Hate speech on the net, a danger for society and democracy

By Sophie Achermann,

CEO of Alliance F and Co-Director of Stop Hate Speech

“Hate speech on the internet has become a global problem. The barrier to hostility and insults has fallen. Never before has it been possible to see so clearly how hatred on the internet can pose a threat to democracy as it was on 6 January 2021. After weeks of inflammatory attacks on the internet, the hatred spread from social media to the offline world, and an attack on the oldest democracy was witnessed in real time by all. However, this monstrous spectacle does not stand alone. For years, we have seen how hatred on the net has repeatedly translated into offline violence. In 2015, during the big wave of refugees, social media was filled with hate against refugees. In Germany, countless refugee homes were set on fire. Even misogyny from the keyboard has turned into attacks from the internet. The mood on the internet is heated. Many people are withdrawing for fear of hate and negativity.

Freedom of opinion and diversity of opinion

Freedom of expression is a fundamental right. Free speech and open debate are core values of a free society. The internet and especially social media have made it possible for people to express their opinions and be heard, regardless of their social or economic status. This is an achievement that must be preserved. But it is important to distinguish between freedom of expression and hatred. Hate is not capable of dialogue. Hate is not capable of discourse. Hate-filled content causes many people to withdraw from public discourse for fear of virtual violence. This restricts the freedom of expression of many people and damages the diversity of opinion. Especially in a country like Switzerland, where we discuss and have a say in

new laws and constitutional amendments every few weeks, hate speech is extremely dangerous and harmful. We need open, tough but fair discussions and everyone must play by the same rules.

Who determines the rules of the game?

Twitter, Facebook and Instagram have pulled the plug on the president of the United Nations. With harshness, they are trying to make up for the failures of recent years. Of course, social media is to blame. But blaming alone is not enough. Politicians have neglected to make democratic, legitimate discussions and decisions, and now private companies or Silicon Valley are deciding on freedom of opinion and diversity of opinion on the internet. And what does society do? After all, the hate on the internet is human-made. So solutions are needed to protect our democracy. Politicians have to deal intensively with the problem. But society also bears its responsibility.

The danger of hate speech for democratically legitimised social media

Research shows that women are sexualised and vulgarly assaulted much more often than men. They receive rape threats or are confronted with stereotypes: Women belong in the kitchen at home. This leads to women reacting differently to hate speech: They withdraw from online discussions and social media more often than men.

As a result, social networks are no longer open to a large part of the population. Women who are already underrepresented in the public sphere will withdraw even more. This is dangerous, because a democracy can only exist if all citizens are given the opportunity to participate in the discourse. This is why the largest women's umbrella organisation in Switzerland – alliance F – has set itself the goal of addressing this issue. With the aim that all people can participate in the internet without fear of violence. alliance F has launched the Stop Hate Speech project.

A Swiss project Stop Hate Speech tries to counter hate by combining technology and social engagement

In Switzerland, a new platform was launched at the beginning of January 2021. In the connection of society and technology, it tries to counter hate and with

the aim of improving discourse on the internet. This should contribute to a stable democracy.

Bot Dog: the little sniffer dog

The project has programmed an algorithm named Bot Dog. The little dog searches the internet for hate. As soon as it finds it, it brings it back to a community of volunteers. These in turn can reject the hate with counter speech.
So far, there is no universally accepted definition of hate speech. Hate is a subjective feeling. And it is always a challenge to build an algorithm on subjective assumptions. To get around this problem, the project has engaged civil society.

Over the past few months, around 1200 volunteers have trained the new "Bot Dog" algorithm from scratch. By assessing whether comments contain hate or not, "Bot Dog" has learned to detect hate speech on the internet itself. For this purpose, the community has evaluated more than 52,000 comments. Each comment was rated at least three times and the community was also weighted according to diversity criteria. Groups of people who were less frequently found in the community were given more weight in their evaluation. This was done with the aim of obtaining a possible cross-section of the population and their feelings towards hate speech.

This was an attempt to actively prevent a bias in the programming of Bot Dog. Bot Dog has a finder rate of 80% (in January 2021 – and will get better over time). Bot Dog tries not only to find hateful messages, but to erure them in the context of an article. Accordingly he rates articles based on their risk of receiving hate comments. This is to enable the earliest possible intervention in the discourse and thus contribute to the prevention of hate speech.

Counter Speech platform

With a counter-speech platform, the community can react to Bot Dogs prey, counteract hate and sustainably improve the culture of discussion on the net. Support in the form of various counter-speech strategies, a reference book on forms of discrimination and a help centre can also be found on the platform. In the coming months, the site will be expanded with the help of community work into a digital know-how centre that will contribute to sensitising society and empowering it to deal competently with hate speech.

Bot Dog (algorithm) and the community are constantly learning from each other and improving.

Scientific support

In the accompanying research project of ETH Zurich and the University of Zurich, various counter speech strategies are being researched for their effectiveness and impact, so that the phenomenon of hate speech can be combated as specifically as possible.
What is the best way to prevent hate speech? And which counter-speech strategies are particularly successful? So far, very little is known about which counter-speech strategies are most effective in which context and in response to which type of hate speech. In this regard, not only in Switzerland, but also our partner organizations and NGOs working against hate speech at home and abroad, are largely in the dark. With the help of the community the project is trying to change that.

The goal of the accompanying research is to ensure that “Stop Hate Speech” can have the greatest possible impact with its limited resources and that anti-hate speech organizations worldwide can provide directly applicable knowledge about which strategies work in which contexts.

Figures about hate on the net

Hate on the Net is not yet sufficiently well researched. Reliable data is lacking for Switzerland, which could also have an influence on finding a political solution. With the help of Bot Dog and the cooperation with universities, the project should also be able to improve the data situation in Switzerland. The Stop Hate Speech project wants to promote the culture of discussion with data, facts and active counter-speech and thus ultimately help to depolarise minds.

Geneva 21 January 2021