Curbing Harmful Online Communication – A Social Responsibility of Online Platform Operators

Sabine Einwiller, University of Vienna and Sora Kim, The Chinese University of Hong Kong
Problem situation

• Online sphere was envisioned to serve as an electronic forum where a plurality of voices engages in rational argument, thus fostering democratization (e.g., Rheingold 1995).

• Yet, this vision is severely hampered.

• Plenty of highly emotional and often aggressive, hateful and thereby harmful voices uttered and disseminated online.

• Dealing with *Harmful Online Communication* (HOC) and ways to deal with it is a pressing social issue.

• Various conflicting priorities, above all: freedom of speech vs. defense of human dignity and safety
Support for free speech

It is very important that without state/gov’t censorship in our country people can say what they want:

But, only 37% of US-Americans would oppose a law that would make it a crime to make public comments intended to stir up hatred against a group based on race, gender, religion, sexual orientation. (YouGov, May 2015)

Note: Global median of 38 countries (Russia and Ukraine not included in European median)
Source: PEW Research Center, Spring 2015, Global Attitudes Survey, Q56b
Harmful online communication (HOC)

• Ways of expression in online environments containing aggressive and destructive diction that violate social norms and aim at harming the dignity or safety of the attacked target, which can be a person, a social group or an organization.

• Legal regulation
  – Differs greatly between countries
  – When laws exist, they are often difficult to apply to online environment
  – Difficult to execute in global online environment
Legal regulation

• USA
  – Constitution, Amendment 1 - Freedom of Religion, Press, Expression
  – unprotected by the first amendment: Obscenity, child pornography, fighting words

• GERMANY
  – Laws on incitement of the masses, defamation, libel, slander
  – New legislation on online hate speech scheduled to be adopted in summer of 2017, but still heated debates; law shall enable Germany to fine social media companies up to 50 million euros ($55 million) for not reacting quickly enough to reports of illegal content or hate speech.

• SOUTH KOREA
  – Act on Promotion of Information and Communication Network Utilization and Information Protection (specific Internet law)
  – e.g., Article 61.1: Any person who has defamed any other person by clearly and openly alleging facts with a slanderous purpose through information and communications networks shall be punished ...

• CHINA
  – Constitution, Article 38: ... Insult, libel, false accusation or false incrimination directed against citizens by any means is prohibited.
  – General Principles Civil Law, Article 47: inciting national hatred or national discrimination, or humiliation in publications or computer information networks shall be detained ...
The social responsibility

• Countering harmful online communication can be considered a social and moral responsibility of organizations that operate online platforms with comments functions in order to protect the target(s).

• At the same time, organizations face the danger of being accused of censoring and prohibiting an open and multifaceted discussion.

  ➢ Dilemma of curbing HOC while allowing free speech
Research question

• How do organizations that provide and run online platforms address HOC and thus exercise their social responsibility to protect the targets of harmful communication while ensuring freedom of expression?
Methodology

• Phase 1
  Identifying and analyzing online comments policies by organizations that operate online platforms with comments options (qualitative and quantitative content analysis)

• Phase 2
  Interviews (qual.) with representatives of the organizations

• Research sample:
  – Online platform operators embedded in different national environments: China, South Korea, Germany, USA, Japan, Austria, UK
  – Online platform operators per country: web portal sites (3), blog hosting sites (3), News media sites (8), Social network sites (4), Community sites (3), Ecommerce sites (3), Corporations (10), Recommendation portals (3)
Sample: Number of policy documents

- Social Network Sites: 167
- Web Portal Sites: 94
- Corporation SNS/Websites: 82
- News Media Sites: 75
- Community Sites: 46
- Blog Hosting Sites: 40
- E-Commerce Sites: 26
- Recommendation Portals: 16

N_{total} = 546
Types of policy documents

Terms of Service/Conditions of Use
- SK = 49%
- CN = 23%
- US = 15%
- GE = 14%

Community Guideline/Netiquette
- SK = 33%
- GE = 27%
- US = 22%
- CN = 17%

Reporting Guideline
- SK = 68%
- CN = 16%
- GE = 9%
- US = 7%

Content Guideline
- SK = 32%
- US = 24%
- CN = 23%
- GE = 21%

In %  N_{total} = 546
Choice of language

Mostly instructing, informing
- US: 30
- GE: 47
- CN: 66
- SK: 225

Instructing and prohibitory
- US: 37
- GE: 28
- CN: 23
- SK: 28

Mostly prohibitory
- US: 20
- GE: 21
- CN: 19

\( \chi^2 (6, N = 544) = 123.97 \)
\( p < .0001 \)
\( V = .34 \)

N_{total} = 546
Reference to law(s)

\[ \chi^2(3, N = 546) = 53.64 \]
\[ p < .0001 \]
\[ V = .31 \]

SK: 51%

US: 11%

GE: 10%

CN: 28%

N_{total} = 546
# Organizations’ dealing with HOC

<table>
<thead>
<tr>
<th>Action</th>
<th>US</th>
<th>GE</th>
<th>CN</th>
<th>SK</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete without explanation/comment</td>
<td>71</td>
<td>68</td>
<td>69</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>Delete, close user's account</td>
<td>52</td>
<td>51</td>
<td>74</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Legal persecution</td>
<td>16</td>
<td>19</td>
<td>34</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Delete or close discussion</td>
<td>3</td>
<td>71</td>
<td>49</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Committee</td>
<td>2</td>
<td>24</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warning of user</td>
<td>10</td>
<td>15</td>
<td>27</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Delete, curtail with expl./comment</td>
<td>61</td>
<td>21</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User (member) management system</td>
<td>1</td>
<td>31</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Tracking system</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharing forum</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*N_{total} = 546*
Opportunities for user action
Mentioned in policy document or on website

- Mark or flag post: 230 (US: 56, GE: 78, CN: 92, SK: 76)
- Notify provider - telephone: 201 (US: 8, GE: 68, CN: 201, SK: 31)
- Notify authority or government: 167 (US: 1, GE: 70, CN: 98, SK: 31)
- Notify provider - postal mail: 176 (US: 4, GE: 176, CN: 4, SK: 4)
- Call for counter speech: 4

\[N_{\text{total}} = 546\]
Best practice – reporting & handling

- NAVER and DAUM, South Korean online portals and search engines (market shares Naver: 77%, Daum: 20%, Google: 2%)
- Provide information about what is considered to be harmful communication online, attempting to educate users
- Provide easy report buttons (or flag functions) so that users can request temporary removals of harmful posts; those being accused can appeal or request a formal objection through the reporting center
- Run reporting centers with dedicated review committees
- Very transparent about what they do to prevent obscene, harmful and illegal postings
NAVER Green Internet: What does NAVER do to prevent Obscene and illegal posting online?
What does NAVER do to prevent Harmful posts?
“What is considered harmful posts”

네이버는 어떤 노력을 하나요?
네이버는 모든 이용자가 안전하게 사용할 수 있는 인터넷 환경 조성을 위해
방송통신심의위원회의 인터넷 내용등급서비스를 참고하여, 유해한 게시물이 노출되지 않도록 노력하고 있습니다

유해게시물에 대한 대응
Report HOC

Report Copyright Infringement

Report Defamation Posts
The Procedure of Handling a Defamation Report  By Daum.com
Discussion and conclusions

• Many platform providers (esp. SNS and web portals) try to show their social responsibility by issuing policies on HOC
• However, execution is challenging not least because of different legal situations and the dilemma of curbing HOC while allowing free speech
• South Korea most advanced regarding legislation and reporting
• More needs to be done regarding the education of users also with respect to counter speech