The Synopsis of Research Carried out by
the "NEVER AGAIN" Association within the OpCode project
in September 2019 – September 2020

Author: Jacek Dziegielewski

**"NEVER AGAIN"
ASSOCIATION**
neveragainassociation.org

A) Brief presentation of your organization role within OpCode project and your research background on hate speech phenomena

The "NEVER AGAIN" Association, founded in 1996, is a non-profit organization that focuses on monitoring hate speech and hate crimes, as well as conducting educational initiatives.

The Open Code for Hate-Free Communication (OpCode) project aims at countering online hate speech using a multidimensional approach consisting of monitoring, analysis, taking action, developing free and open software solutions for moderating user generated content, and by engaging in advocacy and network consolidation (International Network Against Cyber Hate). The project also aims to consolidate the efforts of civil society organizations to participate in European monitoring exercises and continuously monitor the Web, thus providing data as a basis for future policy. Furthermore, our goal is to provide national and European policy-makers as well as IT companies with recommendations on how to address the issue. The project will contribute to European Commission's recommendation on measures to effectively tackle illegal content online that highlights the necessity of cooperation and mutual exchange between tech companies and civil society organizations for the benefit of the public. In frames of the OpCode project in the period September 2019-2020, "NEVER AGAIN" participated in two Monitoring Exercises together with the other European project partners and in the preparation of the special report on online hatred in the context of the coronavirus pandemic.

B) Brief presentation of the national context in the last year (2019-2020) in relation with the hate speech phenomena

Recent years brought an outburst of hate speech and hate crimes. Often the hate is incited by radical political movements (although also mainstream political parties have played an important role in making the public debate more hateful and violent). In 2019 and 2020 Poland held elections to the Parliament and the presidential elections. Incitement to hate occurred in both of the political campaigns. The target of most hate speech by politicians and public figures were the members of the LGBT+ community. Politicians, journalists and celebrities used hatred against this community as a way to strengthen the radical electorate. Many examples of such hate are included in the current edition of "The Brown Book", the "NEVER AGAIN" Association's register of hate crimes and hate speech incidents. 2020 brought the COVID-19 pandemic and hatred towards people of Asian origin, which later spread into discrimination against all non-Poles (individuals or whole groups considered as "others") as well as an outburst of conspiracy theories, often incorporating antisemitic, homophobic and racist elements. Both the LGBT+-targeted hate and the coronavirus-related hate made its way into the streets, resulting in acts of violence and discrimination. It is a common pattern that the hate speech by public figures leads to violence in real life.

C) Main trends and relevant outcomes of the first two Code of Conduct Monitoring Exercises

Social media platforms in Poland are not responsive in a sufficient manner. A big part of the content we submitted for deletion during the Monitoring Exercises was not even analyzed (or we did not receive a sign that it was analyzed), and a vast majority of the content remained available online. Facebook seems to be the most responsive (but usually with negative responses), while Twitter and YouTube in general do not respond at all to our reports. Most of the cases we chose to report to the IT companies included antisemitism, racism, homophobia and incitement to violence.

During the latest Monitoring Exercise, "NEVER AGAIN" reported 58 hateful comments or materials. Most of them were reported to Facebook and Twitter. A very surprising outcome was the lack of reaction on the side of Twitter. We did not receive any information indicating assessment of any of the reported cases. Those tweets included severe homophobia, racism, antisemitism – also referring to the Black Lives Matter movement and the 2020 coronavirus crisis. During the previous Monitoring Exercises we received some feedback from Twitter. In some cases, the hateful content was removed. In others we received information that Twitter admits that certain tweet violates their rules and should be removed – but they did not remove the reported content and it is still accessible on Twitter. Some examples of hate speech that Twitter ignored are: "Hitler was LGBT", "Jews have one purpose - to kill and enslave non-Jews, they can murder, lie and cheat", "we need to destroy LGBT", "Jews killed more Poles than Hitler", "You jews [sic] will perish, you killed Jesus Christ [...] you should cut off your penises and uteruses". Facebook removed some of the reported content, for example comments: "F*ck the Jews", "The biggest evil of the whole world are jews [sic] - murderers of the Son of God", "Away with Jews", "to the bag with him" (comment calling for murder of an immigrant). The Facebook decided not to remove the following comments: "Maybe someone will throw this plague out of Poland" (about Jews), "I pray every day that God helps to cleanse Poland of those traitors" (also about Jews). Facebook did not indicate any assessment and did not remove comment "Hitler is missing. He would gas and shoot all those monkeys" (about black people and immigrants). Facebook also did not block the selling of t-shirts expressing support for Polish racist killer Janusz Waluś, who murdered African anti-Apartheid activist and politician Chris Hani (as well as containing the "white lives matter" slogan). "NEVER AGAIN" also applied the methodology based on control group and test group. The test group consisted of a researcher who took part in previous Monitoring Exercises, therefore his IP address and personal information were known to the IT company. The control group consisted of a different alias of the researcher and used VPN, therefore the person reporting content from this account was "unknown" to the social media company. The results confirmed the thesis that social media companies (especially Facebook) react more eagerly to the reports coming from a person who has submitted reports earlier than from a completely "new" user.

D) Main challenges and limitations during the first two Code of Conduct Monitoring Exercises

The main challenge remains to convince the social media platforms to start responding to our reports. Without a proper response it is hard to combat hate speech effectively. Also the technology and procedures used to assess the reports is vague and unclear; we do not know why one report is assessed and the other is not; we do not know why some content is removed and the other is not. IT companies should design a clear system of reporting hateful and/or illegal content and establish a transparent and efficient way of communicating with users who are reporting this content. It is unacceptable that hateful content, whether covered by law or not, is not removed by the social media companies. We noticed that the IT companies responded mostly to more "severe" acts such as calling for murder, but mostly ignored "common" verbal homophobia, racism and antisemitism. In our opinion also statements glorifying antisemitism or racism pose a great danger (violence usually starts with violent words and statements). So, the people deciding which comment to remove and which to leave intact should react to all reports, not only the most "severe" ones. This applies mainly to Facebook, as Twitter does not respond to vast majority of reports, not removing or even assessing even tweets containing a clear call for violence (for example, a tweet containing a link to article about the Unabomber – Ted Kaczynski, terrorist who, in the opinion of the article's author, did a great job fighting the "lefties").

E) Main trends and critical aspects that you identified within the Corona Hate case study

The Corona Hate research showed that the wave of hatred was first targeted at people of Asian descent. Later, most probably caused by the national lockdown, the hatred "moved" and also broadened the target to all people of non-Polish descent. We also noticed a huge amount of conspiracy theories, often being spread by public figures such as politicians, celebrities, artists, religious leaders etc. Those theories often included antisemitic, racist and homophobic tropes. The rise of conspiracy theories is also highly dangerous as it poses a great risk to public health – as many of them include denial of pandemic or anti-vaccine content. The scheme observed during the pandemic-era hatred is similar to the previous outbursts of xenophobic, racist, antisemitic and homophobic behavior. Some politicians and political parties use the existing fears in society or create a new kind of fear to consolidate votes. In 2015, it was observed in the hateful narrative towards refugees (identifying all refugees / immigrants as terrorists etc.), then the narrative was followed by real-life discrimination and acts of violence. In 2019 and 2020 many politicians used homophobia as political fuel. This attitude created a sense of danger in the more vulnerable part of electorate, who became

scared of "LGBT ideology" or "gender ideology". It also involved fake news such as connecting homosexual orientation to pedophilia or manipulating the WHO directive to present sex education as, for example, "teaching little children how to masturbate". Politicians using hateful homophobic language achieved success in the elections. Similarly, public figures such as politicians, journalists, religious leaders (e.g. catholic priests) promoted conspiracy theories and non-scientific approach towards the topic of global pandemic. Such statements, repeated frequently, created a safe space for people who started using violence or discrimination against people being unfoundedly accused of spreading the virus. It shows a great responsibility of public figures.

F) Main challenges and limitations when documenting Corona Hate case study

First of all, a big problem is the general unwillingness of social media platforms to respond to reports and remove the hateful content. Another issue is the attitude of media – both traditional and social media – because of the fact that they either give platform to people spreading fake news and conspiracy theories, or not pay enough attention when it comes to removing such content that has already been posted on the Internet.

Another significant limitation, as pointed above, is also that the social media platforms do not provide a clear way of communication with users reporting hateful content. We cannot act effectively, if algorithms are vague and it is unclear on what basis it is decided which content is removed and which is not. Although several social media platforms developed strategies to combat fake news and disinformation (mainly Facebook and YouTube), still many videos and posts on those sites present conspiracy theories and information that is unproven and potentially harmful.