Chapter 12
On Technology Against Cyberbullying

Janneke M. van der Zwaan, Virginia Dignum, Catholijn M. Jonker and Simone van der Hof

Contents

Janneke van der Zwaan is a Ph.D. Candidate in Human–Computer Interaction at Delft University of Technology.Virginia Dignum is Associate Professor at the Faculty of Technology, Policy and Management, Delft University of Technology. Catholijn Jonker is Full Professor of Man–Machine Interaction at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology. Simone van der Hof is Full Professor of Law and Information Society and chair of eLaw, the Center for Law in the Information Society, Leiden University.An extended version of this chapter will appear in Van den Hoven et al. (forthcoming) *Responsible Innovation* Volume 1: Innovative Solutions for Global Issues.

J. M. van der Zwaan (✉) · V. Dignum · C. M. Jonker
Delft University of Technology, Delft, The Netherlands
e-mail: J.M.vanderZwaan@tudelft.nl

S. van der Hof
Center for Law in the Information Society, Leiden University, Leiden, The Netherlands

S. van der Hof et al. (eds.), *Minding Minors Wandering the Web: Regulating Online Child Safety*, Information Technology and Law Series 24, DOI: 10.1007/978-94-6265-005-3_12,

12.1 Introduction

Nowadays, many children and adolescents spend a lot of time online.[1] The Internet is used not only as an educational tool, but also for fun, games and to develop and maintain social contacts. One of the risks children and adolescents run online is to become a victim of cyberbullying. Cyberbullying can be defined as 'any behavior performed through electronic or digital media by individuals or groups that repeatedly communicates hostile or aggressive messages intended to inflict harm or discomfort on others'.[2]

Recently, cyberbullying gained a lot of attention. There have been a number of cases involving online bullying with extreme consequences for those involved that have received extensive media coverage in the US and Western Europe. Additionally, in the academic world, studies are conducted to map the problem of cyberbullying and its consequences for victims, bullies, and bystanders. With victimization rates ranging from 20 to 40 %,[3] cyberbullying is a common risk for children and adolescents. In addition, recent findings from the EU Kids Online II survey indicate that cyberbullying has a high impact on victims.[4]

Antisocial behavior such as cyberbullying can be regulated socially, legally, and/or technologically.[5] Social norms play an important role in regulating behavior in general. Law also regulates behavior; "[different] laws all continue to threaten ex post sanction for the violation of legal rights".[6] Technology can control or steer social behavior through functionalities in the software design (coined 'code as law' by Lessig[7]) or through exerting social influence (persuasive technology[8]). The different modalities are connected and their interaction is complex.[9] Generally, complex problems such as cyberbullying cannot be solved by measures from a single modality alone; better solutions may be found in a combination of measures from different modalities.

This chapter focuses on using technology to protect and empower children and adolescents against cyberbullying. So far, this topic has received little attention).[10]

---

[1] The Gallup Organisation 2008; Eurobarometer 2007.

[2] Tokunaga 2010.

[3] Tokunaga 2010.

[4] Livingstone et al. 2010.

[5] Instead of three, Lessig distinguishes four modalities for regulation: social norms, the law, architecture and the market; Lessig 2000. In the case of cyberbullying, the market is not or less relevant as a modality for regulation and will therefore not be addressed in this chapter.

[6] Lessig 2006, p. 124.

[7] Lessig 2000.

[8] Fogg 2002.

[9] Lessig 2006.

[10] Exceptions are Internet Safety Technical Task Force 2008; Szwajcer et al. 2009; Mesch 2009.

However, recently different initiatives have started to investigate the regulation of cyberbullying through technology, such as AMiCA[11] and Friendly ATTAC.[12] Existing work seems to rely on the assumption that general Internet safety technologies can be used as protection against cyberbullying as well. In this chapter, we show that this assumption is mostly unfounded and propose an alternative approach to addressing cyberbullying with technology.

This chapter is organized as follows. In Sect. 12.2, we provide a background on Internet safety technology and cyberbullying. In Sect. 12.3, this information is used to construct a framework of characteristics that technology against cyberbullying should have to be able to protect against cyberbullying. In Sect. 12.4, we use the framework to discuss the expected effectiveness of existing Internet safety technologies against cyberbullying. Finally, in Sect. 12.5, we present our conclusions.

12.2 Background

12.2.1 Internet Safety Technology

Online safety of children and adolescents concerns risks such as harassment, bullying, sexual solicitation, exposure to problematic and illegal content (including pornography, and violence), malicious software (for instance, viruses), hackers, and online delinquency (for example, identity theft). In their review of existing Internet safety technology, the Technology Advisory Board of the Internet Safety Technical Task Force distinguished the following functional goals[13]:

- Limit harmful contact between adults and minors,
- Limit harmful contact between minors,
- Limit/prevent minors from accessing inappropriate content on the Internet,
- Limit/prevent minors from creating inappropriate content on the Internet,
- Limit the availability of illegal content on the Internet,
- Prevent minors from accessing particular sites without parental consent,
- Prevent harassment, unwanted solicitation, and bullying of minors on the Internet.

These goals show that Internet safety technology is restrictive; they clearly intend to restrict online behavior. This view on technology corresponds to the aforementioned ‘code as law’ perspective from Lessig. Web filtering software is an example of restrictive technology; a Web filter blocks access to websites based on certain criteria.

---

[11] See www.clips.ua.ac.be/amica/.

[12] See www.friendlyattac.be/en/.

[13] Internet Safety Technical Task Force 2008.

Different types of Internet safety technologies can be distinguished, including[14]:

- Content and behavior analysis,
- Filtering,
- Monitoring,
- Blocking undesirable contacts,
- Reporting,
- Age/identity verification, and
- Educational technology.

Some technologies, such as age/identity verification, require storing personal data, which raises privacy concerns. Monitoring online behavior or automatically analyzing online communication might also invade privacy. In addition, restrictive technology could violate the rights to freedom of information and expression. Children's privacy and their rights to freedom of information and expression must be balanced against the potential benefits of Internet safety technologies. In some cases it might be appropriate to restrict behavior, for example, to protect younger children, whereas for older children and adolescents protecting their privacy and/or freedom of information and expression might be more important.

12.2.2 Cyberbullying

Research on cyberbullying is still in the early stage. Little is known beyond prevalence, frequency among specific groups, and negative outcomes.[15]

1. Compared to traditional bullying

Cyberbullying—by definition—is a type of bullying. According to Olweus,[16] bullying is aggressive behavior or intentional 'harm doing' which is carried out 'repeatedly and over time' and in an interpersonal relationship characterized by an imbalance of power. Additionally, cyberbullying has some specific characteristics. First, cyberbullies can remain anonymous relatively easy.[17] Another important difference is the lack of physical and social cues in online communication.[18] This prevents the bully from being confronted with the consequences of the harassments[19] and could also lead to misinterpreting messages as cyberbullying when in

[14] Internet Safety Technical Task Force 2008; Szwajcer et al. 2009.

[15] Tokunaga 2010.

[16] Olweus 1999.

[17] Ybarra and Mitchell 2004; Patchin and Hinduja 2006; Kowalski and Limber 2007; Shariff 2008.

[18] Ybarra and Mitchell 2004; Patchin and Hinduja 2006; Kowalski and Limber 2007; Kowalski et al. 2008.

[19] Kowalski et al. 2008.

fact they were not intended to be.[20] A third difference is the 24/7-attainability provided by online communication.[21] Traditional bullying is usually characterized by a confined period of time during which bullies have access to their victims. In most cases, victims of traditional bullying are safe at home. This is no longer the case with cyberbullying. Other differences between traditional bullying and cyberbullying are the quick distribution of electronic messages to (potentially) infinite audiences[22] and the permanent nature of information on the Internet.[23]

2. Types, Media, and Methods

Cyberbullying refers to bullying through electronic communication devices. It happens through different media, such as e-mail, instant messenger applications, social networking websites, blogs, chat rooms, online games, virtual worlds, and mobile phones (sms). Cyberbullying can be communication-based or content-based. Methods used for online bullying include name-calling, gossiping, ignoring, threatening, spreading personal conversations, manipulating and spreading pictures, creating defamatory websites, and sending sexual comments.[24]

3. Victims

Prevalence rates of cyberbullying victimization vary among studies. In a recent review of existing research, Tokunaga reports victimization rates of 20–40 %.[25] Cyberbullying victims tend to be heavier Internet users than youth that is not victimized.[26] Victims of traditional bullying and those that bully others online are also more likely to be cyberbullied.[27]

4. Bullies

Online bullies are typically the same age as their victims.[28] And even though anonymity is often viewed as integral to cyberbullying, it seems that cyberbullying often takes place in the context of social groups and relationships.[29] Online bullying has a strong connection with the offline world; between 44 and 82 % of cyberbullying victims know their bullies offline.[30]

---

[20] Ybarra et al. 2007.

[21] Patchin and Hinduja 2006; Kowalski and Limber 2007.

[22] Kowalski and Limber 2007; Kowalski et al. 2008; Shariff 2008.

[23] Shariff 2008.

[24] Dehue et al. 2008; Vandebosch and Cleemput 2008.

[25] Tokunaga 2010.

[26] Smith et al. 2008.

[27] Ybarra et al. 2006; Li 2007.

[28] Patchin and Hinduja 2006; Wolak et al. 2006; Kowalski and Limber 2007; Hinduja and Patchin 2009.

[29] Mishna et al. 2009.

[30] Wolak et al. 2006; Hinduja and Patchin 2009.

5. Tackling Cyberbulling

Because cyberbullying is a phenomenon that only recently emerged, validated approaches to stop or prevent it do not yet exist. However, some researchers made suggestions on how to tackle the problem.

Many studies stress the importance of education and awareness to reduce and prevent cyberbullying. Ybarra et al. support the idea to include cyberbullying prevention in conventional anti-bullying programs.[31] It is important to educate both children and adults (e.g., teachers and parents).[32] Educating parents and other adults might make it easier for children and adolescents to talk to them about their negative online experiences.[33] Teaching technological skills—again, both to children and adults—deserves special attention, so children and adults know what can be done in certain situations.[34]

12.3 The Framework

In order to discuss the expected effectiveness of existing technology against cyberbullying, we constructed a framework consisting of desired characteristics of technology against cyberbullying. These characteristics are derived from important topics that emerge from the literature on Internet safety technology and cyberbullying when taking the perspective of a (potential) cyberbullying victim and looking at the direct consequences for his/her online experience. To identify the desired characteristics, we started with the following basic questions: what are the online behaviors that can be characterized as cyberbullying?, who are the bullies?, and when do users need protection? While using the literature to answer these questions, we took as a starting point the principle that all Internet users (including bullies and victims) should be restricted in their behavior as little as possible and that it is better to learn potential victims to deal with antisocial behavior such as cyberbullying than to attempt preventing them from coming into contact with these types of behavior at all. Subsequently, we identified some risks associated with online technology in general.

Online behaviors that can be characterized as cyberbullying are diverse; different types, media and methods can be used to cyberbully others. Like traditional bullying, cyberbullying usually is communication-based (for example, name-calling in chat conversations or sending threatening e-mails), but content-based cyberbullying also occurs (for example, creating a fake profile on a social network or posting manipulated pictures). Technology against cyberbullying should take

[31] Ybarra et al. 2006.

[32] Ybarra et al. 2006; Dehue et al. 2008.

[33] Ybarra et al. 2006.

[34] Finkelhor et al. 2000; Smith et al. 2008.

into account different types, media, and methods of cyberbullying and at least target online communication.

Recent studies reveal that many of the online threats experienced by children and adolescents are perpetrated by peers, including sexual solicitation[35] and online harassment.[36] Additionally, victims usually know their bullies in real life.[37] Therefore, technology against cyberbullying should at least take into account relationships with known and unknown peers.

Cyberbullying can occur at any moment. This 24/7 attainability of cyberbullying is enabled by technology. Technology against cyberbullying should also be available at any moment and/or be able to intervene at any moment. In other words, technology against cyberbullying should provide real-time support.

Technology in general has some risks that might limit its suitability to protect against cyberbullying. For example, in Sect. 12.2 we observed that existing Internet safety technology always restricts users in some way. A disadvantage of restrictive technology is that it can be circumvented relatively easily by computer-savvy users. It is very hard to force people to use some technology. Therefore, it is suggested that technology against cyberbullying should rely on voluntary use. Victims (and potentially bystanders) are motivated to use some technology if they have something to gain (they want to stop the bullying), while cyberbullies are less likely to participate voluntarily, because bullying is an intentional act.

Additionally, technology might invade privacy and/or limit freedom of expression. Although these issues are beyond the scope of this chapter, they are very important. Children's privacy and their right to freedom of expression should be balanced carefully against the potential benefits of technology against cyberbullying. Therefore, protection of privacy and freedom of expression are included in the framework.

The desired characteristics are summarized in Table 12.1. We would like to emphasize that this list should not be regarded as the only one possible or as a definitive set of characteristics for technology against cyberbullying. Instead, it is intended as a starting point for discussing technology in the context of cyberbullying. In the next paragraph, these characteristics are used to assess the expected effectiveness of existing technology against cyberbullying.

12.4 Existing Internet Safety Technologies

This paragraph reviews existing Internet Safety technologies and discusses their expected suitability against cyberbullying based on the framework proposed in Sect. 12.3. The following technologies are discussed: content and behavior

[35] Wolak et al. 2006.

[36] Smith et al. 2008; Hinduja and Patchin 2009.

[37] Mishna et al. 2009.

Table 12.1 Desired characteristics for technology against cyberbullying

Characteristics
a. Suitable for different types, media and methods
b. Take peer contact into account
c. Real-time
d. Voluntary use
e. Protecting the user's privacy
f. Protecting the user's freedom of speech

analysis, filtering, monitoring, blocking undesirable contacts, reporting, age/identity verification, and educational technology. Most existing parental control applications, such as Net Nanny[38] or Cyber Patrol,[39] combine multiple technologies, including content and behavior analysis, filtering, and monitoring in one product. Below, we focus on the separate technologies, not complete applications.

12.4.1 Content and Behavior Analysis

Content and behavior analysis are about automatically extracting meaningful information from data, such as text, images, video material, and network traffic. Potentially, these techniques can also be used to detect cyberbullying in text-based conversations.

Preliminary results on related tasks show that it is rather difficult to automatically recognize different types of harassment. Pendar used a statistical approach to automatically distinguish between communication of sexual predators and victims.[40] Classifier performance ranged from 40 to 95 %. Kontostathis et al. also attempted to recognize sexual predation and the resulting classifier correctly predicted predator speech 60 % of the time.[41] These results seem promising, however, these studies have some limitations. First, the datasets used for the experiments were small (701 and 25 conversations respectively[42]). Standard corpora for text classification contain hundreds of thousands texts (e.g., the Reuters corpus[43]). Second, the data used consisted of conversations that were known to be malicious; most online conversations are not. Data imbalance (data sets containing

[38] See www.netnanny.com/.

[39] See www.cyberpatrol.com/.

[40] Pendar 2007.

[41] Kontostathis et al. 2009.

[42] Pendar 2007 and Kontostathis et al. 2009 both used data made available by Perverted Justice (www.perverted-justice.com/).

[43] Lewis et al. 2004.

only a few objects that need to be detected) is a well-known problem in machine learning that leads to suboptimal classifier performance.[44]

In 2009, the Content Analysis for the Web 2.0 Workshop (CAW2.0) offered a shared task on misbehavior detection.[45] Yin et al. trained classifiers to identify harassing messages in chat and online discussion forums.[46] Performance was between 25 and 40 %, so, there is much room for improvement.

Automatically recognizing cyberbullying or other harmful content could be a first step in protecting children and adolescents against these threats. As mentioned before, most applications for parental control employ some form of content analysis. Content and behavior analysis can be used to detect different forms of cyberbullying, both communication-based and content-based. However, related work shows that detecting different types of harassment is not trivial and needs to be improved before it can be used as (partial) protection against cyberbullying. The technology can be applied to all communication, including peer communication. In addition, content and behavior analysis can be both used voluntary and non-voluntary. It can be applied in real-time. Because technology for content and behavior analysis stores and interprets online behavior, which can be considered personal data, the privacy of users might be invaded. Detecting inappropriate data does not limit the freedom of expression per se, but actions taken after something has been detected might.

12.4.2 Filtering

Web-filtering software blocks access to websites with inappropriate content, such as pornography. Filtering techniques include white lists (lists of websites the user is allowed to visit), black lists (lists of websites the user is not allowed to visit), and content analysis (the content analysis algorithm decides whether the user is allowed to visit a website, e.g., based on the occurrence of certain key words). Common problems with Web filtering are underblocking (fail to block access to websites with inappropriate material) and overblocking (block websites that do not contain inappropriate material). Hunter evaluated four commercial Web-filtering applications. He found the applications blocked 75 % of a collection of inappropriate material and 21 % of a collection of appropriate material.[47]

Filtering is a preventive measure. It does not specifically target communication, but filtering incoming and/or outgoing communication could limit or prevent harmful contact between minors and between minors and adults. However, automatically recognizing either communication-based or content-based cyberbullying is not a trivial task (see Sect. 12.4.1). Filtering technology does not exclude

[44] Chawla et al. 2004.

[45] See http://caw2.barcelonamedia.org/.

[46] Yin et al. 2009.

[47] Hunter 2000.

communication between peers. Because users do not get the choice to apply filtering or not before they go online, filtering does not rely on voluntary use. Filtering software may be circumvented. For example, it is very easy to substitute terms that are filtered for unfiltered terms that are equally offending, for example 'loser' becomes 'l o s e r', 'L0S3R', 'looser', etc. Filtering software is real-time technology; websites are blocked and/or communication is filtered instantaneous. Since filtering software does not store personal data to block access to certain online resources, privacy is not at stake. However, blocking communication or preventing access to websites may affect freedom of information and expression.

12.4.3 Monitoring

Monitoring software informs parents about their children's online activities by recording websites addresses and online communication (for example instant messaging). Most parental control software allows monitoring online activities. A recent study found the use filtering and/or monitoring software does not correlate with less cyberbullying victimization.[48]

Monitoring software is preventive and works based on the assumption that users will adapt their behavior if they know their online activities are being watched. Because all online activity is stored, monitoring software theoretically targets all types, media, and methods of cyberbullying. In practice, however, cyberbullying incidents will have to be extracted by hand or automatically (see Sect. 12.4.1). Since cyberbullying might be hard to recognize and cyberbullying may only be a small part of all online activity, this is a tedious job. Because all online activities are recorded, peer communication is taken into account. Monitoring software does not rely on voluntary participation, users usually do not know or notice being monitored. Activities are recorded in real-time, however, action can be taken only after the records have been reviewed by an external party (for example a parent). For monitoring, privacy is an issue, because all online activities, which can be considered personal data, are recorded and stored for reviewing. Freedom of expression is not at stake.

12.4.4 Blocking Undesirable Contacts

Most instant messaging applications (e.g., Windows Live Messenger[49]), chat rooms, and social networking sites (e.g., Facebook[50] and MySpace[51]) give users

[48] Mesch 2009.

[49] See http://explore.live.com/windows-live-messenger.

[50] See www.facebook.com/.

[51] See www.myspace.com/.

the possibility to block other users, in order to prevent them from being contacted by these people. Many social networking sites also provide the possibility to restrict unknown users from contacting them and accessing their profile.

Blocking happens in response to incidents and limit harmful contact between minors and both minors and adults. Blocking contacts is suitable only for communication-based cyberbullying in applications where blocking options are available. It does take into account contact between peers. In fact, blocking bullies is a common advice for stopping cyberbullying.[52] Blocking is a voluntary act that allows users to control who can contact them. Users can block contacts whenever they want; in that sense blocking is real-time. Blocking users does not invade privacy or restrict freedom of expression.

12.4.5 Reporting Content

Many social Web applications (e.g., Facebook and MySpace) provide the possibility to report inappropriate and illegal content, for instance, by clicking a button labeled 'report abuse'. Reports are sent to community moderators that manually review reported content and decide whether or not to remove it. Some social networking sites, chat rooms, online games, and forums also allow users to report others when they break the rules, for example, by cyberbullying. Moderators decide whether and how to punish offenders.

Reporting tools can be useful for limiting access to inappropriate material, including some forms of content-based cyberbullying (for instance happy slapping videos or fake profiles on social networking sites). Reporting communication-based cyberbullying is only possible if moderators are available in the application and communication records exist. Everybody can report content they feel is inappropriate, so this technology relies on voluntary use. Because moderators have to check reports manually, it may take some time before reported content is removed. Therefore, reporting is not real-time. Privacy is not at risk, since no personal data needs to be stored for reporting (reporters may be anonymous). Removing content might interfere with freedom of expression. Therefore, in the case of cyberbullying, content will only be removed in obvious and/or extreme cases of content-based cyberbullying.

12.4.6 Age/Identity Verification

Age and/or identity verification technologies aim at restricting inappropriate contact between minors and adults as well as preventing minors to access

[52] See for example http://cybermentors.org.uk, www.stopcyberbullying.org and www.cybersmart.gov.au/.

inappropriate content. For example, in Second Life,[53] users must be 18 years old to view mature content. These technologies are preventive. Age and/or identity verification often use public or private databases containing information on either minors (for example school records) or adults (such as known sex offenders). People in the database (for instance minors or sex offenders) or people of certain ages (for example adults) are either allowed or not allowed to contact certain other people (such as minors) or access certain material (for example pornography).

Age and/or identity verification technologies do not target various forms of cyberbullying, such as content-based cyberbullying and harmful contact between peers. Age and/or identity verification may rely on voluntary participation (for example by becoming a member of a social network that applies age restrictions). In other cases participation may not be voluntary, for example if a school only allows its pupils to use the school's social networking website. This technology works online. However, since verifying age or identity requires gathering and storing personal data, the privacy of users might be at risk. Freedom of speech is not threatened.

12.4.7 Educational Technology

Education is another approach to improving the online safety of minors. Since the topic of this chapter is technology, the discussion below is limited to educational technology, such as interactive computer games.

FearNot! is an Intelligent Virtual Environment (IVE) in 3D, where synthetic characters act out bullying scenarios.[54] The application was designed for children 8–12 to witness the events from a third-person perspective. After a bullying episode, the victimized character turns to the user to ask for advice. The IVE offers children a safe environment that supports social and emotional learning. A controlled trial conducted in Germany and the UK established a short-term effect of escaping victimization for a priory identified victims of bullying and a short-term overall prevention effect for UK children,[55] demonstrating the potential of IVEs to support anti-bullying activities.

Other applications aimed at educating minors about online safety include Mr Ctrl[56] (not available anymore) and Internet Safety with Professor Garfield.[57] Mr Ctrl was a chatbot that answers questions about online safety. Internet Safety with Professor Garfield is a series of online interactive lessons about different topics concerning Internet safety. This type of applications can be used individually, but

[53] See http://secondlife.com/.

[54] Paiva et al. 2005.

[55] Sapouna et al. 2010.

[56] See http://mrctrl.spaces.live.com/ (in Dutch).

[57] See www.infinitelearninglab.org/.

also provide teaching material for classroom use. To the best of our knowledge, these applications have not been evaluated.

Because education is aimed at stimulating the right behavior in general, it basically targets all types, media and methods of cyberbullying. From the examples given here, it is not clear to what extend peer communication is explicitly taken into account. However, it would be easy to do so. Educational programs are usually mandatory, so there is no voluntary participation. Educational technology is designed to support traditional classroom teaching and not to protect or empower pupils at the same time they use the Internet. Finally, privacy and freedom of expression are not at risk in normal educational settings.

Another concern regarding education and/or educational technology is its effectiveness. Mishna et al. performed a systematic review of interventions against cyber abuse of youth.[58] Three educational programs were selected for review. Mishna et al. concluded that participation in cyber abuse prevention and intervention strategies is associated with an increase in Internet safety knowledge, but changes to Internet risk attitudes and behavior are not significant.[59] So, increased knowledge about safe Internet use does not necessarily correlate with less risk taking (or other behavior changes) online.

12.4.8 Summary

In this section, we discussed the expected effectiveness of different existing Internet safety technologies against cyberbullying. The results of this discussion are summarized in Table 12.2. While all technologies satisfy at least some of the desired characteristics from the framework we proposed, we expect their effectiveness against cyberbullying to be limited. Technologies such as age/identity verification, filtering and monitoring, reporting, and blocking undesirable contacts have not been designed to protect against cyberbullying, but with other online risks in mind. Some of these technologies primarily target access to undesirable content. Their success in protecting against cyberbullying, which is mostly communication-based, is therefore limited. According to our criteria, blocking undesirable contacts is the most promising approach.

One of the most salient features of existing Internet safety technology is its attempt to steer the behavior of users by restricting them. While in certain cases restricting bullies and/or victims might be useful, teaching them to deal with cyberbullying incidents seems a better approach. This viewpoint is supported by the literature. For example, Shariff argues that incidents of cyberbullying potentially are valuable learning experiences.[60] This potential, however, is ignored by

[58] Mishna et al. 2010.

[59] Mishna et al. 2010.

[60] Shariff 2008.

Table 12.2 Match between characteristics of existing technologies and the desired characteristics of technology against cyberbullying

	Different forms	Peer communication	Voluntary use	Real-time	Protect privacy	Protect freedom of expression
Content and behavior analysis	±	+	−	+	±	+
Filtering	±	+	−	+	+	−
Monitoring	±	+	−	−	−	+
Blocking contacts	−	+	+	+	+	+
Reporting	−	+	+	−	+	−
Age/identity verification	−	−	±	+	−	+
Educational technology	+	?	−	−	+	+

+: good match; ±: partial match; −: no match; ?: unknown

existing technologies. Additionally, Thierer claims education (media literacy) is the primary solution against online risks.[61] In his view, the role of technology is to supplement (but never to supplant) education. Educational technology (Sect. 12.4.7) is a primary example of using technology to supplement education.

Our discussion was focused on the separate existing Internet safety technologies. One might argue that combining multiple technologies, as is done in existing parental control software, might increase performance compared to individual technologies. However, the main issues, i.e., using technology that has been designed for other risks and restricting users instead of empowering them, will not be tackled by combining restrictive technologies.

Finally, we would like to emphasize that our discussion is limited to the expected effectiveness of technologies against cyberbullying. We do not claim the technologies discussed in this paragraph should not be used; they might be very effective against other online risks, such as exposure to problematic and illegal content or identity theft. However, based on the characteristics proposed in the framework, we expect that the effectiveness of existing Internet safety technology against cyberbullying is limited.

12.5 Conclusion

This chapter makes two contributions. First, we presented a framework of desired characteristics of technology against cyberbullying based on a review of literature on Internet safety technology and cyberbullying. Second, we discussed the

[61] Thierer 2009.

expected effectiveness of existing Internet safety technologies based on this framework. The results indicate that these technologies are not effective against cyberbullying, mainly because they restrict online behavior that is not related to cyberbullying.

The framework was constructed based on literature on Internet safety technology and cyberbullying. The framework consists of desired characteristics that were formulated by taking the perspective of a potential cyberbullying victim and making sure his/her online experience is not restricted too much. The following desired characteristics for technology against cyberbullying emerged: technology should be suitable for different types, media and methods of cyberbullying (at least communication-based cyberbullying), it should take into account peer contact, it should rely on voluntary use, it should be real-time, and user's privacy and freedom of expression should be balanced against restriction. These characteristics are not a definitive list; rather they should be seen as a first contribution in an ongoing discussion of technology against cyberbullying.

Our review of existing Internet safety technologies shows that all of them satisfy at least some of the characteristics from our framework. However, we conclude that the effectiveness of these technologies against cyberbullying is still limited. Technologies such as age/identity verification, filtering and monitoring, reporting, and blocking undesirable contacts have not been designed to protect against cyberbullying, but with other online risks in mind. Some of these technologies primarily target access to undesirable content. Their success in protecting against cyberbullying, which is mostly communication-based, is therefore limited. Blocking undesirable contacts is the most promising approach. Additionally, apart from education, none of the technologies discussed are designed to empower children and adolescents. Rather, the technologies restrict the behavior of bullies and/or victims (filtering and monitoring, age/identity verification, blocking undesirable contacts). While in some cases restricting the behavior of bullies and/or victims might be useful, incidents of cyberbullying potentially can be valuable learning experiences,[62] which are currently ignored by technology.

The results of our review of existing technologies indicate that prevention and detection of cyberbullying do not suffice. Five online safety task forces agree and conclude that empowerment, i.e., education and awareness, is a primary solution strategy to protect children and adolescents against online risks.[63] Technology can be used to supplement education and awareness. However, it is important to emphasize that technology alone can never solve a complex problem such as cyberbullying. A combination of social, legal, and technological measures is required for best results.

Technology does not have to be restrictive to influence behavior. Persuasive technology steers behavior by exerting social influence. In previous work, we presented a design for a virtual empathic buddy that provides emotional support

[62] Shariff 2008.

[63] Thierer 2009.

and practical advice to children that are victims of cyberbullying.[64] The buddy is a virtual character that 'lives' on the screen of potential cyberbullying victims. At the user's request, it provides emotional support and practical advice on how to deal with the incident.[65] A preliminary study suggests adolescents recognize the emotional cues emitted by the buddy.[66] Further research is required to assess the effectiveness of this kind of technology.

Acknowledgments This work is funded by NWO under the Responsible Innovation (RI) program via the project 'Empowering and Protecting Children and Adolescents Against Cyberbullying'.

References

Chawla NV, Japkowicz N, Kotcz A (2004) Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor Newsl 6(1):1–6

Dehue F, Bolman C, Völlink T (2008) Cyberbullying: youngsters' experiences and parental perception. Cyberpsychol Behav 11(2):217–223

Eurobarometer (2007) Safer internet for children, qualitative study in 29 European countries, national analysis, The Netherlands. http://ec.europa.eu/information_society/activities/sip/surveys/qualitative/index_en.htm

Finkelhor D, Mitchell KJ, Wolak J (2000) Online victimization: a report on the nation's youth. www.unh.edu/ccrc/pdf/Victimization_Online_Survey.pdf

Fogg B J (2002) Persuasive technology: using computers to change what we think and do, chapter computers as persuasive social actors. ACM, New York

Hinduja S, Patchin JW (2009) Bullying beyond the schoolyard: preventing and responding to cyberbullying. Corwin Press, Thousand Oaks

Hunter CD (2000) Internet filter effectiveness (student paper panel): testing over and underinclusive blocking decisions of four popular filters. In: CFP'00: Proceedings of the tenth conference on computers, freedom and privacy, pp 287–294, ACM

Internet Safety Technical Task Force (2008) Enhancing child safety and online technologies: final report of the internet safety technical task force to the multi-state working group on social networking of state attorneys general of the United States. Technical report. http://cyber.law.harvard.edu/pubrelease/isttf/

Kontostathis A, Edwards L, Leatherman A (2009) Chatcoder: toward the tracking and categorization of Internet predators. In: Proceedings of the 7th text mining workshop

Kowalski RM, Limber SP (2007) Electronic bullying among middle school students. J Adolesc Health 41(6, Supplement 1):22–30

Kowalski RM, Limber SP, Agatston PW (2008) Cyber bullying: bullying in the digital age. Wiley Blackwell, Malden

Lessig L (2000) Code and other laws of cyberspace. Basic Books, New York

Lessig L (2006) Code: and other laws of cyberspace, version 2.0. Basic Books, New York

[64] Van der Zwaan et al. 2010.

[65] We would like to emphasize that the buddy is not intended as a replacement for professional help or human support. Instead, the buddy should be seen as an additional, easily accessible support channel for cyberbullying victims.

[66] Van der Zwaan et al. 2012.

Lewis DD, Yang Y, Rose TG, Li F (2004) RCV1: a new benchmark collection for text categorization research. J Mach Learn Res 5:361–397
Li Q (2007) New bottle but old wine: a research of cyberbullying in schools. Comput Hum Behav 23(4):1777–1791
Livingstone S, Haddon L, Görzig A, Ólafsson K (2010) Risks and safety on the internet: the perspective of European children. Initial findings. http://eprints.lse.ac.uk/33731/
Mesch GS (2009) Parental mediation, online activities, and cyberbullying. Cyberpsychol Behav 12(4):387–393
Mishna F, Saini M, Solomon S (2009) Ongoing and online: children and youth's perceptions of cyber bullying. Child Youth Serv Rev 31(12):1222–1228
Mishna F, Cook C, Saini M, Wu MJ, MacFadden R (2010) Interventions to prevent and reduce cyber abuse of youth: a systematic review. Res Soc Work Pract 21(1):5–14
Olweus D (1999) The nature of school bullying: a cross-national perspective, chapter Sweden. Routledge, London, pp 7–27
Paiva A, Dias J, Sobral D, Aylett R, Woods S, Hall L, Zoll C (2005) Learning by feeling: evoking empathy with synthetic characters. Appl Artif Intell Int J 19(3):235–266
Patchin JW, Hinduja S (2006) Bullies move beyond the schoolyard: a preliminary look at cyberbullying. Youth Violence Juv Justice 4(2):148–169
Pendar N (2007) Toward spotting the pedophile telling victim from predator in text chats. In: ICSC'07: Proceedings of the international conference on semantic computing. IEEE Computer Society, pp 235–241
Sapouna M, Wolke D, Vannini N, Watson S, Woods S, Schneider W, Enz S, Hall L, Paiva A, Andre E, Dautenhahn K, Aylett R (2010) Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial. J Child Psychol Psychiatry 51(1):104–112
Shariff S (2008) Cyber-bullying: issues and solutions for the school, the classroom and the home. Routledge, London
Smith PK, Mahdavi J, Carvalho M, Fisher S, Russell S, Tippett N (2008) Cyberbullying: its nature and impact in secondary school pupils. J Child Psychol Psychiatry 49(4):376–385
Szwajcer E, Ebbers W, Oostdijk M, Wartena C, Hulsebosch B (2009) Kinderen en nieuwe media—technische and socio-technische oplossingsmogelijkheden voor gevaren in de online wereld. www.novay.nl/medialibrary/documenten/originelen/Eindrapportage_kinderen_en_nieuwe_media.pdf
The Gallup Organisation (2008) Towards a safer use of the Internet for children in the EU—a parents' perspective. http://ec.europa.eu/public_opinion/flash/fl_248_en.pdf
Thierer AD (2009) Five online safety task forces agree: education, empowerment & self-regulation are the answer. Progress & freedom foundation progress on point paper, vol 16, issue no. 13
Tokunaga RS (2010) Following you home from school: a critical review and synthesis of research on cyberbullying victimization. Comput Hum Behav 26(3):277–287
Van den Hoven J, Koops B-J, Romijn H, Swierstra T, Doorn N (2013, forthcoming) Responsible innovation volume 1: innovative solutions for global issues. Springer, Dordrecht
Van der Zwaan JM, Dignum V, Jonker CM (2010) Simulating peer support for victims of cyberbullying. In: Proceedings of the 22st Benelux conference on artificial intelligence (BNAIC 2010)
Van der Zwaan JM, Geraerts E, Dignum V, Jonker CM (2012) User validation of an empathic virtual buddy against cyberbullying. Stud Health Technol Inform 181:243–247
Vandebosch H, Cleemput KV (2008) Defining cyberbullying: a qualitative research into the perceptions of youngsters. Cyberpsychol Behav 11(4):499–503
Wolak J, Mitchell KJ, Finkelhor D (2006) Online victimization of youth: five years later. www.unh.edu/ccrc/pdf/CV138.pdf
Ybarra ML, Mitchell KJ (2004) Youth engaging in online harassment: associations with caregiver-child relationships, Internet use, and personal characteristics. J Adolesc 27(3):319–336

Ybarra ML, Mitchell KJ, Wolak J, Finkelhor D (2006) Examining characteristics and associated distress related to internet harassment: findings from the second youth internet safety survey. Pediatrics 118(4):1169–1177

Ybarra ML, Diener-West M, Leaf PJ (2007) Examining the overlap in internet harassment and school bullying: implications for school intervention. J Adolesc Health 41(6, Supplement 1):42–50

Yin D, Xue Z, Hong L, Davison BD, Kontostathis A, Edwards L (2009) Detection of harassment on web 2.0. In: CAW 2.0'09: Proceedings of the 1st content analysis in web 2.0 workshop