which relied on a small-scale[3] survey, found that just over a half of survey respondents said they knew how to report bullying on sites they used (almost 60%). Less than a third reported to have done it at least once in their lives, and only 17% of those who report to have been bullied reported to a social media platform the last time such abuse had happened. Less than a third of children reported to have seen companies' Safety or Help centers, and even fewer-13%- reported to have seen Facebook's advanced policy of social reporting. One of the less common examples of an attempt at providing publicly available evaluation of companies' efforts against cyberbullying resulted from collaboration between a UK-based non-governmental organization, the NSPCC, and the telecommunications company O2 (NewsO2, 2016). Relying on a survey with children as well as parent panels, the project asks whether they are aware of major companies' reporting tools and whether they find them useful. The project then provides the rating for each platform's safety in a visually simple way for parental and children's use. While the results vary by platform from satisfactory to less satisfactory, they are also limited to the United Kingdom. For instance, Facebook's reporting tools were rated as more satisfactory (as indicated with a "happy smiley") than Instagram's, Snapchat's and Kik's ("an indifferent or straight-face smiley").

The amount of published accounts regarding how children conceptualize company responsibility for helping them in bullying cases, an issue relevant from the

---

3 N= 152; Age of respondents: 11-19; the school that children were sampled from was international in character and it included children from a variety of countries and cultures. English was the primary language in the school.

perspective of children's rights (Livingstone & Bulger, 2014; Livingstone, Carr & Byrne, 2015), is limited. The above-mentioned exploratory study in Norway asked the children via survey and focus groups whether they thought companies should take responsibility for bullying on their platforms. While the respondents assigned a significant degree of responsibility to the companies, some of the focus group respondents perceived that there was only so much that the companies could do in the face of pervasiveness of bullying on their sites.

Conclusion: Narratives in the face of Evidence

Narratives related to industry responses to cyberbullying converge around the complex problem of responsibility and effectiveness of platforms' efforts. While companies tend to be legally exempted from liability, they nonetheless develop a variety of mechanisms, sometimes as part of sustained efforts on behalf of policy makers and experts within self-regulatory initiatives, or as part of self-organizational structures that could fall under the umbrella of companies' corporate social responsibility efforts. The palpable data on effectiveness of these mechanisms (statistics on how many bullying reports are received on a given platform per unit in time, how fast these are processed and how satisfied children and parents are with this processing and available mechanisms in general) is to a great degree lacking when it comes to more established and newer companies alike. Nonetheless, older and more established social media companies that participate in self-regulatory initiatives and police their platforms in an effort to exhibit best behavior can be perceived by regulators as less of a problem than the start-ups whose user base may outgrow their e-safety expertise. The voices that appear to be consistently missing among the many narratives on industry efforts, are those of children. To this date, there exists a limited amount of continuous effort internationally that would ask children, across platforms, whether they find the mechanisms as able to address bullying problems, why or why not this may be the case, and then consider adequate measures in accordance with children's responses.

References

Bangemann Report, Europe and The Global Information Society: Recommendations to the European Council. (1994). Retrieved from: http://www.cyber-rights.org/documents/bangemann.htm

Bazelon, E. (2013b). *Sticks and stones: Defeating the culture of bullying and rediscovering the power of character and empathy*. New York, NY: Random House.

DeNardis, L. E. (2014) *The global war for Internet governance*. New Haven, CT: Yale University Press.

de Haan, J., van der Hof, S., Bekkers, W., & Pijpers, R. (2013). Self-regulation. In: B. O'Neill, E. Staksrud & S. McLaughlin, *Towards a better internet for children? Policy pillars, players and paradoxes* (pp. 111-129). Göteborg, Nordicom

Citron, D. K. (2014). *Hate Crimes in Cyberspace*. Harvard University Press: Cambridge, MA.

Compassion Research Day. (2013). Retrieved from http://new.livestream.com/facebooktalkslive/events/2564173

Communications Decency Act of 1996, (CDA), Pub. L. No. 104-104 (Tit. V), 110 Stat. 133 (Feb. 8, 1996), codified at 47 USC 223, 230.

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*, *18*(3), 410–428. doi:10.1177/1461444814543163

Donoso, V. (2011). Results of the Assessment of the Implementation of the Safer Social Networking Principles for the EU. Individual Reports of Testing of 14 Social Networking Sites. Retrieved from https://lirias.kuleuven.be/bitstream/123456789/458077/1/Individual+Reports+SNS+Phase+A.pdf.

Dormehl, L. (2015, January 28). 300 hours of footage per minute: Google explains why policing YouTube is so tough. *Fast Company*. Retrieved from http://www.fastcompany.com/3041622/fast-feed/300-hours-of-footage-per-minute-google-explains-why-policing-youtube-is-so-tough

European Commission. (2009) Safer social networking principles for the EU. Retrieved from: https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/sn_principles.pdf

Facebook (2016). Community Standards. Helping to Keep You Safe. Retrieved from: https://www.facebook.com/communitystandards#

Facebook Help Center (2016). How to report things? Retrieved from: https://www.facebook.com/help/181495968648557?helpref=faq_content

Facebook Reporting Guide (n.d.). Retrieved from: https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-xpa1/t39.2178-6/851563_293317947467769_1320502878_n.png

Gasser, U. & Schulz, W. (2015). Governance of Online Intermediaries: Observations from a Series of National Case Studies. Retrieved from: https://publixphere.net/i/noc/page/NoC_Online_Intermediaries_Research_Project_Synthesis

Gillespie, T. (2010). The Politics of 'Platforms'. *New Media & Society, 12*(3), 347-364.

Greater Good Science Center. (2015). Greater good. Retrieved from http://greatergood.berkeley.edu/

Greenberg, A. (2016, September 9). Inside Google's Internet Justice League and its AI-powered War on trolls. The Wired. Retrieved from: https://www.wired.com/2016/09/inside-googles-internet-justice-league-ai-powered-war-trolls/?mbid=social_twitter

Hinduja, S., & Patchin, J. W. (2009). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying.* Thousand Oaks, CA: SAGE.

Hopkins, N. (2017, May 22). How Facebook allows users to post footage of children being bullied. *The Guardian*. Retrieved from: https://www.theguardian.com/news/2017/may/22/how-facebook-allows-users-to-post-footage-of-children-being-bullied

ICT Coalition (n.d.). ICT coalition. Retrieved from http://www.ictcoalition.eu/

Jones, L. M., & Mitchell, K. J. (2015). Defining and measuring youth digital citizenship. *New Media & Society,* doi:10.1177/1461444815577797

Latzer, M., Just, N., & Saurwein, F. (2013). Self –and co –regulation. Evidence, legitimacy and governance choice. In E. P. Monroe, S. G. Verhulst, & L. Morgan (Eds.), *Routledge handbook of media law* (pp. 373–397). Oxon: Routledge.

Law Reform Commission. (2016). *Final Report on Harmful Communications and Digital Safety*. Dublin: Law Reform Commission.

Lievens, E. (2010). *Protecting children in the digital era: The use of alternative regulatory instruments*. Leiden, Netherlands: Martinus Nijhoff.

12

Lievens, E. (2016). Is self-regulation failing children and young people? Assessing the use of alternative regulatory instruments in the area of social networks. In S. Simpson, H. Van den Bulck, & M. Puppis (Eds.), *European media policy for the twenty-first century: Assessing the past, setting agendas for the future* (pp. 77–94). New York, NY: Routledge.

Lipton, J. D. (2011). Combating cyber-victimization. *Berkeley Technology Law Journal, 26*(2), 1103-1155.

Lipton, J., D. (2013). Cyberbullying and the First Amendment. *Florida Coastal Law Review, 14*(99), 99-130.

Livingstone, S., & Bulger, M. (2014): A Global Research Agenda for Children's Rights in The Digital Age, Journal of Children and Media, DOI: 10.1080/17482798.2014.961496

Livingstone, S., Carr, J. & Byrne, J. (2015). One in Three: Internet Governance and Children's Rights. Global Commission on Internet Governance. Retrieved from: https://www.cigionline.org/publications/one-three-internet-governance-and-childrens-rights

Magid, L. (2014, August 14). IAC's Ask.com buys Ask.fm and hires a safety officer to stem bullying. Forbes. Retrieved from http://www.forbes.com/sites/larrymagid/2014/08/14/iacs-ask-com-buys-ask-fm-and-hires-a-safety-officer-to-stem-bullying/

Marsden, C.T. (2012) Internet co-regulation and constitutionalism: Towards European judicial review, *International Review of Law, Computers & Technology*, *26*(2-3), 211-228, DOI: 10.1080/13600869.2012.698450

Matias, J. N., Johnson, A., Boesel, W. E., Keegan, B., Friedman, J., & DeTar, C. (2015). Reporting, reviewing, and responding to harassment on Twitter. *Women, Action, & the Media*. Retrieved from http://womenactionmedia.org/twitter-report

McLaughlin, S. (2013). Regulation and legislation. In: B. O'Neill, E. Staksrud & S. McLaughlin, *Towards a better internet for children? Policy pillars, players and paradoxes* (pp. 77-91). Göteborg, Nordicom

Milosevic, T. (2016). Social Media Companies' Cyberbullying Policies.*International Journal Of Communication, 10*, 22. Retrieved from http://ijoc.org/index.php/ijoc/article/view/5320

Newman, A.L. & Bach, D. (2004). Self-regulatory trajectories in the shadow of public power: Resolving digital dilemmas in Europe and the United States. *Governance: An International Journal of Policy, Administration and Institutions, 7*(3), 387—413.

News O2. (2016, April 6). Children reveal riskiest social media sites. Retrieved from: http://news.o2.co.uk/?press-release=children-reveal-riskiest-social-media-sites-new-net-aware-guide

O'Neill, B. (2013). Who cares? Practical ethics and the problem of underage users on social networking sites. *Ethics and Information Technology, 15*(4), 253-262.

O'Neill, B. (2014). First report on the implementation of the ICT Principles, Retrieved from: http://www.ictcoalition.eu/gallery/75/ICT_REPORT.pdf

O'Neill, B., & Staksrud, E. (2012). Policy implications and recommendations: Now what? In S. Livingstone, L. Haddon & A. Görzig, *Children, risk and safety on the Internet: Research and policy challenges in comparative perspective* (pp. 339-354). Bristol: The Policy Press

Parliament of Australia (n.d.). Parliament of Australia website system maintenance. Retrieved from http://outageaph.aph.gov.au/

Report of the Round Table on Advertising. (2006). Retrieved from: http://ec.europa.eu/consumers/archive/overview/report_advertising_en.pdf

Safer Social Netowkring Principles for the EU. (2009). Retrieved from: https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/sn_principles.pdf

Schneider, S. K., Smith E., & O'Donnel, L., (2013b). Social media and cyberbullying: Implementation of school-based prevention efforts and implications for social media approaches. Retrieved from http://www.promoteprevent.org/sites/www.promoteprevent.org/files/resources/Social_Media_and_Cyberbullying_FinalReport-EDC_0.pdf

Sacco, D., Silbaugh, K., Corredor, F., Casey, J., & Doherty, D. (2012). *An Overview of state anti-bullying legislation and other related laws.* Retrieved from

http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/State_Anti_bullying_Legislatio n_Overview_0.pdf

Snapchat. (n.d.). Safety Center. Retrieved from: https://www.snapchat.com/safety

Staksrud, E. (2013). *Children in the Online World: Risk, Regulation and Rights*. Ashgate: London.

Staksrud, E., & Lobe, B. (2010). Evaluation of the implementation of the Safer Social Networking Principles for the EU, Part 1: General report (Study commissioned by the European Commission), Retrieved from http://ec.europa.eu/information_society/activities/social_networking/docs/final_report/first_p art.pdf

Staksrud, E., & Livingstone, S. (2009). Children and online risk: powerless victims or resourceful participants? *Information, Communication & Society 12(*3), 364-387. Retrieved from http://eprints.lse.ac.uk/30122/1/Children_and_online_risk_%28LSERO_version%29.pdf

Svantesson, D.J.B. (2005). The characteristics making internet communication challenge traditional models of regulation. What every international jurist should know about the internet. *International Journal of Law and Information Technology*, 13(1) 39-69.

Twitter (2017). Our Trusted Partners. Retrieved from: https://support.twitter.com/groups/57-safety-security/topics/274-handling-issues-online/articles/20171366-trusted-resources   UK Council for Child Internet Safety. (n.d.).

UK Council for Child Internet Safety (n.d.). Retrieved from: https://www.gov.uk/government/groups/uk-council-for-child-internet-safety-ukccis

van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford, UK: Oxford University Press.

Van Royen, K., Poels, K., & Vandebosch, H. (2016). Help, I am losing control! Examining the reporting of sexual harassment by adolescents to social networking sites. *Cyberpsychology, Behavior and Social Networking*, *19*(1), 16–22.

Vaas, L. (2014, October 30). Snapchat escapes Australian cyberbullying crackdown, for now. *Naked Security*. Retrieved from: https://nakedsecurity.sophos.com/2014/10/30/snapchat-escapes-australian-cyberbullying-crackdown-for-now/

Wauters, E., Lievens, E., & Valcke, P. (2016). Empowering children through labeling in social networks: Illusion or Solution? In M. Walrave, K. Ponnet, E. Vanderhoven, J. Haers, & B. Segært (Eds.). *Youth 2.0: Social Media and Adolescence: Connecting, Sharing & Empowering*, (pp. 227–249). Springer.

Yale Center for Emotional Intelligence (2013). Introducing the Facebook Bullying Prevention Hub. Retrieved from http://ei.yale.edu/introducing-the-facebook-bullying-prevention-hub/