

WORKING PAPER SERIES
Oxford Digital Diplomacy Research Group
Oxford Department of International Development
University of Oxford

Corneliu Bjola and Ilan Manor

Combating Online Hate Speech and Anti-Semitism

DigDiploROx Working Paper No 4 (Jan 2020)



#DigDiploROx



I. EXECUTIVE SUMMARY

The spread of hate speech and anti-Semitic content has become endemic to social media. Faced with a torrent of violent and offensive content, nations in Europe have begun to take measures to remove such content from social media platforms such as Facebook and Twitter. However, these measures have failed to curtail the spread, and possible impact of anti-Semitic content. Notably, violence breeds violence and calls for action against Jewish minorities soon lead to calls for violence against other ethnic or racial minorities. Online anti-Semitism thus drives social tensions and harms social cohesion. Yet the spread of online anti-Semitism also has international ramifications as conspiracy theories and disinformation campaigns now often focus on WWII and the Holocaust.

On Nov 29, 2019, the Oxford Digital Diplomacy Research Group (DigDiploROx) held a one-day symposium at the European Commission in Brussels. The symposium brought together diplomats, EU officials, academics and civil society organizations in order to search for new ways to combat the rise in online anti-Semitism. This policy brief offers an overview of the day's discussions, the challenges identified and a set of solutions that may aid nations looking to stem the flow of anti-Semitic content online. Notably, these solutions, or recommendations, are not limited to the realm of anti-Semitism and can help combat all forms of discrimination, hate and bigotry online.

Chief among these recommendations is the need for a multi-stakeholder solution that brings together governments, multilateral organisations, academic institutions, tech companies and NGOs. For the EU itself, there is a need to increase collaborations between units dedicated to fighting online crime, terrorism and anti-Semitism. This would enable the EU to share skills, resources and working procedures. Moreover, the EU must adopt technological solutions, such as automation, to identify, flag and remove hateful content in the quickest way possible. The EU could also redefine its main activities - rather than combat incitement to violence online, it may attempt to tackle incitement to hate, given that hate metastases online to calls for violence.

Finally, the EU should deepen its awareness to the potential harm of search engines. These offer access to content that has already been removed by social media companies. Moreover, search engines serve as a gateway to hateful content. The EU should thus deepen its collaborations with companies such as Google and Yahoo, and not just Facebook or Twitter. It should be noted that social media companies opted not to take part in the symposium demonstrating that the solution to hate speech and rising anti-Semitism may be in legislation and not just in collaboration.

The rest of this brief consists of five parts. The first offers an up-to-date analysis of the prevalence of anti-Semitic content online. The second, discuss the national and international implications of this prevalence. The third part stresses the need for a multi-stakeholder solution while the fourth offers an overview of the presentations made at the symposium. The final section includes a set of policy recommendations that should be adopted by the EU and its members states.

II. THE SPREAD OF ANTI-SEMITIC CONTENT ONLINE

Recent surveys suggest that Jewish communities throughout Western Europe and North America feel increasingly insecure while some believe that they must even hide their faith. This was made evident in a plea by a French Rabbi who urged Jews not wear the Kippah in public¹. The anxiety, uncertainty and fear now felt among Jewish communities stems from an increase in the number of anti-Semitic incidents over the past year, including the horrid attack on a synagogue in Pittsburgh that left eleven dead and many more injured. According to the Kantor Centre, 2019 saw a 70% increase in anti-Semitic incidents in Germany and France alone. A recent poll found that nine out of ten Jewish students in France have experienced anti-Semitism at least once during their studies. In the U.K., there has been a 16% increase in anti-Semitic incidents in 2019, with 1,652-documented case².

The rise in offline anti-Semitism is directly linked to a steady increase in online anti-Semitic content. As the French interior ministry recently stated, “not one day (passes) without an Anti-Semitic act online”. The Kantor Centre recently reported a “surge in online calls for the killing of Jews, for the extermination of Jews worldwide, and images of Jews being killed”.³ It is now estimated that some form of anti-Semitic content is published on social media every 83 seconds. A study by the Anti-Defamation League found that 4.2 million anti-Semitic Tweets were disseminated on Twitter in English during 2018. One week alone saw the publication of 181,700 anti-Semitic Tweets.⁴ When commenting on this finding, the CEO of the World Jewish Congress stated “We knew that Anti-Semitism online was on the rise, but the numbers revealed give us concrete data on how alarming this situation really is”.⁵

The past year has also seen the growing use of anti-Semitic images and tropes in online political discussions. As one scholar has noted, anti-Semitism is now used in mainstream political debates. One example is the use of anti-Semitic images by French internet users debating the Yellow Vests’ protests. Another example, documented by the Institute for the Future, is the use of “Age-old Anti-Semitic tropes and conspiracies...particularly among Twitter users that identify as Republicans and/or supporters of President Trump”.⁶

In a 2018 report, the European Commission against Racism and Intolerance found that anti-Semitic attacks including threats on rabbis, attacks on Jews wearing religious symbols and bullying in schools were actively promoted online in European Union (EU) member states. The report

¹ Haaretz Newspaper, <https://www.haaretz.com/jewish/dont-wear-a-kippa-unless-you-usually-do-1.5390995>

² Time Magazine, 2019, <https://time.com/5580312/kantor-center-anti-semitism-report/>

³ <http://www.kantorcenter.tau.ac.il/general-analyses-antisemitism-worldwide>

⁴ Anti-Defamation League, 2018, <https://www.adl.org/resources/reports/quantifying-hate-a-year-of-anti-semitism-on-twitter>

⁵ Jewish Daily Forward, 2017, <https://forward.com/fast-forward/367015/study-anti-semitic-content-posted-on-social-media-every-83-seconds/>; the New York Times, 2019, <https://www.nytimes.com/2019/04/04/world/europe/antisemitism-europe-united-states.html>

⁶ Institute for the Future, 2019, http://www.iftf.org/fileadmin/user_upload/downloads/ourwork/IFTF_Executive_Summary_comp.prop_W_05.07.19_01.pdf

further found that offline calls for violence against Jews were commonplace among radicalized Muslims, including schoolchildren, and in university campuses. These calls were propagated both offline and online. Even more disconcerting was the finding that online Holocaust denial and glorification of the Nazi past was on the rise in several EU states. The report concludes that anti-Semitic content is now published by both extreme right-wing and left-wing parties.⁷

A 2017 report by the Israeli government found that anti-Semitic content on social media deals primarily with four issues: incitement to offline violence against Jews and calling for the deaths of Jews (5%), Holocaust denial (12%), traditional anti-Semitic tropes and conspiracy theories (49%) and attacks on the State of Israel (34%). The report notes that on Facebook, anti-Semitic content is removed three months after it was published, on average. On YouTube, the removal of such content can take up to ten months with an estimated 90% of anti-Semitic content not being removed at all. On Twitter, incitement to violence against Jews is the least likely form of hate speech to be removed.⁸

Even though governments, multilateral institutions such as the EU and social media companies are aware of the volume of anti-Semitic content shared online, few concrete steps have been taken to limit the flow of such content. This inaction stems from several factors. First, multilateral institutions and governments have limited resources through which they are currently attempting to counter the spread digital disinformation, strategic misinformation and hate speech⁹. Anti-Semitic content is thus one more form of digital content that must be grappled with. Given the fear of digital disinformation, brought about by the nefarious digital activities attributed to Russia, Western governments are also allocating most of their resources to countering disinformation, while hate speech is treated as a secondary threat. Even the Israeli MFA has only one diplomat who is tasked solely with countering the spread of anti-Semitism online.

Moreover, social media companies have conflicting interests when it comes to removing anti-Semitic content. On the one hand, Facebook and Twitter's reputation has been tarnished given the ease with which hate speech can spread on their platforms. Moreover, recent terror attacks against Jews and Muslims have either been inspired by social media content or have been shared on social media in near-real time. On the other hand, companies such as Facebook and Twitter depend on their number of users for selling advertisements. Shutting down accounts that spread hate speech, including anti-Semitism, would reduce the number of users these companies have subsequently impacting their revenue.¹⁰

Another challenge facing governments and multilateral institutions hoping to counter the spread of anti-Semitism online is the need to collaborate with social media companies. Yet the latter are reluctant to share their data with governments, to explain how malicious content is removed or to shut down accounts that have been designated by governments as 'harmful'. Indeed, every month the Israeli MFA sends Facebook entire spreadsheets listing thousands of profiles that share anti-Semitic content. However, only a few of these are eventually shut down in what is a lengthy and drawn out process.

⁷ Anti-Semitism: Overview of the data available in the European Union 2007-2017, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-antisemitism-update-2007-2017_en.pdf

⁸ Measuring the Hate, <https://mfa.gov.il/MFA/ForeignPolicy/AntiSemitism/Documents/Measuring-the-Hate.pdf>

⁹ Manor, I. (2019). *The Digitalization of Public Diplomacy*. Springer International Publishing.

¹⁰ Manor, I., & Soone, L. (2018). The digital industries: Transparency as mass deception. *Global Policy*.

III. DOMESTIC AND INTERNATIONAL IMPLICATIONS

The spread of anti-Semitic content online affects European societies at large as anti-Semitism drives social tensions, harms social cohesion and often translates into an increase in violence. Moreover, as Martin Luther King stated, ‘violence only breeds more violence, adding darkness to a night already devoid of stars’. In other words, violence against Jews may lead to acts of violence against other minority groups be they Muslim communities, members of the LGBTQ community or racial minorities. In some countries, such as France, anti-Semitism has translated into an exodus of Jewish people to Israel. These waves of emigration deprive European countries of their cultural heritage, harm their economies and decrease their heterogeneity. In the face of rising anti-Semitism, European countries have also had to increase their expenditure on internal security. Yet those most affected by anti-Semitism are, of course, Jewish communities who face an onslaught of hate and violence on a daily basis.

The rise of anti-Semitism also contributes to creating a mutually reinforcing environment of public hate and suspicion, which has the potential to transcend borders and disrupt diplomatic relations. Recent digital disinformation campaigns affiliated with Russian authorities have tried, for instance, to exploit the historical trauma of the Holocaust in order to undermine the political standing of the Baltic governments in front of their own populations, their European partners or of the international community. A study conducted by the Oxford Digital Diplomacy Research Group has found that the political memory of the Holocaust has been weaponised through the strategic use of historical information on social and traditional media in an attempt to tarnish the reputation and national image of Lithuania and instigate diplomatic tensions with countries in the European Union as well as with Israel.¹¹ By deforming the context of discussion of the Holocaust through disinformation, such campaigns devalue the historical significance of the Holocaust and undermine international efforts to counter anti-Semitism.

IV. THE NEED FOR PARTNERSHIPS

Given the complex nature of the problem, the challenge of rising anti-Semitism on social media cannot be addressed by national authorities or international organisations alone. It requires a multi-stakeholder solution involving a coalition of actors at the national and international level, such as governments, multilateral organisations, academic institutions, tech companies, and NGOs. Technology companies can offer, for instance, innovative tools for identifying, mapping and removing anti-Semitic content. These may also help automate the struggle against anti-Semitism online. Academic institutions may assist in better defining the boundaries between hate speech and anti-Semitism. Non-governmental institutions can augment governments’ efforts to work opposite social media companies whose platforms have morphed into hubs of hate and calls for violence. Finally, networks of Jewish organizations may assist in flagging anti-Semitic content that should be removed from online platforms. A networked approach to combating online anti-Semitism may thus prove most effective.

As a first step towards developing this network, DigDiploRox convened on Nov 29, 2019, a one-day symposium at the European Commission in Brussels. The symposium was supported financially by the Philigence Foundation, Switzerland. The symposium brought together

¹¹ C Bjola and I Manor, “Managing Lithuania’s Historic Image through Strategic Communication,” *Policy Report*, July 2019.

academics, EU officials, and diplomats to share and discuss theoretical contributions, case studies and policy approaches that can better assist governments and international organisations in their efforts to counter the spread of anti-Semitism content online. The symposium addressed two main themes as follows:

I. *Mapping Anti-Semitism Online - New trends and Approaches*

Keeping up with new manifestations of online anti-Semitism is not easy as the digital medium is designed to primarily reward low-quality messages that cultivate controversy in a visually and emotionally engaging manner. Clear conceptual benchmarks are therefore needed to ensure that that online manifestations of anti-Semitism are accurately and timely identified.

- Drawing on definitions adopted by the International Holocaust Remembrance Alliance (IHRA) and the EU's Fundamental Rights Agency (FRA), Lesley Klaff, *Senior Lecturer in Law at Sheffield Hallam University*, offered a comprehensive overview of the legal aspects of the fight against digital anti-Semitism. She reviewed and responded to the main objections to these definitions, discussed the legal distinction between old and new anti-Semitism that tended to focus on the state of Israel, and presented a list of 'contemporary' examples that could serve as illustrations of online anti-Semitic expressions.
- Agne Kaarlep, *Policy Officer Directorate-General Migration and Home Affairs at the European Commission* highlighted the usually overlooked link between combating anti-Semitism and combating online terrorist content as both had been involved in promoting and generating violence. However, it also appears that states are more adamant when dealing with terrorist content, as opposed to anti-Semitic content. Notably, several Horizon 2020 funded projects are developing tools for mapping content shared on the Dark Web. While these tools are still in the prototype stage, they promise to increase the efficacy of identifying and removing terrorist-related content.
- Raphael Cohen-Almagor, *the Director of the Middle East Study Group (MESG) at the University of Hull*, explained the social responsibility of Internet intermediaries in tackling online hate and anti-Semitism. Internet intermediaries should closely monitor violent hate and terroristic forums, and issue alerts to warn Netusers and readers of the problematic content. This can be accomplished by two means: registration, such as when problematic material will have restricted access and people will have to sign up for reading it, providing some details about their identity and why they wish to read this particular piece of information; Stop! Messages, such as when individuals wishing to access dangerous information, will be presented with a warning message about the harmful nature of hate and terrorism. These messages aim to reduce the permissibility of violence speech. Cohen-Almagor also commented on the enduring relevance of the concept of deliberative democracy as a possible approach to confronting the Internet's dark side.
- Cornelii Bjola and Ilan Manor, members of the *Oxford Digital Diplomacy Research Group*, discussed the analytical deficiencies of the concept of echo-chamber for understanding the patterns of online dissemination of anti-Semitic content. They instead suggested that more attention should be paid to the mechanisms by which "rhetorical communities of hate" form, grow and interact with each other. They also advanced a novel methodology, based on social network analysis, that could better assist governments and social media companies to track and dismantle online hate communities.

II. *Combating Anti-Semitism Online – Solutions and Challenges*

As social media companies have been generally slow, unable or unwilling to take firm action against hate speech in general, and anti-Semitic content in particular, governments and international bodies have come under increasing pressure to identify alternative options that could assist them with this task. These options generally fall in two broad categories: regulatory frameworks and technological solutions.

- Katharina Von Schnurbein, the *European Commission’s Coordinator on combatting Antisemitism* provided an overview of the Commission’s efforts to engage social media companies, such as Facebook and Twitter, to persuade them to adopt policies that would allow hateful content to be removed from their platforms. New regulation is currently considered by the Commission to geofence the EU information space against hateful speech, especially from terrorist sources. The EU works to combat hate through educational activities, collaborations with local law enforcement agencies and by promoting legislation in member states.
- Daniel Allington, *Senior Lecturer in Social and Cultural Artificial Intelligence at Kings College London*, spoke of the need to pass legislation that would define social media platforms as publishers. Once that happens, social media companies will be legally liable for the content that they publish. Liability will increase the motivation of tech companies to combat hate speech since they may otherwise face the risk not only of criminal prosecution but also of civil action. Perhaps more importantly, it would establish a culture in which the owners of social media platforms are understood to hold civic responsibilities analogous to those of legacy media organisations such as periodical publishers. Currently, platforms such as Facebook facilitate the distribution of harmful content to vast audiences with no possibility of public oversight (for example, via ‘closed’ and ‘secret’ groups).
- Andre Oboler, *Lecturer in cyber-security, privacy and surveillance at La Trobe University Law School, Australia*, pointed out that current automatic detection programs are inaccurate, expensive and cannot cope with the volume of online anti-Semitic content. He instead proposed a hybrid approach that would draw on experts, crowd reviewers and Artificial Intelligence (AI) to better tackle this challenge.
- Elad Ratson, *Head of Research & Development at the Israeli Ministry of Foreign Affairs*, called attention to the slow and inefficient mechanism of reporting hate speech on social media platforms. He presented a new tool for flagging and reporting hateful content, which can facilitate up to 60% removal of reported tweets or the permanent suspension of hate users. Notably, Ratson stated that the Israeli MFA has attempted to collaborate with social media companies by flagging and reporting anti-Semitic content. However, these have failed to remove the majority of the content flagged by the MFA. Ratson also spoke of the need to create a coalition of countries that will collectively manage ties opposite tech giants. Such a coalition may be more effective in forcing social media companies to remove content.
- Sebastian Polzin, *Expert on anti-Semitism in the German Ministry of Foreign Affairs*, explained why the mere digitisation of conventional approaches to countering anti-Semitism would not work. Recent initiatives of the German government have instead focused on facilitating international collaboration via a network of special representatives and on developing off-the-shelf solutions for countering anti-Semitism via hackathons, digitization of archives and online educational tutorials.

V. RECOMMENDATIONS

The symposium has identified a series of important themes, ranging from institutional measures and analytical tools to regulatory frameworks and technological solutions, which together may contribute to addressing the issue of online anti-Semitism more effectively. As the challenge of anti-Semitism is a European-wide issue thus requiring close collaboration between all members of the European Union, the following set of recommendations are mainly tailored for the European Commission, but they are relevant for individual governments as well.

Institutional measures

1. Within the EU, and among member states, the issue of anti-Semitism often falls under the umbrella term of 'hate speech'. However, anti-Semitism is a unique form of hatred that rests on historical, ethnic, religious and political foundations. Successfully combating the spread of anti-Semitism online thus requires that resources be dedicated to this specific form of hatred, including by establishing a special unit, with its own resources, budget and staff, for combating anti-Semitism.
2. The EU should encourage collaborations between EU units dedicated to combating terrorism, crime and anti-Semitism. By sharing resources, skills and working procedures, the EU can increase the efficacy of its efforts to combat anti-Semitism online.
3. There is a recurring pattern to the emergence of anti-Semitic sentiments. An increase in social tensions, and crises in national democracies are often accompanied by a rise in anti-Semitic rhetoric. The EU may allocate its resources more efficiently by focusing on countries that are in the midst of or about to experience political and social upheaval.
4. While 72% of the content flagged by EU citizens as anti-Semitic is removed from social media platforms, very little content is actually flagged. The EU should create a network of diplomats, NGOs, civil society organizations, crowd reviewers and EU citizens who collaboratively identify and flag anti-Semitic content online. These efforts could be facilitated by the establishment of a network of special representatives following the German model.

Analytical tools

5. Anti-Semitism stems from stereotypes, conspiracies theories and misinformation related to both classical and contemporary forms of hate. It is important to distinguish between different forms of antisemitism and to tailor the response strategy accordingly. According to Irwin Cotler, a Canadian human rights law professor, classical anti-Semitism refers to the discrimination against, denial of, or assault upon the rights of Jews to live as equal members of the society they inhabit. The new or contemporary anti-Semitism involves the discrimination against, denial of, or assault upon the right of the Jewish people to live as an equal member of the family of nations, with Israel as the targeted "collective Jew among the nations".
6. Building on the work of the Israeli Ministry of Foreign Affairs, the EU may refocus its activities by attempting to answer three questions- what is the scale of anti-Semitic content shared online in EU member states? How can such content be identified in a minimal amount of time? How can such content best be removed?
7. Network analyses can help the EU and members states identify online 'rhetorical communities of hate' that spread misinformation about Jews. These analyses can be used for understanding the resonance of anti-Semitic messages, their patterns of dissemination, as well as their potential for fostering online radicalisation and offline violence. Moreover,

social media platforms should open up large online fora to public scrutiny: for example, many ‘closed’ and ‘secret’ Facebook groups have a reach which may rival that of legacy media organisations, yet their content is protected from public oversight in a way that would not be possible within the legacy media.

Regulatory frameworks

8. Current EU legislation focuses on combating incitement to violence online. The EU could redefine its activities and focus on combating ‘incitement to hate’. Such a redefinition will enable the EU to prevent online content from metastasizing into calls for violence. Indeed, a Facebook post that promotes a Jewish conspiracy theory soon evolves into a call to arms and, finally, to a call to violence. Moreover, by focusing on incitement to hate, the EU will be able to flag a higher volume of anti-Semitic content for removal.
9. While all EU member states ban Holocaust distortion, only a few enforce this. The EU should encourage member states to take steps to actively enforce the IHRA Working Definition of anti-Semitism, while supplying member states with the tools, resources and insight necessary to do so.
10. The most effective way to combat all forms of hate online, including anti-Semitism, is for member states and the EU to designate social media companies as publishers who are liable for the content shared on their platforms. This designation would force social media companies to remove all forms of hate from their platforms, or risk litigation.
11. Social media companies and Internet intermediaries should be encouraged to take seriously their social responsibility to tackle online hate and anti-Semitism and to make public the measures they have adapted to this end, on a regular basis. Moreover, independent third-party organisations should be engaged to audit and evaluate the effectiveness of such self-regulatory practices.

Technological solutions

12. While the EU has flagged and sought to remove hateful content from social media, it has yet to expand its efforts to search engines. Search engines offer access to hateful content that has already been removed from social media platforms. It is thus imperative that the EU work more closely with popular search engines such as Yahoo and Google. Notably, search engines are based on algorithmic filtering and the same query may yield different results in different nations. Thus, the EU should also encourage its member states to collaborate with search engines on the removal of hateful content at the national level.
13. The EU should also examine the use of automated solutions. Algorithms and artificial intelligence may be used to continuously scour social media platforms, identify content that elicits hate or promotes stereotypes against Jews and automatically reports such content to social media companies. Automation may be the best tool for reporting and combating online anti-Semitism given the sheer scale of hateful content shared online. However, for automated tools to function effectively, they must be able to access content within fora that are currently protected (such as ‘closed’ or ‘secret’ Facebook groups, especially where these are too large to be plausibly regarded as ‘private’ spaces).
14. Anti-Semitic content is often accompanied by code words (e.g., NRWN meaning New Race War Now). Mapping code words and using a lexicon of code words may increase the efficacy of EU efforts to identify and flag anti-Semitic content on social media.

Collaborations with NGOs and Jewish organizations may help the EU create, and continuously update such a lexicon.

15. To keep up with the evolving nature of online anti-Semitism, hackathons and off-the-shelf initiatives should be more frequently used as tools for exploring technical solutions to identifying, analysing and countering new expressions, channels and patterns of dissemination of anti-Semitic content.