Centre for
**Global
Cooperation
Research**

Briefing provided as a supplement to the 11th Käte Hamburger Dialogue 'Resisting Online Defamation: Prospects for Global Cooperation' on 11 December 2018 in Duisburg

Briefing

Online Defamation

Keywords: hate speech, criminal defamation, slander, libel, harrassment, insult, blasphemy, lèse-majesté

Topic and terms: an overview

Defamation is the communication of a false statement that harms the reputation of an individual, a company business, product, group, government, religion, or entire nation. Defamation is regulated in civil and/or criminal law, depending on national legislation. Often comments are racist or sexist and hence target certain people or groups. Online defamation is thus a generic term for the phenomenon of group-related misanthropy or sedition on the Internet and social media spaces.

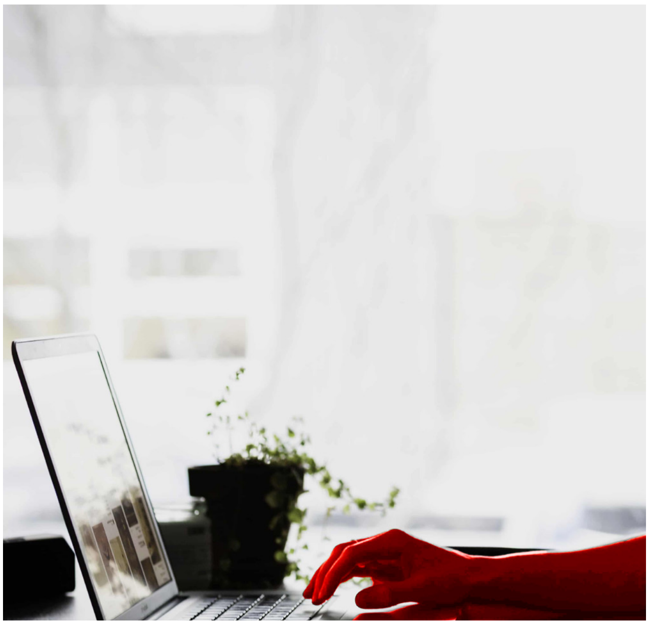
Marginalized groups are favoured objects of defamation. However, established groups or individuals can also become the target of defamatory campaigns.

Defamations are attributions by others. The actors themselves regard their statements as true or legitimate. This topic is therefore largely about the social understanding of what is perceived as defamation and what is to be fought against. The online 'pillory' functions as an extended form of the public sphere. This ranges from private peer groups in social networks to globally active influencers, NGOs and state actors.

They all benefit from:

- Real-time communication over any distance
- Anonymity of the (number of) actors
- cross-border ranges
- multi-agent systems
- Target group communication (targeting)

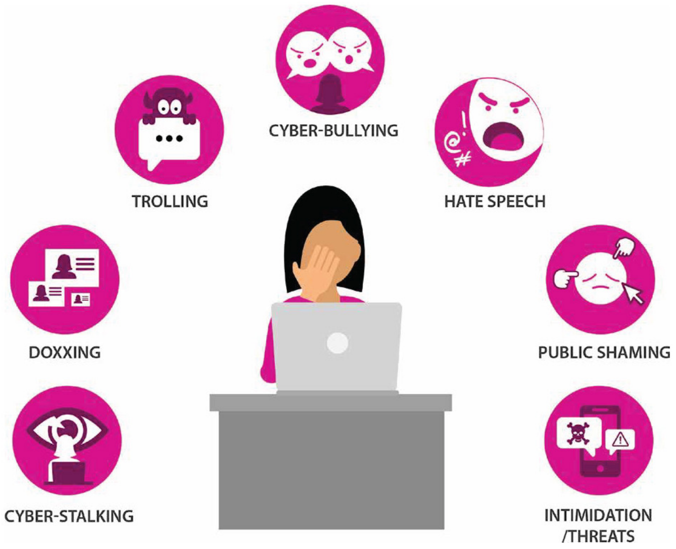
Online defamation is increasingly recognized as a social problem that official policymakers must address along with civil society actors. Racist-sexist hate campaigns launched by nationalist circles are currently the most obvious form of organized online defamation - not only in this country. As part of the forth estate, women journalists are to some extent doubly discriminated against on the basis of gender and occupation and are exposed to a circle of defamatory strategies (figure).



The Internet as a medium of globalization has become both the setting and vehicle for populist anti-globalist strategies. The observation of such tendencies in different countries within and outside Europe and the cross-border networking of relevant actions have put this topic on the agenda of global cooperation research.

Without claiming to be exhaustive, this briefing compiles aspects that may be relevant for public discussion.

Online harrassement of women journalists



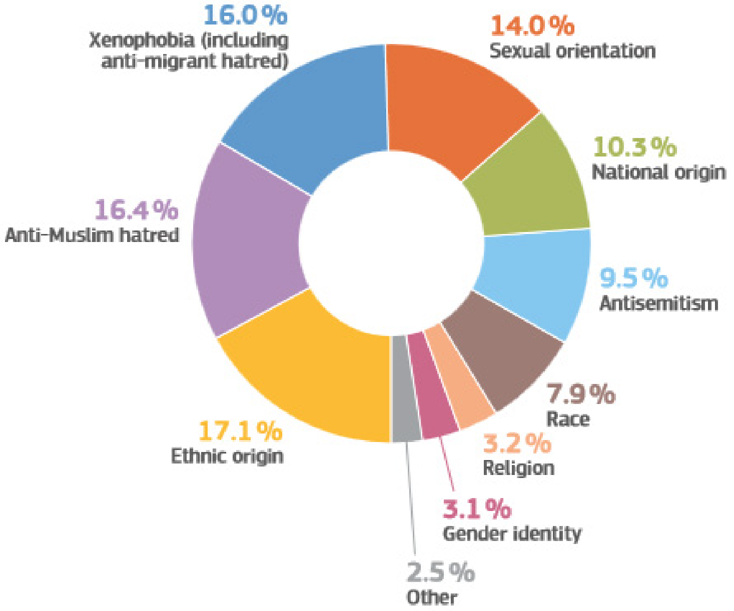
The challenge in Europe and Germany

EU: Code of conduct on countering illegal hate speech online

At the EU Internet Forum in May 2016, the EU Commission agreed a code of conduct with the leading IT companies Facebook, Microsoft, Twitter and YouTube to combat illegal hate speech on the Internet. On a voluntary basis, Internet companies should ensure that reports from users are processed within 24 hours, that users receive better support in reporting defamatory content on the Internet and that they are better informed about the rules in the online community.

- Instagram, Google+, Snapchat and Dailymotion also joined the agreement in 2018.
- The agreement will be evaluated on a regular basis. Results of the last evaluation from January 2018 include:
  - On average, 70% of the hate speeches reported (a total of 2,982 reports) were deleted by the IT companies.
  - 81% of the reports of hate speeches could be reviewed within 24 hours.
  - Reporting hate speeches was simplified and reporting procedures were made more transparent.
  - Cooperation with civil organisations to raise awareness of hate speech on the Internet was intensified.
  - Hate Speech mostly refers to racist hate speech against ethnic minorities, migrants* and refugees. The most frequent reasons for reporting hate speech were ethnic origin (17.1%), anti-Muslim hatred (16.4%) and xenophobia (16%).

Target Minorities: EU Code of Conduct third monitoring (2018): Notifications per ground of hate speech (in %)



Target: Youth

A recent survey, conducted prior to the elections in Hesse confirms findings from other studies: Young people between 18 to 24 years of age, are particularly affected by hate comments and threats in online networks. 42% report emotional stress and sexual harassment. Over 50% of the 18 to 44-year-olds–three age cohorts–report that they have been insulted once or several times. The threat of physical force experienced 40% of the younger ones at least once.

Target: Journalists

A study by the Institute for Interdisciplinary Research on Conflict and Violence entitled 'Perceptions of and Experiences with Attacks among Journalists' reports that a total of 42% of the interviewed journalists experienced hateful attacks from their audience in 2016. More than one fifth of the interviewees (26%) report that they were attacked from 'several times' to 'regularly'. The study also refers to the police crime statistics of the Federal Ministry of the Interior of the years of 2015, which show that criminal hate commentaries on the Internet have increased by 176% compared to the previous year. A survey of a total of 66 newspaper editorial offices revealed that 27 of them no longer publish certain content on Facebook and that more than half of them are overwhelmed with moderating online forums.

AfD: Study finds Social media can act as a propagation mechanism between online hate speech and real-life violent crime

The study by Müller/Schwarz scrutinized likes related to AfD's facebook account. Hate crimes increased in times of high anti-refugee coverage. The authors establish a similar correlation between Donald Trump's Tweets and the number of hate crimes targeting Muslims in US counties with high Twitter usage.

Debated: the German Netzwerkdurchsetzungsgesetz (NetzDG)

In order to force social networks to react to illegal content, the so-called 'Network Enforcement Act' (NetzDG) came into force on 1 January 2018. The law is controversial. Many critics fear that the operators of the networks would delete too much as a precaution due to the short 24-hour deadline, thus restricting the freedom of expression of users. If a contribution or an account is deleted without justification, those affected hardly have a chance to take action against the networks. The NetzDG does not provide for the possibility of appeals.

Satire or irony is often not recognized (danger of overblocking). In addition, the NetzDG places the authority to make decisions in the hands of the corporations.

Global Trends in Freedom of Expression and Media Development

Two types of arrangement between global digital platforms and national regulation can be observed:

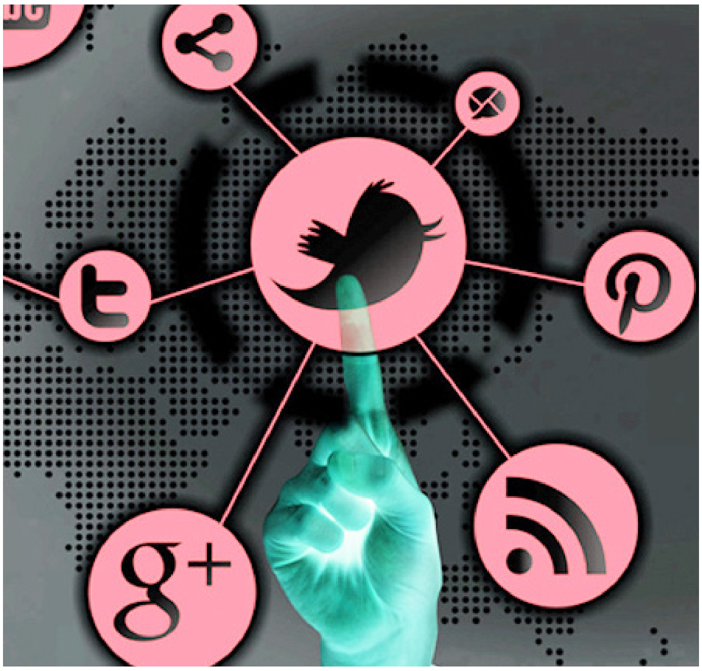
- Most parts of the world have allowed the growth of global enterprises of scale which transcend the regulatory state and for which modes of self- regulation in consultation with governments is the trend.
- In fewer places, there is a reassertion of sovereignty and an effort to 'domesticate' the platforms in line with tight controls on local legacy media. As part of this latter trend, internet providers are government-owned or controlled, or are in the hands of businesses close to the government, and data localisation is mandated.

Defamation challenge: two legacies

Many countries de-criminalized defamation as part of a trend to open up society and free speech. However, in recent years online developments gave way to an opposite trend. According to an UNESCO-report, countries in every region have started to increasingly criminalize instances of defamation by expanding their respective legislation to online content. Cybercrime and anti-terrorism laws have been passed throughout the world.

The independence of media is increasingly under pressure, due to complex interconnections between political power and regulatory authorities.

At the same time algorithmic cleansing works by industry proprietary solutions outside transparent governance arrangements. This may constitute a challenge for the legitimation of procedures agreed on in a global context.



Online defamation of Rohingyas in Myanmar

In March, a UN Fact-finding Mission on Myanmar announced that social media had 'substantively contributed to the level of acrimony' amongst the wider public, against Rohingya Muslims and that online hate speech is 'certainly, of course, a part of that'. Yanghee Lee, Special Rapporteur on the situation of human rights in Myanmar, specified: "We know that the ultra-nationalist Buddhists have their own Facebooks and are really inciting a lot of violence and a lot of hatred against the Rohingya or other ethnic minorities.' Reuters scrutinized the platforms' lasting difficulties to monitor their contents and their reluctance to properly tackle the issue.

Africa

In Africa at least four Member States decriminalized defamation between 2012 and 2017.

The African Court in Kanaté v. Burkina Faso observed that defamation laws are a ‚remnant of colonialism'.

'No one shall be found liable for true statements' (African Commission on Human and Peoples' Rights).

Fake, false, non-objective information

'General prohibitions on the dissemination of information based on vague and ambiguous ideas, including "false news" or "non-objective information", are incompatible with international standards for restrictions on freedom of expression … and should be abolished.'

(UN Special Rapporteur on Freedom of Opinion and Expression, OSCE, OAS and African Commission on Human and Peoples' Rights Special Rapporteurs for Freedom of Expression, 2017.)

An analysis of Arab defamation laws

Scrutinizing provisions in Egypt, Jordan, Kuwait, Lebanon, Libya and the United Arab Emirates, Duffy/Alkazemi explain that penalties for defamation usually are a fine or prison up to two years (all countries). But UAE cybercrime law provides for three to fifteen years in prison and ‚stiff fines' for publishing anything to deride the reputation of a long list of leaders.

True but defamatory

Many legislations see truth as a remedy against defamation accussations. However, in the Lebanese criminal law, article 583, people accused of slander are denied justifying themselves by showing the truthfulness of the slanderous information and disseminating it. The law was written to dissuade true, but defamatory statements (Duffy/Alkazemi).

Comments and the Chilling Effect - *the challenge*

Between 2013 and 2016, 20% of online publishing platforms either shut down comments or put them offline after even time consuming measures did not work (World Editors Forum). Numerous users have a similar experience on social media: the amount of toxic material feels unbearable, and people exposing themselves in defamation environments face emotional and physical threats.

The Guardian surveyed the 70 million comments recorded on its website between 1999 and 2016. Of these comments, approximately two per cent were blocked for abusive or disruptive behaviour.

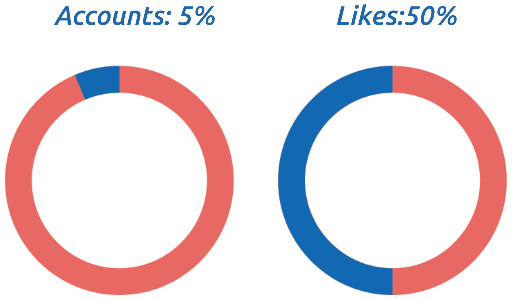
Notably, out of the 10 staff journalists who received the highest levels of abuse and 'dismissive trolling', eight were women, and two were black men. The chilling effect comes here: how can you expect from anyone to stand that?

But there is also success: especially in the field of problem analysis and communication strategies.

A quantitative analysis of likes for hateful speech-posts showed that 5% of accounts produced 50% of all likes in the dataset (diagram pictured). Numbers are not always, what they look like. A majority is suggested this way and legitimacy proclaimed (state actors included).

Volunteers engage in online discussions to uncover fake news or hateful content. #wirsindhier is a group acting in Germany.

Worldwide, content publishers are trying to promote quality comments from their users. According to a 2016 survey (World Editors Forum) a plurality (22%) embraced increased moderation, not to scare away those readers. Only 8% closed commenting completely.



Organizations and Platforms against Hate Speech

Hoaxmap - Neues aus der Gerüchteküche
German platform that tries to uncover false reports: https://hoaxmap.org. On an interactive map you can find numerous false reports, which were locally assigned and disproved.

No Hate Speech Movement Deutschland
https://no-hate-speech.de

Task Force "Umgang mit rechtswidrigen Hassbotschaften im Internet"
https://www.fair-im-netz.de

Canada
Canadian Anti-Hate Network
https://www.antihate.ca

United States

Southern Poverty Law Center: https://www.splcenter.org
Color of Change: https://colorofchange.org

Sources

European Commission: Countering illegal hate speech online http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300.

Siegert, S., 2016: Exklusive Journalist-Umfrage. Nahezu jede zweite Zeitungsredaktion schränkt Online Kommentare ein. In: journalist online, 01.03.2016.

Institut für Demokratie und Zivilgesellschaft (IDZ) im Auftrag von Campact (2018). #Hass im Netz: Der schleichende Angriff auf unsere Demokratie. Eine repräsentative Untersuchung in Hessen.

Müller, Karsten and Schwarz, Carlo, Fanning the Flames of Hate: Social Media and Hate Crime (May 21, 2018). http://dx.doi.org/10.2139/ssrn.3082972.

Müller, Karsten and Schwarz, Carlo, Making America Hate Again? Twitter and Hate Crime Under Trump (March 30, 2018). http://dx.doi.org/10.2139/ssrn.3149103.

Netzwerkdurchsetzungsgesetz (NetzDG) and Critique (in German): https://www.bpb.de/dialog/netzdebatte.

UNESCO. 2018. World Trends in Freedom of Expression and Media Development: 2017/2018 Global Report, Paris.

Matt J. Duffy & Mariam Alkazemi (2017). Arab Defamation Laws: A Comparative Analysis of Libel and Slander in the Middle East, Communication Law and Policy, 22:2, 189-211, DOI: 10.1080/10811680.2017.1290984.

Online Defamation of Rohingyas
https://www.bbc.com/news/technology-43385677;
https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/.

Philip Kreißel, Julia Ebner, Alexander Urban, Jakob Guhl (2018). Hass auf Knopfdruck. Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz. Hrsg. Institute for Strategic Dialogue (ISD) und #ichbinhier. https://www.isdglobal.org/wp-content/uploads/2018/07/ISD_Ich_Bin_Hier_2.pdf.

Code of Conduct on countering illegal hate speech online. Results of the 3rd monitoring exercise. January 2018.

World Editors Forum (2016). Do Comments Matter? Global Online Commenting Study 2016. Frankfurt.

Centre for
Global
Cooperation
Research

Käte Hamburger Kolleg / Centre for Global Cooperation Research (KHK/GCR21)
A Central Research Institute of the Unversity of Duisburg-Essen
Director: Sigrid Quack
Co-Directors: Tobias Debiel, Dirk Messner, Jan Aart Scholte
Schifferstraße 44, 47059 Duisburg
Phone: +49 (0)203 379 5230 Fax: +49 (0)203 379 5276
E-Mail: info@gcr21.uni-due.de

Briefing 11/2018 - GCR21 Public Relations (Editor: Martin Wolf. Recherche: Daniela Weißkopf, Eva Marie Trösser)