



Harmonizing freedom and protection: Adolescents' voices on automatic monitoring of social networking sites



Kathleen Van Royen*, Karolien Poels, Heidi Vandebosch

Department of Communication Studies, University of Antwerp, Belgium

ARTICLE INFO

Article history:

Received 13 September 2015
Received in revised form 24 February 2016
Accepted 25 February 2016
Available online 27 February 2016

Keywords:

Social networking sites
Adolescents
Automatic monitoring
Online protection
Qualitative research

ABSTRACT

Automatic monitoring of user-generated content on social networking sites (SNSs) aims at detecting potential harm for adolescents by means of text and image mining techniques and subsequent actions by the providers (e.g. blocking users, legal action). Evidently, current research is primarily focused on its technological development. However, involving adolescents' voices regarding the desirability of this monitoring is important; particularly because automatic monitoring might invade adolescents' privacy and freedom, and consequently evoke reactance. In this study, fourteen focus groups were conducted with adolescents ($N = 66$) between 12 and 18 years old. The goal was to obtain insights into adolescents' opinions on desirability and priorities for automatically detecting harmful content on SNSs. Opinions reflect the contention between a need for protection online versus the preservation of freedom. Most adolescents in this study are in favour of automatic monitoring for situations they perceive as uncontrollable or that they cannot solve themselves. Clear priorities for detection must be set in order to ensure the privacy and autonomy of adolescents. Moreover, monitoring actions aiming at the prevention of harm are required.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Adolescents spend a substantial amount of time on social networking sites (SNSs) (Lenhart et al., 2011). On these sites, adolescents are exposed to risks such as cyberbullying and sexual risks which can result in harm (Lenhart et al. 2011; Ybarra & Mitchell, 2008). Automatic monitoring of user-generated content on SNSs aims at detecting potential harm for adolescents by means of text and image mining techniques and subsequent actions by the providers (e.g. blocking users, legal action). Research in this area has been focusing on its technological development (Cano, Fernandez, & Alani, 2014; Dadvar, Trieschnigg, Ordelman, & de Jong, 2013; Lightbody, Bond, Mulvenna, & Bi, 2014; Van Hee et al., 2015). However, involving adolescents' voices regarding the desirability of this monitoring is important, particularly because automatic monitoring might invade adolescents' privacy and freedom, and consequently evoke reactance (Brehm, 1966). Therefore, in this study we involve adolescents in order to obtain insights into opinions on desirability for automatically detecting harmful content on SNSs.

2. Theoretical background

2.1. Social networking sites: between opportunities and risks

SNSs provide adolescents with several opportunities to actively participate in public culture and to gather and connect with friends

(boyd, 2014). In addition, SNSs are important venues for adolescents to experiment with their identity and manage their privacy and intimacy (Livingstone & Brake, 2010). Throughout these activities, youth create new forms of expression and develop social norms in negotiation with their peers (Ito et al., 2008). Several of these created social norms involve the use practises ranging from gossip, flirting, arguing, joking, ostracising and name-calling (Marwick & boyd, 2011). New ways of communication, experimental forms of self-display and affordances of SNSs, may transform the ample opportunities such as identity performance, intimacy and sociability easily into risks (Livingstone, 2008). Adolescents engage in risk-taking behaviours such as the disclosure of personal information and experimental peer communication, which may exacerbate online risks (Livingstone & Brake, 2010). The EU Kids Online network classified online risks to children into three types: content risks (content on the web, in which the child is the recipient), contact risks (in which the child is a participant, e.g. being groomed, being bullied) and conduct risks (the child is the actor in creating risks e.g. bullying/harassing) (Hasebrink, Livingstone, Haddon, & Ólafsson, 2009). Commonly encountered risks for adolescents on SNSs include cyberbullying and sexual risks (including sexual solicitations and cyber grooming¹) (Lenhart et al. 2011; Ybarra & Mitchell, 2008). However, experiencing risk does not imply harm since harm occurs depending on certain risk factors and protective factors (Staksrud, Ólafsson, & Livingstone, 2013). This study focuses on adolescents' opinions regarding the protection against harm of cyberbullying and sexual risks on SNSs.

* Corresponding author at: Sint-Jacobstraat 2, 2000 Antwerp, Belgium.
E-mail address: kathleen.vanroyen@uantwerpen.be (K. Van Royen).

¹ Cyber grooming: online solicitation of minors by adults for sexual purposes.

2.2. The role of SNSs in providing safety on their platform

In addition to digital literacy and public policy initiatives, one way to protect children is through industry-oriented measures such as monitoring these risks by automatically reviewing user-generated content on SNSs. Industry has been allocated a role in minimizing harm for youngsters on their platforms (Coyne & Gountsidou, 2013). Moreover, the challenge for parents to monitor their children's increasingly mobile media use beyond their view (Haddon & Vincent, 2014) suggests a more prominent role for public policy and industry players in protecting children. Online safety for young users is an important action undertaken by the Digital Agenda of the European Commission and aims at fostering multi-stakeholders dialogue and the self-regulation of SNS providers active in Europe (European Commission, 2010). At the European level, the self-regulating 'Safer Social Networking Principles' were formulated, encouraging SNS providers to ensure young users' safety (EC, 2009). These principles serve as a guideline and each provider can apply these recommendations in function of the nature of their network. Principles indicate several options for SNS providers to ensure safety and include providing: clear educational messages and user policies to allow users to navigate their services safely; age-appropriate services (e.g. delete under-age users and promote parental controls); tools and technologies to assist youngsters in managing risks (e.g. privacy settings); easy-to-use reporting mechanisms to alarm content violating the Terms of Service; responses to illegal content or conduct; settings for and information on privacy; and assessments of the available services to identify potential risks (EC, 2009). At present, SNSs review their content to detect illegal or prohibited user-generated content, using human moderators or semi-automated forms of detection (Staksrud & Lobe, 2010). However, to keep track of the vast daily user-generated content, automatic detection systems are suggested (Pachenko, Beaufort, Naets, & Fairon, 2013; Qi & Han, 2007; Reynolds, Kontostathis, & Edwards, 2011). Such systems apply automatic text- and image-categorization techniques using machine learning (Delort, Arunasalam, & Paris, 2011). These are similar to existing effective applications such as spam filtering (Sebastiani, 2002) and can detect potentially harmful content to inform the human moderators of SNSs (who conduct an in-depth analysis of these cases). In addition to words and emoticons expressing insults, profanity and typical contextual words, machine learning models have the capacity to automatically determine gender, age and personality (Schwartz et al., 2013). After detection, SNS providers can further deal with the detected content by taking actions such as removing the content, blocking users, legal action, and offering support. Currently, various efforts are being taken to optimize the detection of content that might represent risks or (associated) mental health problems of adolescents, such as cyber grooming (Bogdanova, Rosso, & Solorio, 2012; McGhee et al., 2011; Peersman, 2012) and cyberbullying (Dadvar et al., 2013; Dinakar, Jones, Havasi, Lieberman, & Picard, 2012; Kontostathis, Edwards, & Leatherman, 2010; Ptaszynski et al., 2010; Van Hee et al., 2015). Since automatic monitoring can take many forms, opinions on desirability will differ depending on what is being detected, who is being monitored and what happens after detection.

In addition to tackling practical and legal difficulties for the role of industry in online protection (Coyne & Gountsidou, 2013) as well as fostering the technological feasibility of automatic monitoring, insights into opinions on the monitoring of SNSs are vital.

2.3. Children's digital rights

This study adopts a child's rights-based approach. The United Nations Convention on the Rights of the Child (UNCRC) (United Nations, 1989) conveys the child's right to be protected against abuse and neglect, urging in online contexts for protection against harm encounters (Third, Bellerose, Dawkins, Keltie, & Pihl, 2014). In particular, beyond protection rights, the UNCRC places equal emphasis on a child's right to participate and the right to freedom of expression (United Nations, 1989). This

enshrines the idea to leave children free in seeking, receiving and imparting information online (Livingstone & O'Neill, 2014) and thus, automatic monitoring should embrace this right. Moreover, the automatic monitoring of online interactions may invade youngsters' privacy (van der Zwaan, Dignum, Jonker, & van der Hof, 2014) and consequently, evoke reactance. In advertising, for instance, privacy invasions can lead to reactance (Brehm, 1966; Tucker, 2014). Nevertheless, privacy is vital to adolescents' lives for developmental goals, including the achievement of autonomy, development of identity, exploration of sexuality and creation of intimacy (Peter & Valkenburg, 2011). This need to respect the child's privacy is asserted by Article 16 of the UNCRC (United Nations, 1989). Respect for privacy and freedom of speech was also argued for by experts in the field of cyberbullying when questioned on the automatic detection of cyberbullying (Van Royen, Poels, Daelemans, & Vandebosch, 2014). Similarly, van der Zwaan et al.'s (2014) framework of desired characteristics for the effectiveness of technologies against cyberbullying (e.g. monitoring, filtering, or educational technology) embodies ethical aspects. Particularly in regard with automatic monitoring technologies, they note that the users' privacy and voluntary use might be at stake (van der Zwaan et al., 2014). Thus, given these ethical aspects, discussions whether to increase the protection of youth online must include children and adolescents' voices to assure their rights (Third et al., 2014). To date, adolescents' perceptions of the desirability of protective strategies to be applied by the industry remain an understudied area, with few exceptions (Byrne & Lee, 2011). Thus far, adolescents have been consulted occasionally regarding their digital rights (Nordic Youth Forum, 2012; Third et al., 2014), parental mediation (Livingstone & Bober, 2003; Media Awareness Network, 2004) or their perceived usefulness of mediation sources (Tomkova, 2012).

In addition to insights into the desirability of automatic monitoring, it is important to investigate what adolescents consider harmful content on SNSs. Not all risks encountered in the online environment result in harm (Staksrud et al., 2013), and it is argued that eliminating all risks is neither feasible nor desirable (Livingstone & O'Neill, 2014). Moreover, focusing on the most harmful situations appears important—as Rooney (2010) suggests, surveillance technologies can negatively impact children by creating a risk-free environment that does not reflect the real world. The 'resilience to risk' approach argues that children can only develop resilience through exposure to risk or stress (Coleman & Hagell, 2007). Children need freedom to experiment and explore the online world, make mistakes and learn to cope, and eventually develop towards resilient individuals (Livingstone & O'Neill, 2014). Therefore, insights into the perceived severity of SNS risks can be useful in this regard.

To summarize, the objectives of this study are to solicit adolescents' views regarding the automatic monitoring of harmful conduct on SNSs and which content should (not) be monitored.

3. Methodology

3.1. Focus group protocol

Focus groups were conducted to obtain data on adolescents' perceptions on desirability of automatic monitoring and perceived harmful content on SNSs. In total, there were 14 focus groups with adolescents aged 12–18 years ($N = 66$) and the group sizes ranged from three to six individuals.

The adolescents were sampled from seven schools in Flanders (Belgium) by means of convenience sampling at the school level. First, the school principal was contacted by mail or telephone and briefed on the purpose of the study through a face-to-face meeting. After obtaining verbal consent from the school principal, adolescents with an active account on an SNS were voluntarily recruited from classes to participate. The researchers briefly explained the study's objectives in each classroom and the adolescents were asked to indicate whether they were willing to participate. For each class, maximum 6 pupils were recruited. Parents of the participants received an information letter, explaining to them the

Download English Version:

<https://daneshyari.com/en/article/345831>

Download Persian Version:

<https://daneshyari.com/article/345831>

[Daneshyari.com](https://daneshyari.com)