



הארגון הבינלאומי של עורכי-דין ומשפטנים יהודים (ע"ר)
THE INTERNATIONAL ASSOCIATION OF JEWISH LAWYERS AND JURISTS (R.A.)

IAJLJ Policy Paper Series 2019/1

Self-Monitoring of Antisemitism in Social Media

(July 2019)



About the IAJLJ:

The IAJLJ is an international non-governmental organisation committed to issues that are on the agenda of the Jewish people and works to combat racism, xenophobia, antisemitism, and Holocaust denial. The IAJLJ is an accredited ECOSOC Special Consultative NGO at the United Nations.

The IAJLJ was established in 1969 by three distinguished human rights leaders - the 1968 Nobel peace prize laureate Rene Cassin, Justice Haim Cohen from the Israeli Supreme Court, and Justice Arthur Goldberg from the US Supreme Court.

Date of Publication: July 24th, 2019

Written by: Ido Rosenzweig (ido@iajlj.org)

International Association of Jewish Lawyers and Jurists
10, Daniel Frisch st,
Tel-Aviv, 6473111
Israel
Tel: +972-3-6910673
Fax: +972-3-6953855
Email: office@iajlj.org
Website: www.intjewishlawyers.org



Table of Contents

1. Introduction	3
2. Online Antisemitism And Its Real-Life Consequences	3
3. The Definition of Antisemitism and New Forms of Antisemitism.....	5
4. Freedom of Expression and Limitations.....	6
5. The Existing Use of Self-Monitoring in Social Media.....	8
6. Conclusions and Recommendations.....	11



Self-Monitoring of Antisemitism in Social Media

1. Introduction

Along with the rise of extremism and antisemitism around the world, there is an increased surge in the number of online antisemitic incidents, and it is of vital importance to pay due attention to this deeply troubling phenomenon along with its real-life implication.

In this submission, the IAJLJ presents its point of view on the correlation between online antisemitic hate speech and real-life violence, and emphasizes the importance of self-regulation mechanisms by social media platforms. The IAJLJ also provides a list of recommendations to move forward in handling these problems.

While this policy paper focuses mainly on antisemitism, relevant elements and recommendations can be read in a wider context to address other forms of online hate speech, including xenophobia and Islamophobia.

2. Online Antisemitism And Its Real-Life Consequences

Recent reports by the Anti-Defamation League (ADL), FBI, OSCE-ODIHR, and the EU Agency for Fundamental Rights show a rise in antisemitic incidents in recent years. According to the ADL, there have been 1,986 antisemitic incidents in 2017 and 1,879 in 2018 in the United States of America (US) alone.¹ This constitutes a significant increase in comparison to previous years (1,267 in 2016, 942 in 2015 and 751 in 2013). Such increase is also evidenced by the annual report of the FBI which states that approximately 58 percent of the 1,679 religious bias hate crimes that were reported to law enforcement agencies in 2017 were anti-Jewish in nature.² The OSCE-ODIHR annual report shows a similar trend also in countries/regions other than the US with 2,139 antisemitic incidents reported in 2017.³ As reported by the Campaign Against

¹ www.adl.org/media/12857/download

² <https://ucr.fbi.gov/hate-crime/2017/topic-pages/incidents-and-offenses>

³ <http://hatecrime.osce.org/what-hate-crime/anti-semitism>



Antisemitism, antisemitic crimes grew by almost 15 percent in 2016 compared to 2015 and 44.5 percent compared to 2014.⁴

According to a report by the Online Hate Prevention Institute, which reviewed 2,000 antisemitism publications in the mainstream social media⁵ (Facebook, YouTube and Twitter), within the time span of ten months, only 20 percent of the publications that were followed have been removed.⁶

A report by the World Jewish Congress found that during 2016, a total of 382,000 antisemitic posts were posted online on social media.⁷ Sixty-three percent out of all posts were found on Twitter.

An analysis conducted by the ADL, estimates 4.2 million antisemitic tweets on twitter between the dates of February 2017 and January 2018.⁸

When addressing online antisemitism, especially on social media, one must acknowledge that online hate speech is not only a hate crime on its own, but also has real-life consequences that can lead to physical hate crimes.⁹ The European Commission against Racism and Intolerance called in its General Policy Recommendation No. 15 for speedy reaction by public figures to hate speech, and promoting self-regulation by media.¹⁰

⁴ <https://antisemitism.uk/wp-content/uploads/2017/10/National-Antisemitic-Crime-Audit-2016.pdf>

⁵ In this submission, the IAJLJ refers mainly to the mainstream internet and social media. This submission does not address the much more unregulated area of the dark-web.

⁶ Examples for such social media posts are available in the report at <https://ohpi.org.au/measuring-antisemitism/>

⁷ www.worldjewishcongress.org/en/news/an-anti-semitic-post-is-uploaded-to-social-media-every-83-seconds-wjc-research-finds-3-5-2017; The full report: www.worldjewishcongress.org/download/RVsVZzRXTaZwO41YbzIWwg?

⁸ <https://www.adl.org/resources/reports/quantifying-hate-a-year-of-anti-semitism-on-twitter#major-findings>

⁹ Müller, Karsten and Schwarz, Carlo, Fanning the Flames of Hate: Social Media and Hate Crime (November 30, 2018). Available at SSRN: <https://ssrn.com/abstract=3082972>

¹⁰ <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.15>



3. The Definition of Antisemitism and New Forms of Antisemitism

On 26 May 2016, the International Holocaust Remembrance Alliance (IHRA) adopted a working definition of antisemitism. According to that definition:

“Antisemitism is a certain perception of Jews, which may be expressed as hatred toward Jews. Rhetorical and physical manifestations of antisemitism are directed toward Jewish or non-Jewish individuals and/or their property, toward Jewish community institutions and religious facilities.”¹¹

This working definition was adopted and endorsed by several States and international bodies.¹²

Throughout the years, antisemitism has had different shapes and practices. In recent years a troublesome development can be identified: the development of antisemitism in the guise of anti-Israel and anti-Zionism. These are attempts to mask antisemitism under the premise of legitimate criticism against the state of Israel and its government.

The IHRA working definition attends such development, and includes the following examples of antisemitism:

- (1) Accusing Jewish citizens of being more loyal to Israel, or to the alleged priorities of Jews worldwide, than to the interests of their own nations;
- (2) Holding Jews collectively responsible for actions of the state of Israel;
- (3) Denying the Jewish people their right to self-determination, e.g., by claiming that the existence of a State of Israel is a racist endeavor; and
- (4) Drawing comparisons of contemporary Israeli policy to that of the Nazis.

According to the ADL, anti-Zionism can be described as the following:

¹¹ <https://www.holocaustremembrance.com/working-definition-antisemitism>

¹² Including, inter alia, United Kingdom, Israel, Austria, Scotland, Romania, Germany, Bulgaria, Belgium, Lithuania, Republic of North Macedonia, the Netherlands, Slovakia, Czech Republic, Greece, Hungary, and France. See <https://www.holocaustremembrance.com/working-definitions-and-charters>



“Words or actions related to Israel are anti-Semitic when they blame all Jews for the actions of the state, single out Israel in denying the country’s right to exist as a Jewish state and an equal member of the global community, use anti-Jewish stereotypes or conspiracy theories (such as accusations of Jewish world domination), or traditional anti-Semitic imagery or comparisons to Nazis.”¹³

Excluding this development from the scope of antisemitism enables antisemitism activity to blossom and operate under the false disguise of legitimate criticism.

This should not be interpreted as a restriction on legitimate criticism of the State of Israel or its policies; just like criticism against any other State, voicing criticism of Israel enjoys protection under freedom of expression, and specifically political expression.¹⁴ However, as mentioned above, that right is not unlimited, and article 20 of the ICCPR explicitly holds that the freedom of expression cannot be used to permit discriminatory advocacy and hate speech.

4. Freedom of Expression and Limitations

On 5 March 2019, the Special Rapporteur on Freedom of Religion or Belief presented a report to the Human Rights Council on the interaction between the freedom of thought, conscience, religion or belief,¹⁵ and the freedom of opinion and expression.¹⁶ The report noted that “the fates of these two rights are entwined”.¹⁷ Although fundamental, the right to freedom of expression is not unlimited. As stipulated in articles 19(3)(b) of the ICCPR and even more so in article 20(2), limitations of expression can be legally imposed by the State “[F]or the protection of national security or of public order (*ordre public*), or of public health or morals”,¹⁸ and for the prevention

¹³ www.adl.org/resources/tools-and-strategies/what-is-anti-israel-anti-semitic-anti-zionist

¹⁴ ICCPR, article 19.

¹⁵ ICCPR article 18

¹⁶ ICCPR article 19

¹⁷ A/HRC/40/58, March 5, 2019, para. 5; available at: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/060/28/PDF/G1906028.pdf?OpenElement>

¹⁸ International Covenant on Civil and Political Rights (ICCPR) article 19(3)(b)



of “[A]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence”.¹⁹

The CCPR committee addressed this specific notion in its General Comment 34, holding that limiting the freedom of expression must be done cautiously, and in accordance with State legislation. Moreover, any such limitations must conform to the requirements of necessity and proportionality.²⁰

Recently, the European Court of Human Rights (ECtHR or the Court) have issued a judgment related to the notion of freedom of expression under article 10 of the European Convention on Human Rights (ECHR). The Court held that the right for freedom of expression must be carefully balanced with the rights of others to have their religious feelings protected.²¹ In that case, the Court addressed an expression against the Muslim prophet Mohammad, and held that the limitation on the relevant expression was justified. The reasoning was explicitly based on the prohibition on incitement of hatred, including a reference to article 20(2) of the ICCPR, and General Comment 34. According to the Court:

“Where such expressions go beyond the limits of a critical denial of other people’s religious beliefs and are likely to incite religious intolerance, for example in the event of an improper or even abusive attack on an object of religious veneration, a State may legitimately consider them to be incompatible with respect for the freedom of thought, conscience and religion and take proportionate restrictive measures (citations omitted). In addition, expressions that seek to spread, incite or justify hatred based on intolerance, including religious intolerance, do not enjoy the protection afforded by Article 10 of the Convention.”²²

Moreover, the Court also reiterated that:

¹⁹ ICCPR article 20(2)

²⁰ CCPR committee general comment 34, available at: <https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf>

²¹ E.S. v. Austria (application no. 38450/12), of 25 October 2018 available at: <https://hudoc.echr.coe.int/eng#%7B%22itemid%22:%5B%22001-187188%22%5D%7D>

²² E.S. v. Austria, para. 43.



“... a religious group must tolerate the denial by others of their religious beliefs and even the propagation by others of doctrines hostile to their faith, as long as the statements at issue do not incite to hatred or religious intolerance”.²³

Social media enables and encourages freedom of expression, which is enshrined by the Universal Declaration of Human Rights²⁴ and the ICCPR. However, social media has also provided fertile ground for the manifestation of hate speech, xenophobia, and antisemitism. Limitations are required to prevent such illicit activity.

5. The Existing Use of Self-Monitoring in Social Media

In January 2018, the German law “Network Enforcement Act” (NetzGD),²⁵ enacted in 2016, came into force.²⁶ According to this legislation, companies that fail to remove illegal hate speech from social media platforms within 24 hours of receiving a complaint on such material, can be fined by up to EUR 50 million. However, while other European states have legislation that covers prohibitions on illicit content, the German legislation is considered unique with its legislative approach with a dedicated act designated for social media.²⁷

During the time that the German NetzGD was or enacted, the European Commission signed a Code of Conduct with Twitter, Facebook, Microsoft and YouTube.²⁸ According to this Code of

²³ E.S. v. Austria, para. 52.

²⁴ Universal Declaration of Human Rights, article 19

²⁵ Germany’s Network Enforcement Act (Netzwerkdurchsetzungsgesetz or NetzDG)

²⁶ Tworek and Leerssen, An Analysis of Germany’s NetzDG Law, available at:
https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf

²⁷ Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions On The Mid-Term Review On The Implementation Of The Digital Single Market Strategy A Connected Digital Single Market For All, p. 27, 10 May 2017, available at:
<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52017DC0228&from=EN>

²⁸ https://ec.europa.eu/info/sites/info/files/code_of_conduct_on_countersing_illegal_hate_speech_online_en.pdf



Conduct, these companies have committed to review within 24 hours any received valid notification on illegal hate speech within EU states.

According to the most recent implementation report of the Code of Conduct (February 2019), in approximately 89 percent of the cases, the companies reviewed notifications within 24 hours, and an additional 6.5 percent within 48 hours. It is also noted that over 70 percent of the notified content is removed from social media; the most commonly reported ground for hate speech is xenophobia (17 percent), followed by sexual orientation (15.6 percent). With regard to religion and origin grounds, anti-Muslim hatred (13.0 percent), followed by anti-Roma (12.2 percent) and antisemitism (10.1 percent).

While progress has been made, a number of issues still require attention:

- (1) Transparency and definitions - There is a lack of transparency on the grounds upon which a decision to maintain or remove a notified content by the companies is being made.²⁹ Thus, it is impossible to know what is the relevant working definition of antisemitism for each of the companies. Due to the lack of transparency, it is unclear for example, whether Anti-Zionism is included within the acknowledged hate speech. This is relevant both to the code of conduct and national legislation alike.

This lack of working definition, which is not unique to the code of conduct, has been approached by the European Parliament, by adopting the IHRA working definition of antisemitism on 29 May 2017, and calling all Member States to adopt and use this working definition.³⁰

- (2) Self-Monitoring - The Code of Conduct operates on a voluntarily basis, and does not require the social media companies to conduct self-monitoring for potential hate speech items and antisemitism.

²⁹ https://ec.europa.eu/info/sites/info/files/code_of_conduct_factsheet_7_web.pdf

³⁰ <http://ep-wgas.eu/2017/06/01/960/>



The framework of automatic self-monitoring using artificial intelligence (AI) already exists within social media. In June 2017, both Google and Facebook published separately the internal mechanisms aimed to fight terrorism online.³¹ These mechanisms, combining AI with human overview, aim to provide a balance between the potential limitations on privacy, freedom of expression and the commitment to fight online terrorism and violence.

It is important to note that although there are technical limitations to the ability to monitor every form of hate speech,³² social media companies should make all necessary efforts to continue to prevent and remove hate speech and antisemitism even without users' notification.

Therefore, due to the direct link between online antisemitism and physical hate crimes against Jewish communities and institutions, these existing mechanisms should also apply on self-monitoring on hate speech, including antisemitism.

- (3) Regional scope - The Code of Conduct is geographically limited the EU Member States. However, since online antisemitism is not limited to this region, and since social media includes billions of users, it is impractical to effectively fight online antisemitism hate speech without effectively monitoring and removing such content globally.

Therefore, it is important to expand the applicability of the Code of Conduct and similar frameworks beyond the EU Member States.

- (4) Platform limitations - Ongoing efforts by the mainstream social media companies to address hate crime and antisemitism are commendable, especially due to their massive exposure and popularity. However, these platforms are not the only available platforms. Forums such as 4chan, 8chan and Gab serve as a growing arena for unregulated and unmonitored hate speech and antisemitic discussions. While online hate speech, and antisemitism are at least partially monitored in the mainstream social media, these platforms are not under any voluntary

³¹ Facebook's notification is available here: <https://newsroom.fb.com/news/2017/06/how-we-counter-terrorism>;
Google's notification is available here: <https://blog.google/around-the-globe/google-europe/four-steps-were-taking-today-fight-online-terror>

³² Schmidt and Wiegand, 'A survey of hate speech detection using natural language processing' in Proceedings of the Fifth International Workshop on natural Language Processing for Social Media, Spain, April 2017, available at: <https://www.aclweb.org/anthology/W17-1101>



agreement for self-monitoring and thus are obliged only in accordance with relevant national legislation applicable to them. Such monitoring becomes even more difficult when taking into consideration alternative means of online communication such as the encrypted online apps such as WhatsApp and even more so, Telegram.

Therefore, it is recommended that States take active measures to ensure that all online platforms are subject to relevant monitoring.

6. Conclusions and Recommendations

Bearing in mind the exponential rise of antisemitism incidents in recent years and online hate speech, the IAJLJ recommends that the relevant UN bodies, including HRC, the relevant special rapporteurs, the Office of the High Commissioner of Human Rights, as well as Governments and the EU:

- 1. Acknowledge the direct link between online hate speech, online antisemitism, and antisemitic attacks.**
- 2. Acknowledge the distinction between legitimate criticism of Israel and antisemitism disguised as anti-Zionism.**
- 3. Operate to promote the public adoption of the working definition of antisemitism, by States and social media companies, and incorporate it in self-monitoring activities against hate speech.**
- 4. Recommend a stricter enforcement of existing mechanisms for self-monitoring on hate speech, including antisemitism.**
- 5. Recommend taking active measures to identify existing similar national and regional legislation, and expand the applicability of the Code of Conduct and similar frameworks beyond the EU Member States.**
- 6. Recommend taking active measures to ensure that all online platforms are subject to monitoring.**