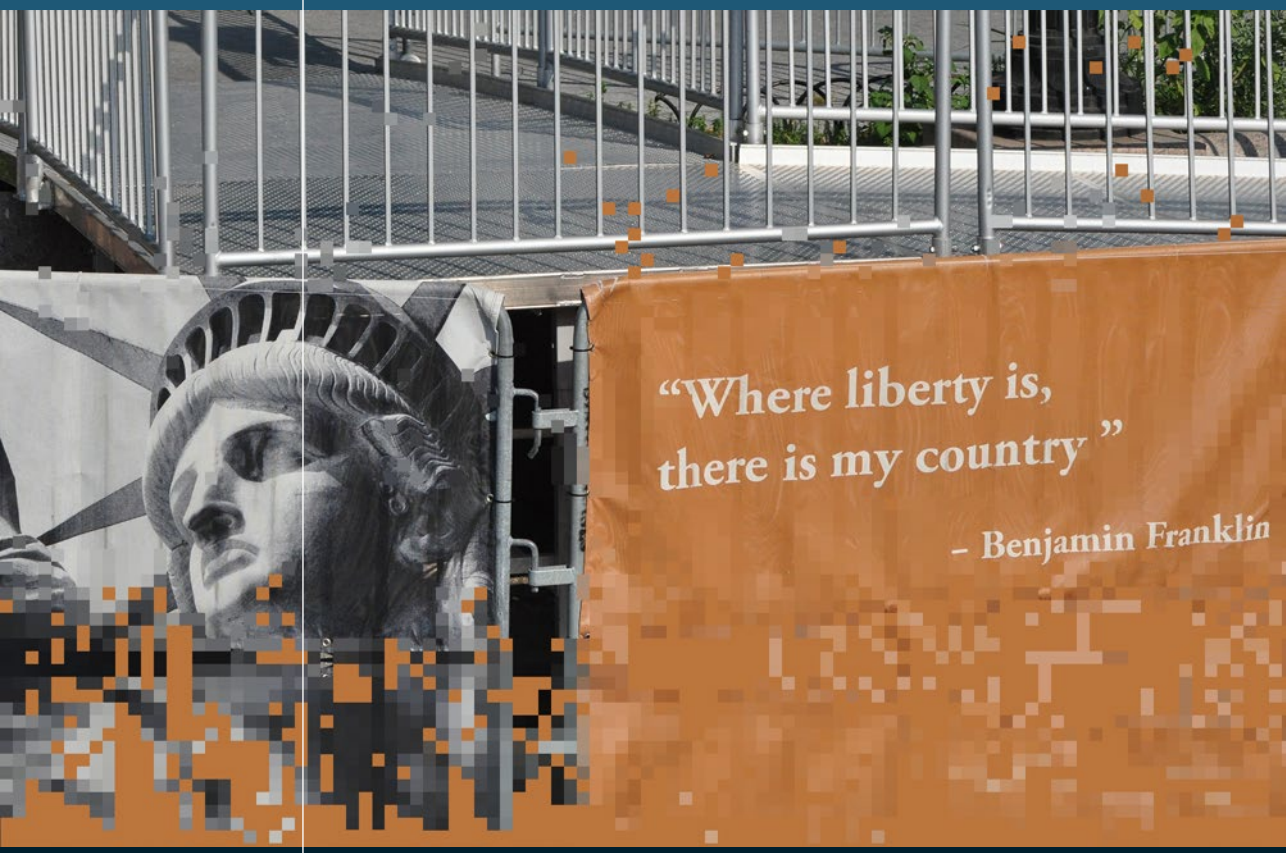
Radicalization at High Speed
Terrorism and the Media in the Age of Artificial Intelligence

Katarzyna Maniszewska



21 Collegium Civitas

STUDIES AND ANALYSES

Katarzyna Maniszewska

Radicalization at High Speed:
Terrorism and the Media in the Age of Artificial Intelligence

Warsaw 2025

Academic reviewers:
Jarosław Stelmach, WSB Merito University in Wrocław, Poland
Tomasz Aleksandrowicz, Police Academy in Szczytno, Poland

General editor, typography, and typesetting: Marek W. Gawron
Proofreading: Renee Stillings
Cover design: Aleksandra Jaworowska, studio@lionpath.pl
Cover photograph by Katarzyna Maniszewska

Volume 21 publication series: Collegium Civitas – Studies and Analyses

e-ISBN: 978-83-66386-53-2
DOI: https://doi.org/10.6084/m9.figshare.2821838

Publisher: Collegium Civitas
Ministry of Education and Science Identification Number: 19800

Palace of Culture and Science, 12th floor, 00-901 Warsaw, plac Defilad 1
phone: +48 500 807 895 | e-mail: wydawnictwo@civitas.edu.pl | http://www.civitas.edu.pl

Freedom of (Hate) Speech?

The question of the legality of inspirer actions – that is, spreading disinformation, hate speech, and inciting violence – is of fundamental importance here. To understand how relevant this point is, let us first take a brief look at the regulations within the European Union countries and then compare them to United States policies. The main difference between the EU and the USA in this regard is that hate speech is illegal under European Union law (of Council Framework Decision 2008/913/JHA 28 November 2008), while in the US, it is constitutionally protected by the First Amendment.

Thus, in the U.S. private companies and their goodwill and commitment to imposing internal regulations banning hate speech provide the backbone of radicalization prevention, with legal measures additionally in place. Of course, we do need to consider the degree to which owners of social media channels are interested in preventing radicalization. The example of Telegram and its owner, Pavel Durov, in custody in France at the time of this writing (Davies 2024) shows that this is not always the case.

Twitter/X, under the leadership of Elon Musk, is another example where the owner is reluctant to implement hate speech protection measures and enforce the X community rules of forbidding hate speech and incitement to violence (X Corp. 2023). This was shown by a report of the Polish association monitoring hate speech "Nigdy Więcej/Never Again." Even when there are laws against hate speech in a country where X operates, the platform is hesitant (to say at least) to take down hateful content. The report details hundreds of cases of hate speech (against different groups: Jews, Muslims, Ukrainian minority in Poland, LGBTQ+ community) in the Polish language reported by the association Nigdy Więcej to X in 12 months, with the platform refusing to remove or ignoring the vast majority of reports (approx. 90% of reported hateful comments were kept online by X (as estimated by Nigdy Więcej/Never Again Association 2024: 2).

In the European Union hate speech is punishable by law and there is a catalogue of forms of conduct that are considered hate speech, among them:

- "public incitement to violence or hatred directed against a group of persons or a member of such a group defined on the basis of race, color, descent, religion or belief, or national or ethnic origin [Including public dissemination or distribution of tracts, pictures or other material – KM];
- "publicly condoning, denying or grossly trivializing crimes of genocide, crimes against humanity and war crimes as defined in the Statute of the International Criminal Court (Articles 6, 7 and 8) and crimes defined in Article 6 of the Charter of the International Military Tribunal, when the conduct is carried out in a manner likely to incite violence or hatred against such a group or a member of such a group" (Council of the European Union 2008: Summary).

Further, on the European Union level, attempts to regulate the dissemination of terrorist content online have been made with Regulation 2021/784 of the European Parliament and of the Council of 29 April 2021 (European Union 2021). The Regulation came into force in 2022, and since then, Internet companies in the EU must take swift measures to prevent the misuse of their services for the dissemination of terrorist content (Wahl 2022). Regulation 2021/784 emphasizes that terrorist content online has serious negative consequences for societies ("While not the only factor, the presence of terrorist content online has proven to be a catalyst for the radicalization of individuals which can lead to terrorist acts") and online service providers hosting such content "since it undermines the trust of their users and damages their business models" (European Union 2021).

However, as Ramin Farinpour noted in 2021 in the article *A snapshot of recent developments regarding EU counterterrorism policies and legislation*, the Regulation is seen as controversial, and the question of freedom of speech is being brought up. "Several NGOs continue to see the new Regulation as a significant threat to freedom of expression. In particular, the broad understanding of 'terrorist content' poses the risk that orders for political purposes will be abusively issued under the guise of combating terrorism" (Farinpour 2021: 368).

As already stated, in the U.S., hate speech is protected by the 1st Amendment: „Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press;

or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances" (U.S. Bill of Rights, ratified 1791). The First Amendment prevents prosecution of hate speech, unless „it directly incites imminent criminal activity or consists of specific threats of violence targeted against a person or group" (American Library Association 2017). Provisions such as codified in the EU Regulation 2021/784 would not have been possible under the 1st Amendment.

The report on online extremism by the United States Government Accountability Office (GAO) issued in January 2024 sheds more light on radicalization through hate speech that can potentially lead to offline terrorist attacks. The authors define hate speech as "derogatory speech against individuals or groups based on their actual or perceived characteristics such as race, color, religion, ethnicity, national origin, gender, gender identity, disability, or sexual orientation" (United States Government Accountability Office 2024: 7).

The report includes a comprehensive review of studies on hate speech and hate crimes in the U.S. conducted in recent years, as well as data and interviews with representatives of companies operating online platforms, including four social media platforms (for purposes of the report, they are anonymized). The interviews offer some interesting insights, such as: "[a]ccording to officials from a company that operates one of the four social media platforms, violative hate speech content from the U.S. represented about 39 percent of violative hate speech content globally from October 2022 through December 2022" (Ibid.: 30). Americans represent about 4% of the global population – 4% of the global population versus 39% of hate speech content. Of course, due to the anonymity, we cannot analyze it thoroughly – we lack information on the share of users globally per country and the absolute numbers of users (theoretically, it is possible that, e.g., a smaller number of users being more prolific is responsible for the hate speech content production). Still, it shows a disturbing, however not surprising picture, given the constitutional protection of hate speech in the U.S. We can add to this picture the data on arms in private possession in the U.S. Bruce Hoffman and Jacob Ware cite the following numbers in their book *God, Guns, Sedition*: U.S.

citizens own 40% of firearms globally, that is approx. 400 million weapons (Hoffman, Ware 2024: 7).

The U.S. is overrepresented in both areas: the amount of hate speech content online and in weapons possession. At the same time, the research cited in the GAO report suggests that the occurrence of hate crimes is associated with hate speech on the Internet. The findings suggest that levels of hate speech online are correlated with hate crimes. An increase in hate speech observed online was synchronized with the time of terrorist attacks, giving the examples of the attack during the rally in Charlottesville (2017) and at the Tree of Life Synagogue in Pittsburgh (2018).

Further, the GAO report is aligned in its conclusions with Gilles Kepel's jihadism of atmosphere theory, namely that individuals radicalized on the Internet can perpetrate violence as lone actors (GAO 2024: 43). It shows how – paradoxically – irrelevant the ideology behind radicalization is. In the American context, the most common motif is the far-right, which is broadly understood (with submovements). However, the patterns are characteristic of those identified by Kepel within jihadism.

The author's hypothesis is that a similar phenomenon occurs with left-wing extremists and a more contemporary trend of the "salad bar extremism"[18] or mixed ideology extremism. Alexander Meleagrou-Hitchens and Moustafa Ayad analyzed the spectrum of mixed ideologies in the report *The Age of Incoherence? Understanding Mixed and Unclear Ideology Extremism,* published in June 2023 by the Program on Extremism. In the report, Meleagrou-Hitchens and Ayad focus primarily on the mixture of jihadist and, conspiracy and alt-right ideologies.

This broader ideological mixture of extremist left-wing, jihadism, and alt-right deserves further attention. These ideologies are aligned in their antisemitism, which, after the October 7, 2023 Hamas attacks on Israel, seems to be a theme uniting extremists of different ideological hues. The European Police Office

[18] The term was first used in 2020 by FBI Director Christopher Wray to describe the nature of some of the recent violent extremist threats. See: Ayad, Conroy, Meleagrou-Hitchers 2023.

Europol has also identified the trend of mixed ideologies. The Terrorism Situation and Trends Report 2023 Europol states that "[t]he lines between diverse types of terrorism, including right-wing, left-wing, anarchist, jihadist, and other ideologies, are likely to become more blurred in the future. Points of convergence have already been observed among terrorist and violent extremist actors across the whole ideological spectrum" (Europol 2023: 73). The same phenomenon was noticed on both sides of the Atlantic. More research is needed to determine the possible scope and effects of this "terrorism mix," especially post-October 7, which, as the author suspects, may be one of the turning points in the development of the "terrorism mix," with antisemitism as ideological glue for the otherwise fragmented and often contradictory ideologies.

How specifically right-wing extremists are capitalizing on the rise of hatred and hostility towards Jews has been shown by the study *From Memes to Mainstream: How Far-Right Extremists Weaponize AI to Spread Antisemitism and Radicalization* where authors affiliated at the International Institute for Counter-Terrorism, Reichmann Institute, Israel analyze the usage of an anti-Israel narrative "to radicalize individuals across the ideological spectrum" with AI-enhanced memes being one of the tools of radicalization (Koblentz-Stenzler, Klempner 2024). Memes are weaponized (Goldenberg, Finkelstein 2020), and as John Giesea points out, memetic warfare is conducted by state and non-state malign actors. Memes are used as the "currency of propaganda" (Giesea 2016: 68). Hamas is one of the examples of efficiency in the weaponization of memes.

It should be explained that in this article, the key terms antisemitism and genocide are understood according to the definitions by the International Holocaust Remembrance Alliance (IHRA) and as adopted by the UN in the Convention on the Prevention and Punishment of the Crime of Genocide. The working definition of antisemitism as adopted by IHRA reads: „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" (International Holocaust Remembrance Alliance 2016: 1). IHRA

makes it also clear that „criticism of Israel similar to that leveled against any other country cannot be regarded as antisemitic" (Ibid.). This part is of vital importance, especially in the context of the antisemitic rhetoric that re-gained prominence after the October 7, 2023, attacks on Israel, capitalizing also on false narratives against Israel (Center on Extremism 2024; Eisele, Steinwehr 2023).

The second crucial term in this context is genocide. Polish-Jewish lawyer Raphael Lemkin, who coined this term in 1944 in response to the atrocities of World War II, defined it as "a coordinated plan of different actions aiming at the destruction of essential foundations of the life of national groups, with the aim of annihilating the groups themselves" (Lemkin 1944: 79). The UN Convention adopted in 1948 defines genocide as:

> "any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such:
>
> (a) Killing members of the group;
>
> (b) Causing serious bodily or mental harm to members of the group;
>
> (c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part;
>
> (d) Imposing measures intended to prevent births within the group;
>
> (e) Forcibly transferring children of the group to another group" (UN Convention on the Prevention and Punishment of the Crime of Genocide, 1948).

A study *Antisemitic Attitudes Across the Ideological Spectrum*, published in 2023 (written prior to the October 7, 2023 attacks), showed that the epicenter of antisemitic views in the U.S. was with young adults on the far right. However, the study also looked at left-wing antisemitism, where anti-Israel attitudes are also present (Hersh, Royden 2023).

More research and content analysis are needed, but the links between antisemitism and left-wing politics are visible in many articles in the American media post-October 7. To name a few examples: *Antisemitism has moved from the right to the left in the U.S. – and falls back on long-standing stereotypes* by Arlie Perliger (2023) in *The Conversation*, *How the Activist Left Turned On Israel* by Charlotte Alter in *Time Magazine*, *The Long Story of Left-Wing Antisemitism* by Dan Hannan

(2024) in the *Washington Examiner*, *Will Progressives Confront Left-Wing Antisemitism?* by Will Marshall (2024) in *The Hill*, *'I Just Couldn't Take It': How a Jewish Politician Decided to Confront Left-Wing Antisemitism* by Alexander Burns (2023) in *Politico*, *The Golden Age of American Jews Is Ending* by Frankin Foer (2024) in *The Atlantic*, How Hamas Won Hearts and Minds on the American Left by Lorenzo Vidino (2023b) in the *Wall Street Journal*.

The coverage of left-wing trends towards antisemitism, along with the extensive coverage of protests on American campuses, including the antisemitic attitudes of some of the universities' staff members, show that antisemitism is on the rise across the ideological spectrum. Columbia University can serve here as an example with the well-documented case of three staff members who were removed from their positions "after finding that text messages they exchanged during a campus discussion about Jewish life 'disturbingly touched on ancient antisemitic tropes'" (Associated Press 2024).

Similar developments can be observed in the European Union countries; even though hate speech is not legal – as was shown at the beginning of this section – the extremism of the atmosphere is becoming visible in societies both online and offline. In recent months, it was especially visible during the pro-Palestinian protests, manifesting in the form of clandestine or openly expressed antisemitism, support for terrorists (perceived as "freedom fighters"), chanting of songs calling for the annihilation of Israel ("from the river to the sea…").[19] How the attack on October 7, 2023, amplified radicalization tendencies in the United States, the United Kingdom, France, Germany, and Italy was presented in the Soufan Center Special Report *Accelerating Hate: The Impact of October 7 on Terrorism and Political Violence in the West* (Broekaert et al. 2024).

The radicalization dynamics show that the protests, even if they may start peacefully, in many cases end up in violence. The radicalization processes may

[19] Which the author of this report could see for herself in Washington, D.C. in August and October 2024.

be broken down into three main phases – radicalization stages, regardless of the ideology that drives the protesters:

- Active demonstration of social disagreement;
- Aggressive demonstration of disagreement, which includes elements of physical violence;
- Terrorism.

What can be historically seen as a characteristic feature of the radicalization pattern in democracies and open societies is that with each stage, public support declines (Maniszewska 2024). This can also be translated to the online sphere. The vast majority of Internet and social media users are not engaged in propagating content supporting terrorism. In addition, Mia Bloom emphasizes that there is a difference between radical speech and action, and even a smaller percentage of social media users would go out to the offline world to perpetuate a violent act.[20] Obviously, it does not mean that we should underestimate those who can be radicalized to the point of committing a crime. The aforementioned attack in Solingen, Germany, where a lone actor stabbed to death three people and wounded eight, was claimed on a Telegram account by the Islamic State as part of the revenge for Muslims in Palestine and everywhere, can be seen as part of the radicalization cycle and effect of the "terrorism of atmosphere" (Maher, Tolba 2024).

[20] Mia Bloom. Personal Interview. 23 October 2024.