

Monika Kaczmarek-Śliwińska

University of Technology in Koszalin

Managing crisis communication in the social media space in the context of hate speech

Zarządzanie komunikacją kryzysową w przestrzeni mediów społecznościowych w kontekście zjawiska hejtingu

ABSTRACT

Crisis management is often hampered by the occurrence online aggression. Environment of new media and social media is characterized by features which in not-supporting conditions can effectively threaten the crisis management. Hence essential an awareness of the phenomenon, but also potential effects and ways of reacting are being shown. It is worthwhile also having a fact that a presence and an activity are missing the person on the account or isn't guaranteeing the organization avoiding connected problems from online aggression.

KEYWORDS:

online aggression, crisis management, social media

ABSTRAKT

Zarządzanie sytuacją kryzysową często utrudnione jest przez zjawisko hejtingu, który staje się głównym lub dodatkowym aspektem negatywnych zjawisk w otoczeniu osoby lub organizacji. Środowisko nowych mediów i wyrosłych na ich gruncie mediów społecznościowych charakteryzują się cechami, które w warunkach niesprzyjających mogą skutecznie zagrażać zarządzaniu sytuacją kryzysową. Stąd istotną okazuje się świadomość zjawiska, ale także potencjalne skutki oraz sposoby reagowania. Warto też mieć na względzie fakt, że brak obecności i aktywności osoby lub organizacji nie gwarantuje uniknięcia problemów związanych z hejtingiem.

SŁOWA KLUCZOWE:

hejting, zarządzanie sytuacją kryzysową, media społecznościowe

“Hate speech” is defined as any of the various forms of online utterances or publications that are considered aggressive because of the form and/or a lack of reasonable argumentation. Most of them are texts (such as comments), but there are also graphics (e.g., famous memes). This phenomenon can refer to, or affect, anyone who communicates through the Internet, whether anonymous, private individuals, organizations, celebrities, famous institutions, or well-known brands.

While reading or viewing a post, it is often difficult to distinguish between hate speech and criticism. There are two main attributes that constitute the difference: the reasoning and the form.

Firstly, criticism should refer to facts and events, often documented or verifiable with citations of or links to sources of information. Secondly, criticism is supposed to be polite and non-infringing.

On the other hand, hate speech communicates unsubstantiated claims and can be overly emotional, rude, or even illicit (infringing on the rights of the addressee).

Now, from the point of view of the audience, these two types of utterances can have different effects. An attack on a public figure, institution, or organization can propagate to broad circles of the public and can cause anything from a minor problem to a major crisis.

An “ordinary problem” caused by hate speech for an organization, brand, or public figure can be considered a natural part of their being public, but a “crisis,” defined as a “time when the mission, vision, or goals of the entity or person affected becomes threatened,”¹ can put an end to their very existence or career. It is important to notice the point at which a problem turns into a crisis because this escalation calls for managerial measures that are crucial for protecting the future, the image, and the reputation of the entity or person concerned.

Therefore, distinguishing between hate speech and criticism and between its consequences (problem or crisis) is essential to managing the content of social media by those affected by them.

LEGAL ASPECTS

The hate speech we encounter online takes two main forms: text (comments) and graphics (memes).

Janina Fras distinguishes between the following four types of hate speech:²

1. Abuse – an utterance or behavior that offends another person
2. Insult – a negative opinion expressing emotions or values, intended to humiliate another person

¹ Kaczmarek-Śliwińska, “Public relations organizacji” 13.

² Dobek-Ostrowska, Fras, and Ociepka, 96.

3. Libel – a negative opinion about another person, containing a false claim that inspires moral contempt

4. Invective – an offensive allegation, abuse, or insult

This classification applies both to textual and graphic assaults.

Each of them can be recognized as an offence according to the Criminal Code or the Civil Code: an offense against dignity, an offense of defamation, or an offense against personal rights.

According to the Criminal Code:³

- Art. 212(1): Whoever slanders another person, a group of persons, an institution, a corporate entity, or an unincorporated organization for a behavior or attribute that can humiliate them in view of the public opinion or can put them at risk of losing trust that is essential for a job, profession, or activity is punishable by a fine or restriction of liberty.
- Art. 212(2): If the perpetrator of an act defined in Art. 212(1) committed the act using mass media, they are punishable by a fine, restriction of liberty, or a prison sentence of up to one year.
- Art. 213(2): None of the acts defined in Articles 212(1) and 212(2) is recognized as an offense if the publicly raised or broadcast allegation is true and the allegation:
 - involves a public officer or
 - serves to protect a legitimate public interest.

Where an allegation involves private or family life, the proof of truth may be sought only where the allegation was supposed to prevent a threat to the life or health of a person or to the depravation of a minor.

- Art. 214: The exclusion of incrimination under Art. 213 is not a waiver of the criminal liability of a perpetrator attracted by the form of raising or broadcasting the allegation.

Further, violations involving hate speech may be prosecuted under the Civil Code,⁴ Articles 23 and 24 of which refer to personal rights:

- Art. 23 names the following personal rights: “health, freedom, dignity, free conscience, name or nickname, image, privacy of correspondence,

³ Criminal Code.

⁴ Civil Code.

inviolability of home, and creativity (in the domains of research, art, inventions, and improvements).

- Art. 24 defines legal measures against the violation of these rights:
 - Art. 24(1): A person whose personal right has been, or may become, threatened by someone else's activity may demand discontinuation of the activity, unless it is not unlawful. If a violation has been committed, the person may demand the perpetrator to cause remedial of the consequences of the violation, particularly by making a statement adequate in terms of its content and form. The person may also demand pecuniary compensation for themselves or a contribution to a specific social purpose, pursuant to the relevant provisions of the Civil Code.
 - Art. 24(2): If the violation of the personal right has caused damage to property, the person harmed may demand a remedy on general principles.
 - Art. 24(3): The foregoing regulations are without prejudice to the rights provided for in other laws and regulations, including, without limitation, the Copyrights Law or the Inventions Law.

Going back to the Criminal Code:

- Art. 190 applies to offenses against freedoms and covers much more than the forms of hate speech described above:
 - Art. 190(1): Whoever threatens another person by expressing an intention to commit an offense to the detriment of that person, or their most closely related person, and the threat is perceived as reasonably credible, is punishable by a fine, restriction of liberty, or a prison sentence of up to 2 years..
 - Art. 190a(1): Whoever, by persistent harassment of another person or their most closely related person, gives them a sense of being threatened, justified by the circumstances, or significantly invades their privacy, is punishable by a prison sentence of up to 3 years.
 - Art. 190a(2): Whoever uses an image or the private data of another person and purports to be that person to cause damage to property of, or personal harm to, that person is liable to the same punishment.⁵

⁵ Criminal Code.

- Art. 216 refers to the offense of defamation:
 - Art. 216(1): Whoever insults another person, whether in the person's presence or absence, either publicly or with an intention to reach the person with the insult is punishable by a fine or restriction of liberty.
 - Art. 216(2): Whoever insults another person through mass media is punishable by a fine, restriction of liberty or a prison sentence of up to 1 year.⁶

As shown above, contrary to common opinions, there are legal instruments in place against hate speech. Thus, a person or organization harmed by it should know how to use these remedies; for instance, which law to choose (criminal or civil) or how to formulate claims.⁷ Unfortunately, most social campaigns speaking on the subject are confined to just defining and condemning hate speech, without mentioning the legal instruments and the consequences of using them.

HATE SPEECH-FACILITATING FEATURES OF THE NEW MEDIA

Among all media where hate speech is theoretically possible, it is the “new” or “social” ones, with their mass availability and the unprecedented freedom that give the problem a menacing dimension. Martin Lister and his colleagues pointed out three particularly important attributes of social media⁸: digitality, transformability, and interactivity of information. Going further, Adam N. Jonson described a phenomenon of disinhibition, Danah Boyd studied online audiences, and Michel Walrave and Wannes Heirman described aggressive behavior in terms of the “cockpit effect.”

The first of Lister's attributes, digitality, means that information exists in an intangible form; in contrast to its physical manifestations, it can be stored more conveniently and for longer, it is generally accessible, and it is cheaper to produce and modify or otherwise process.

⁶ Ibid.

⁷ For more see Głowacka 11.

⁸ Pyżalski 20–24.

Because it exists in a digital form, information is transformable, which is the second attribute of social media highlighted by Lister and his team. Particularly graphic content is susceptible to manipulation into hate speech because photographs and other images can be easily downloaded (often in violation of copyrights) and reworked or captioned with text.

And, finally, we have the third of Lister's attributes, interactivity, as a hate speech facilitator. Jan van Dijk examines four facets of interactive communication: multilateralism, the free choice of synchronicity or asynchronicity, the free choice of the role (sender/recipient), and the mental link with interaction partners (the understanding of their actions and contexts).⁹ Although these possibilities might seem to benefit both parties of a communication exchange evenly, van Dijk notes that "the balance of power tips in favor of the recipient" yet admits that "this potential is not used to the full in the contemporary digital media, so the sender's message continues to prevail."¹⁰

Two other, apparently conflicting, possible facilitators of hate speech are "dissociation" and "networking." The former means that the communicating parties do not meet face to face, instead they use electronic devices (PCs, smartphones, or tablets) to share content in cyberspace.¹¹ The latter consists in the building of a networked public,¹² centered around an online forum, such as a social networking service. In certain circumstances, this public can be an ideal environment for hate speech.

⁹ Ibidem 23.

¹⁰ van Dijk 23.

¹¹ Pyżalski 25.

¹² Kaczmarek-Śliwińska, "Public relations organizacji" 67.

Next, there is the disinhibition phenomenon described by A. N. Jonson. This situation occurs when “a user behaves in cyberspace as he or she would have never done in the ‘real reality’ because of the assumed or actual absence of social control mechanisms.”¹³ This sense of liberty, combined with a (mostly false) presupposition of anonymity and the “cockpit effect,” can make a perpetrator of hate speech feel safe and confident that his or her actions will go unpunished.¹⁴

The cockpit effect takes its name from the experience of bomber pilots.¹⁵ Just as the pilots could not see the vastness of the suffering and destruction they caused, someone who spreads hate speech cannot see the suffering of his or her victim.

HOW HATE SPEECH CAN CAUSE A SOCIAL MEDIA CRISIS

Certain relationships developing in online social media can lead to a crisis. Manuel Castells notes that the Internet is a low-cost environment for disseminating information or doing business, offering a lot of opportunities for sharing contacts and ideas within online communities, which would be difficult to achieve in the real world, even if ties established between the communicating parties are weak and can be broken at any time.

According to Edwin Bendyk, even these weak ties can bond online communities and provide people of different social statuses with access to information.¹⁶ In many cases such a social mix could not come to existence in the real world. The disadvantage for, or potentially a threat to, a person or organization who maintains a social platform is that this “social inclusion” can extend to ill-disposed or even hateful individuals. Accordingly, the prevention and management of crises triggered by hate speech should be strongly focused on screening the public.¹⁷

Potential perpetrators of hate speech can come from inside or outside an organization, so both proactive and reactive crisis management measures should include fact-finding about those who wish to join a social medium.

Krystyna Wójcik defines the environment of an organization in the context of image-building activities as “a public that already is, or may become, important for

¹³ Pyżalski 42.

¹⁴ Kaczmarek-Śliwińska, “Public relations organizacji” 67.

¹⁵ Pyżalski 138.

¹⁶ Szpunar 100.

¹⁷ Kaczmarek-Śliwińska, “Public relations organizacji” 67.

the organization because of the public's influence on the achievement of targets by the organization and because of the organization's reliance on contributions from, and resources of, the public."¹⁸ It is obvious that such an environment can attract individuals or groups pursuing incompatible goals or representing unacceptable attitudes, from the point of view of the organization, which can invite criticism or even hate speech.¹⁹ And the environment can evolve, not only because of the churn within the environment, but as a consequence of a change within the organization itself, creating a headache for those within the organization who are responsible for public communication and image protection. In fact, any instance of inspiring a member of the public, for better or worse, can trigger an attack on the organization. This is why the environment should be monitored to minimize such risks.²⁰

Again, this monitoring of the environment for the risk of hate speech can follow the classic, most natural, distinction between the "inner" and the "outer."²¹ The inner environment is involved in the pursuit of the organization's goals, typically formally (e.g., through an employment contract). It consists of people who work today and who worked in the past for the organization at any level, in any internal unit, as well as their relatives and friends. Hateful utterances made by them are often provoked by communication problems within the organization. Members of the outer circle have a different kind of relationship with the organization: it is mostly informal and based on either congruent or conflicting interests. The latter scenario can inspire criticism or, in the event of an escalation of bad feelings, stir hate speech.

SUGGESTION FOR A HATE SPEECH TYPOLOGY

An attempt to identify causes of hate speech and find countermeasures could benefit from a systematization of the phenomenon.²² Among many possible factors, the following three have been considered as an input to such a study: subject matter, focus, and authors.

¹⁸ Wójcik 64.

¹⁹ Kaczmarek-Śliwińska, "Public relations organizacji" 60.

²⁰ Ibid.

²¹ For more on the environments in crisis management see Kaczmarek-Śliwińska, "Public relations organizacji" 61–64.

²² Compare the proposed typology to the typology of victims and the nature of political content of social media in Kaczmarek-Śliwińska and Pyżalski (65).

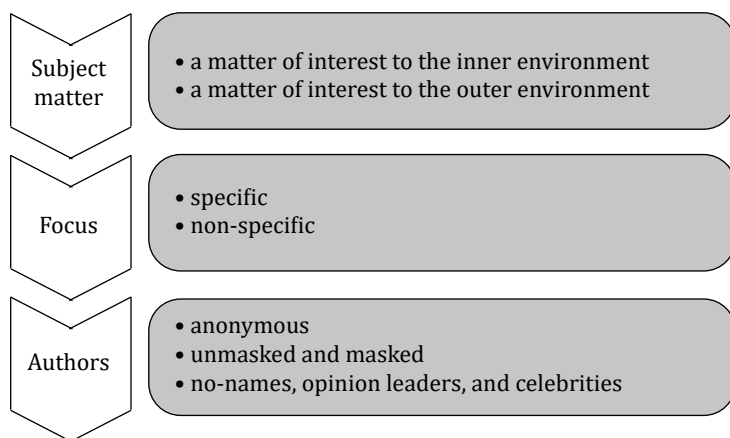


Fig. 1. Hate speech typology based on subjects, targets, and sources

The subject matter can depend on the mutual proximity of the communicating parties (the arms' length as a cut-off measure) and on the focus. Considering the former criterion, the subject matter can be

- interesting to the inner environment, which is mostly the case with less controversial matters that are important to a small group of stakeholders or
- interesting to the outer environment, typically engaging for a large community and having the potential to get through to "institutional media" and spark a crisis.

The focus can be placed on

- a single or specific matter or
- a number of matters, in which case the wider public may be concerned and the risk of a crisis may be higher.

Finally, from the point of view of hate speech authorship, we might be dealing with different identity profiles, such as

- anonymous authors (e.g., on a forum that does not require registration), typical for online mimicry and mimesis,²³

²³ Online communication mimicry is defined as the online activities of individuals with a concealed or false identity, who intend to inspire a behavior or conviction in the target audience. Online communication mimesis consists of building a non-authentic environment in online communications. These activities are used to pass unethical communication aimed at discrediting a brand, product, service, person, or organization (Kaczmarek-Śliwińska, "Public relations w przestrzeni" 63-66).

- people with either genuine or false identities (e.g., on a service that requires registration), or
- individuals unknown to the larger community, opinion leaders, and public figures (celebrities, politicians, journalists, etc.).

HOW TO MANAGE HATE SPEECH IN A CRISIS

Emergency planning for a crisis should consider the consequences of exposure to hateful utterances and both preventive and corrective measures.

Possible consequences include all kinds of problems that an utterance can cause to a person, brand, organization, or institution. While crisis managers do their best to silence the problem, the perpetrator works to the contrary. In an extremely bad scenario, the crisis can escalate and spread onto other business areas or reverberate loudly in institutional and/or social media.²⁴

An outbreak of hate speech should trigger preventive and corrective measures, which can throw a wrench in the works of the organization. The first response can be often handled by a dedicated unit (such as the PR team), but if the situation turns into a real crisis, the organization will need to establish a crisis center or appoint a dedicated team of experts.

However, no matter how well-qualified or skillful the problem-handlers may be, they will work under pressure and stress because their actions will be reactive, that is, unplanned for. Also, they will be distracted from their core activities, which can have a disorganizing or otherwise detrimental effect on the overall performance of the organization.

A hateful attack can harm the image of the organization. If we define an “image” as a totality of the public perception of the organization within the time continuum, each “hate point” can send the ratings down. If such hate events are sporadic, the damage can be made up for by an appropriate anti-crisis campaign. It is worse if such exposure is permanent or planned for (for instance by the competition): this can destabilize, harm, or even ruin an image.

Hate speech can affect not only the future, but also the present of the organization. If the organization takes measures that benefit its standing, but encounters internal and/or external opposition, the organization may revise its action plan and take a step back.

²⁴ Note the variety of attitudes of the recipients towards situations that can turn into a crisis and the causes and consequences of scandals. (See Kepplinger 12).

One more interesting thing worth mentioning about hate speech is the fact that an organization may face it even if it is absent in social media. Indeed, the level of activity of an organization in social media can be correlated to the magnitude of the risk, but inactivity or even absence is not a guarantee that there will be no hate speech; therefore, in either case, media should be monitored for hate events and their trends.

Regarding responses to hate, there are no proven patterns of action that could eliminate the phenomenon or even mitigate the reputational damage caused by it. Such activities, well-judged to maximize synergies and eliminate incoherencies, should be a part of the crisis management plan.

The following, for instance, can be done to face a crisis caused by or contributed to by hate speech:

- making no visible response for the inner and/or outer environment: working “in the background,” with no comments or statements to silence the attackers
- substituting the subject matter, raising a different topic as a shield
- demanding that the perpetrator (or social medium operator) remove specific comments or block the commenting feature (under Art. 21 of the Criminal Code)

One final thing worth considering is outsourcing a crisis management consultant to eliminate the risk of acting under emotions.

CONCLUSION

Apart from creating a number of great opportunities, the “new media,” including social media, have brought to the world new phenomena with the potential of steering a person or organization into a crisis.

Hate speech is impossible to eradicate. Safeguards – such as social media policies, education campaigns, or legal instruments – are often ineffective because social media is an environment full of chaos and, incidentally, the same chaos is the most noxious part of any crisis.²⁵ This is why the keywords include prevention, immunization against hate speech, media monitoring, early warning, and countermeasures.

²⁵ Kaczmarek-Śliwińska, “Public relations w przestrzeni” 129–139.

To cite the classic authors:

- Stanisław Lem: “Each new technology, without any exception, has the heads of benefits and the tails of new, hitherto unknown sores.”
- Henry Jenkins: “When people take media into their own hands, the results can be wonderfully creative; they can also be bad news for all involved.”

This is why we should learning about the new media and be consistent in identifying good communication practices.

REFERENCES

Civil Code of 23/04/1964.

Criminal Code of 06/06/1997.

Dobek-Ostrowska, B., Frasz, J., and Ociepka, B. *Teoria i praktyka propagandy*. Wrocław, 1999.

Głowacka, D. *Praktyczny przewodnik po art 212 kk*. Warsaw, 2012.

Kaczmarek-Śliwińska, M., and Pyżalski, J. “Media społeczne (social media) jako narzędzie realizacji agresji elektronicznej wobec polityków.” *Edukacja dla bezpieczeństwa. Cywilizacyjne problemy bezpieczeństwa*, edited by K. Augustyniak and A. Piotrowski, Poznań, 2011.

Kaczmarek-Śliwińska, M. *Public relations organizacji w zarządzaniu sytuacjami kryzysowymi organizacji. Sztuka komunikowania się*, Warsaw, 2015.

Kaczmarek-Śliwińska, M. *Public relations w przestrzeni mediów społecznościowych. Działania organizacji i jej pracowników*, Koszalin, 2013.

Kepplinger, H. M. *Mechanizmy skandalizacji w mediach*, Krakow, 2008.

Pyżalski, J. *Agresja elektroniczna i cyberbullying jako nowe ryzykowne zachowania młodzieży*, Krakow, 2012.

Szpunar, M. “Alienacja i samotność w sieci vs. grupowość i kapitał społeczny w Internecie.” *Oblicza Internetu. Architektura komunikacji sieci*, edited by M. Sokołowski, Elbląg, 2007.

Wójcik, K. *Public relations. Wiarygodny dialog z otoczeniem*, Warsaw, 2013.

van Dijk, J. *Społeczne aspekty nowych mediów*, Warsaw, 2010.

About the autor

Monika Kaczmarek-Śliwińska is a doctor of economic sciences; a communications and PR coach and consultant (image-building strategies, conventional and new media, and crisis management); the author of 3 monographs on online PR and more than 90 research and journalistic papers; and a coordinator and member of communication and new media research teams. Visit www.monikakaczmarek-sliwinska.pl, paczakutek.blox.pl, or monikakaczmareksliwinska.natemat.pl for more information.