Digital Security Lab

IMS
GOOD JOURNALISM | BETTER SOCIETIES

For whom the bell tolls:
responsibility for disinformation in wartime

Tetiana Avdieieva
Maksym Dvorovyi

Digital Security Lab Ukraine
2022

For Whom The Bell Tolls: Responsibility For Disinformation In Wartime (Mapping report) - Tetiana Avdieieva, Maksym Dvorovyi.
Digital Security Lab Ukraine, Kyiv, 2022.

This mapping report explores the status of false information in the armed conflict and its legal framing therein. Specifically, the report provides an extensive overview of the human rights, public international law, and international criminal law regulations applicable to the phenomenon of disinformation. It defines the place of disinformation within the existent notions of propaganda for war, incitement to genocide and other international crimes, hate speech and calls for the violation of State sovereignty. The report also outlines the legal responses available for tackling disinformation delivered during the armed conflict based on the numerous historical examples of the invoked liability.

This publication was made possible through support provided by IMS (International Media Support). The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of IMS.

TABLE OF CONTENTS

SUMMARY

Russian illegal invasion of Ukraine became an emblematic example of warfare that takes place both on the physical battlefield and within the audiovisual and online media, printed press, and Internet intermediary platforms. The avalanches of misleading materials, targeting domestic and international audiences, not only affect the psychological state of the readers but also contribute to the improper legal qualification of the armed activities, distort the attributability of the violations, as well as facilitate the escalation of hostilities. This, in turn, frequently results in deaths of the civilian population, the commission of war crimes, crimes against humanity, and other less grave violations.

The aim of this mapping report implies the exploration of the status of false information in the armed conflict and its legal framing therein. Specifically, the report provides an extensive overview of the human rights, public international law, and international criminal law regulations applicable to the phenomenon of disinformation. It defines the place of disinformation within the existent notions of propaganda for war, incitement to genocide and other international crimes, hate speech and calls for the violation of State sovereignty. The report also outlines the legal responses available for tackling disinformation delivered during the armed conflict based on the numerous historical examples of the invoked liability.

The analysis of the practice of the international courts dealing with human rights violations, State responsibility, and individual criminal liability enabled to define the tendencies in the qualification of false information shared with an intent to harm, as well as the threshold for establishing a breach. These findings will help focus on the necessary elements of conduct and speech during the collection of evidence of atrocities committed by Russia in Ukraine for future international proceedings aimed at bringing the responsible perpetrators to justice.

Digital Security Lab Ukraine also provided recommendations for Ukraine and other relevant stakeholders regarding the potential ventures for triggering the international responsibility of Russia, its citizens, and other persons involved in the coordinated disinformation campaigns within the armed conflict launched against Ukraine. The recommendations, in particular, address:

- the mechanisms for collection of evidence and legal assessment of available materials by the international human rights bodies, including the judicial and quasi-judicial ones, as well as proper implementation of the international law prohibitions on propaganda for war, illegal incitements, and discriminatory speech;

- the applicable legal instruments to be invoked for ensuring the liability of Russia as a State for the dissemination of disinformation before and within its armed aggression in Ukraine;
- available instruments for invoking the individual criminal liability of persons spreading disinformation, propaganda for war and illegal incitements, including the ICC legal framework and the creation of a special criminal tribunal.

TABLE OF ABBREVIATIONS

AC	Appeals Chamber
ACHR	American Convention on Human Rights
AComHPR	African Commission on Human and People's Rights
AI	Artificial intelligence
AP	Additional Protocol
ARSIWA	Articles on Responsibility of States for Internationally Wrongful Acts
CERD	International Convention on the Elimination of All Forms of Racial Discrimination
CERD Committee	Committee on the Elimination of Racial Discrimination
CJEU	Court of Justice of the European Union
CoE	Council of Europe
DPR	Donetsk People's Republic (temporarily occupied territories in Donetsk region of Ukraine)
DRC	Democratic Republic of Congo
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
EU	European Union
HRC	United Nations Human Rights Committee
IAComHR	Inter-American Commission on Human Rights
ICC	International Criminal Court
ICCPR	International Covenant on Civil and Political Rights
ICJ	International Court of Justice
ICRC	International Committee of the Red Cross
ICTR	International Criminal Tribunal for Rwanda
ICTY	International Criminal Tribunal for Yugoslavia
IHL	International humanitarian law
IHRL	International human rights law
IMT	International Military Tribunal
IMT for the Far East	International Military Tribunal for the Far East
LPR	Luhansk People's Republic (temporarily occupied territories in Luhansk region of Ukraine)
NATO	North Atlantic Treaty Organization
NGO	Non-governmental organisation

OAS	Organisation of American States
OHCHR	Office of the United Nations High Commissioner for Human Rights
OSCE	Organization for Security and Co-operation in Europe
p/pp	Page/Pages
para/paras	Paragraph
PCIJ	Permanent Court of International Justice
PIL	Public international law
RTLM	Radio Télévision Libre des Mille Collines
SCSL	Special Court for Sierra Leone
TC	Trial Chamber
UK	United Kingdom of Great Britain and Northern Ireland
UN	United Nations
US	United States of America
USSR/Soviet Union	Union of Soviet Socialist Republics

Introduction

Without a firearm, machete or any physical weapon,
you caused the deaths of thousands of innocent civilians.
(c) The ICTR in The Media case

Disinformation fairly falls within the notion of *"nothing new under the sun"*, with a serious legal debate around it dating back to at least the beginning of the 20th century. A particular interest at that time, however, was attracted not by the pre-election manipulative narratives or the conspiracy theories surrounding the pandemics. Human rights defenders, academic circles, international lawyers, and the States themselves were interested in regulating the data flows related to the armed conflicts. The first comprehensive doctrinal studies on the international responsibility for foreign propaganda are ascribed to the US scholars, who focused on the procedural mechanisms of State responsibility.[1] It is needless to say that the governments expressed extensive support for regulations in this area to explicitly define the frames of acceptable behaviour. The desire to develop a legal framework only strengthened with the emergence of the Internet and other facilitative tools for the speedy delivery of messages, especially ones allowing for anonymity.

The classical definition of disinformation, designed by the CoE in their brief on information disorder, outlines it as **a deliberate misleading message delivered with an intent to cause harm**.[2] In practice, disinformation is mostly attributed to the specific types of content, serving as a form of speech rather than its substance. For example, materials harming public health can be either false or true, but they nevertheless remain prohibited. The same rule applies to non-consensual sharing of the details of sexual life of the candidates at the elections, which despite their factual basis (or its absence), are still restricted in circulation. And finally, the same approach shall be maintained towards disinformation spread in the context of the geopolitical crisis of various scales – from social unrest to international armed conflicts.

> If disinformation serves as a constituent element of the illegal incitements, propaganda for war, war crimes or crimes against humanity, calls for a breach of sovereignty or other violations of the State's national security – **it is already prohibited** as a part of such crimes or types of prohibited content. Akin to that, it may become an aggravating circumstance, evidence of complicity or aid in the commission of a crime.

In the dimension of armed conflicts, disinformation sporadically constitutes a part

[1] Vernon Van Dyke, 'The Responsibility of States for International Propaganda' (1940) 34(1) AJIL 58

[2] Claire Wardle, Hossein Derakhshan, Information Disorder: Toward an interdisciplinary framework for research and policy-making (CoE, 2017) 20 <https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77> accessed 22 December 2022

of propaganda, incitements to genocide, psychological operations, influence and information operations, cognitive warfare *etc*.[3] All of the listed categories are generally perceived as negative or, at least, as inhabitants of the 'grey zone' of freedom of expression laws. To exemplify, propaganda is frequently treated as *"a deliberate attempt to alter or maintain a balance of power that is advantageous to the propagandist"*,[4] which during wartime implies interference with the internal affairs of the other State. As one might have noticed, the definition is vague and lacks specific legal regulation, which constitutes one of the biggest problems when the issue of responsibility arises. Similarly, information operations are not explicitly allowed or outlawed but are rather subsequently penalised depending on the circumstances. A couple of historical examples only prove the practical difficulty of qualifying false content within armed conflicts:

- Starting from the **Nazi propaganda machinery** – Hitler resorted to the Volksdeutschen notion *"to transform the reality of German aggression into an illusion of the opposite"*.[5] Particularly, Germans were depicted as victims of European maltreatment, suffering from underestimation and cruelty.[6] As it turned out in practice, the propaganda efforts were quite successful, causing protests against the peaceful co-existence in Europe and calls for revenge after the "suppressive attitude" following World War I.[7] Accordingly, misleading media campaigns formed a social perception of hostility, serving a ground for Nazi ideas and their fruitful development in German society of the 1930th, which resulted in the launch of World War II. Despite information serving as a root cause of the German aggression, though, only a couple of propagandists were practically brought to justice, while the State itself bore no legal responsibility for information operations (will be analysed in detail further);
- The spreading of disinformation and manipulation of legal concepts around the **Soviet Union invasion of Poland** in 1939, still celebrated as a liberation campaign in modern Russia,[8] is another example of malicious information operations which met no international response. In this respect, lots of people in territories of the former Soviet Union still support such an interpretation of the Soviet Union's participation in World War II, creating a fertile ground for pro-Russian, anti-European propaganda, hostile attitudes towards the West and fears around the expansion of its influence;
- Analysing the **war in Iraq**, scholars outlined a set of problems caused by misleading propaganda, *i.e.* it limited the availability of facts, undermined

[3] Eian Katz, `Liar's war: Protecting civilians from disinformation during armed conflict` (2020) 102 (914) IRRC, 659–682, 663

[4] Garth S Jowett and Victoria O'Donnell, Propaganda and Persuasion (5th edn.) (SAGE Publications, 1992) 2

[5] Doris L. Bergen, 'Instrumentalization of «Volksdeutschen» in German Propaganda in 1939: Replacing/Erasing Poles, Jews, and Other Victims' (2008) 34 (3) GSR, 447-470

[6] German Foreign Office, Documents on the Events Preceding the Outbreak of the War (University Press of the Pacific, 19 June 2004) 447, 462

[7] John Reid, 'Putin, Pretext, and the Dark Side of the "Responsibility To Protect"' (War on the Rocks, 27 May 2022) <https://warontherocks.com/2022/05/putin-pretext-and-the-dark-side-of-the-responsibility-to-protect/> accessed 27 December 2022

[8] МИД России, Twitter post (Twitter, 17 September 2021) <https://twitter.com/MID_RF/status/1438768364353114115?s=20> accessed 27 December 2022

the transparency of political motivation, created a climate of prejudice open to violent responses, and polarised societies at various levels.[9] In this respect, the demonisation of the Iraqi government and population by the US Bush administration significantly affected the perception of the US forces by the locals, leading to the escalation of clashes. In practice, the main difficulty implied qualification of the role of media materials as inciting to illegal actions since no legal framework existed at the relevant time (or, at least, was binding for the US);

- A similar reaction happened in the **Afghani conflict**, where the US propaganda regarding the prisoners of war often depicted the latter in a false light, thus undermining the trust in the US State institutions.[10] At the very last, in 2002, Pentagon tried to create a coordinated centre on propaganda delivery, which faced lots of criticism from local and foreign media.[11] This, in turn, led to the escalation of tensions, both on the inter-State level and among the population of the conflicting parties. However, no solid legal response was developed since information operations were *de facto* lost among more severe violations, such as physical violence and breach of sovereignty issues.
- Another case is an embrace of the 'responsibility to protect' concept[12] to justify the **Russian illegal invasion of Georgia**, *i.e.* an attempt to view the aggression under the sauce of international obligations to prevent internal clashes and human rights breaches in the foreign territories.[13] It is worth mentioning, though, that information operations in Georgia, especially those, aimed at ethnic Russian population, were rarely addressed when analysing the ways to bring Russia to justice for violation of Georgian sovereignty.

But, if violence gives birth to violence, disinformation campaigns lead to more false narratives being cultivated and shared. In this respect, the recent aggressive invasion of Russia of Ukraine is one of the brightest examples. Namely, large-scale disinformation operations were used in all possible manifestations and, probably, for any possible purpose one can imagine:

- **Disinformation spread before the war** (before the 2014 occupation of the Crimean Peninsula). In the context of Russian aggression against Ukraine, alternative facts, denial of legal concepts, and reference to the non-existent violations of international law on behalf of the counterparty to the conflict have long been employed by the Russian side, *e.g.* the myth about Ukrainian shelling of Donbas or discrimination of Russian-speaking population in Eastern

[9] Elizabeth Willmott-Harrop, 'Iraq: propaganda's war on human rights' (Liberty & Humanity, June 2005) <https://libertyandhumanity.com/themes/international-human-rights-law/iraq-propagandas-war-on-human-rights/> accessed 27 December 2022

[10] Anup Shah, 'War on Terror Mainstream Media and Propaganda' (Global Issues, 1 August 2007) <https://www.globalissues.org/article/352/mainstream-media-and-propaganda> accessed 27 December 2022

[11] Ibid

[12] 'Responsibility to Protect' (United Nations) <https://www.un.org/en/genocideprevention/about-responsibility-to-protect.shtml> accessed 27 December 2022

[13] Ivana Stradner, 'A Responsibility to Protect Ukraine' (AEI, 22 February 2022) <https://www.aei.org/foreign-and-defense-policy/a-responsibility-to-protect-ukraine/> accessed 27 December 2022

Ukraine.[14] *De facto,* these self-invented violations were proclaimed as the main grounds for interference with Ukrainian internal affairs in 2014, the illegal referendum in Crimea and the occupation of Eastern Ukraine. This narrative circulates within the occupied territories up to this date, in parallel praising the Russian economy, contributions to the restoration of the occupied cities after its separation from Ukraine and other related misleading propagandistic materials.

- **Disinformation spread on the eve of the full-scale invasion**. Rapid integration of Ukraine with the EU and NATO forced the Russian government to look for other factual and legal manipulations to strengthen the political and military pressure over Ukraine and the region. For instance, the reference to the development of chemical weapons by Ukrainian authorities under the auspices of the US was used as a precondition for additional mobilisation in Russia.[15] Though, it is essential to mention that no such military or scientific research happened at the time.[16] Another reason was even more general and vague in its nature, elaborating on the need to stop the expansion of NATO across the Russian borders and prevent its potential invasion of Russia itself.[17] This, in turn, was supplemented by the comparison of Ukrainians with Nazis while amplifying the praise of the Red Army and their fight against Germany during World War II.
- **Disinformation over the course of the aggression**. The narrative regarding the development of biological weapons was advanced to the international level, reaching the Geneva Conference.[18] Although most of the participants qualified Russian accusations as a part of its disinformation campaign, the scale of such a campaign and its outreach grew significantly, given the increased attention to Russia as an aggressor. Apart from claims under international law, the Russian side also manipulates the number of dead soldiers and destroyed military equipment and vehicles,[19] distorting the visual image of the course of the illegal invasion. Apart from that, Russia tries to distort the Ukrainian image abroad: distributing misleading information on inexistent crimes committed by Ukrainian refugees, provocations allegedly prepared by Ukraine, concluded or prepared peace treaties *etc*.[20] To exemplify, after the inhumane

[14] New York Times and Office of the President of Russia, 'Putin Vows to 'Actively Defend' Russians Living Abroad' (NATOSource, 02 July 2004) <https://www.atlanticcouncil.org/blogs/natosource/putin-vows-to-actively-defend-russians-living-abroad/> accessed 27 December 2022; Meg Sullivan, 'Justifying Crimea: President Putin Invokes R2P' (BPR, 11 April 2014) <https://brownpoliticalreview.org/2014/04/justifying-crimea-president-putin-invokes-r2p/> accessed 27 December 2022

[15] Steven Lee Myers, 'TU.S. Rebukes Russia for Claims of Secret Bioweapons in Ukraine' (The New York Times, 13 September 2022) <https://www.nytimes.com/2022/09/13/technology/russia-ukraine-bioweapons.html> accessed 27 December 2022

[16] SECURITY COUNCIL, 'United Nations Not Aware of Any Biological Weapons Programmes, Disarmament Chief Affirms as Security Council Meets to Address Related Concerns in Ukraine' (United Nations, 11 March 2022) <https://press.un.org/en/2022/sc14827.doc.htm> accessed 27 December 2022

[17] Sarah Morris, 'Disinformation, Propaganda, and the War in Ukraine' (The Carter Center, 21 March 2022) <https://www.cartercenter.org/news/features/blogs/2022/disinformation-propaganda-and-war-in-ukraine.html> accessed 27 December 2022

[18] Tina Tvauri, 'New Wave of Kremlin Disinformation Regarding the Biosafety Cooperation Between the Us and Ukraine' (MYTH DETECTOR, 23 September 2022) <https://mythdetector.ge/en/new-wave-of-kremlin-disinformation-regarding-the-biosafety-cooperation-between-the-us-and-ukraine/> accessed 27 December 2022

[19] Joshua Keating, 'Why it's so hard to know how many Russian soldiers have been killed in Ukraine' (Grid News, 16 August 2022) <https://www.grid.news/story/global/2022/08/16/how-many-russian-soldiers-have-been-killed-in-ukraine-what-we-know-how-we-know-it-and-what-it-really-means/> accessed 27 December 2022

[20] DisinfoChronicle, 'Kremlin disinformation about the military offensive in Ukraine' (Детектор медіа, 16 may 2022) <https://detector.media/propahanda_vplyvy/article/196936/2022-02-25-disinfochronicle-kremlin-disinformation-about-the-military-offensive-in-ukraine/> accessed 27 December 2022

atrocities in Bucha, Russian authorities distributed the information about the artificial character of images taken there, depicting it as manipulation from the Ukrainian side.[21] Disinformation frequently concerns other States' support of Ukraine, *e.g.* Russian media repeatedly dwelled upon the US invasion of Mexico in the wake of the Russian invasion of Ukraine to counterbalance the powers and political impact.[22] The expected result implied an absence of solidarity with Ukrainian people from the Mexican side.

- **Lawfare** (manipulation of legal concepts to justify or misqualify the State's actions). Following the full-scale invasion of 24 February 2022, Russia also resorted to the concept of responsibility to protect, trying to justify the unlawful use of force by an international obligation to prevent genocide allegedly committed by Ukraine in its own territory. Scholars view the application of the 'responsibility to protect' notion as an attempt to frame illegal activities within the PIL and IHL standards, escaping or, at least, softening the potential liability.[23] In fact, such actions are aimed mainly at justification of Russian actions in the eyes of its own population.[24] Yet, international allies of Russia are heavily relying on such manipulations of legal standards to avoid accusations of complicity or assistance to the aggressor-State. For example, Chinese media called the invasion a 'special military operation', referring to the qualification of aggression developed by Russia at the very start of the full-scale invasion.[25] Also, Georgia for a long time abstains from joining the sanction process, referring, apart from economic reasons, to the fact that Ukraine itself *"failed to avoid war"*,[26] which could be reasoned by the mixed perception of the situation given the massive disinformation campaigns and previous Russian military activities in Georgia.

The response to Russian disinformation campaigns, in turn, was multi-dimensional: starting with statements condemning aggressive information operations and *"manipulative behaviour"*,[27] and ending with the sanctions on the actors in cultural and media spheres. Nowadays, various powerful media literacy campaigns both inside Ukraine and worldwide engine the debunking of the Russian false narratives.

[21] Guy Faulconbridge, 'Kremlin says Bucha is 'monstrous forgery' aimed at smearing Russia' (The Reuters, 4 April 2022) <https://www.reuters.com/world/europe/putin-ally-says-bucha-killings-are-fake-propaganda-2022-04-05/> accessed 27 December 2022

[22] Digital Forensic Research Lab, 'Russian Hybrid War Report: Russia retaliates against anti-war celebrities as social platforms crack down on Russian media' (Atlantic Council, 28 February 2022) <https://www.atlanticcouncil.org/blogs/new-atlanticist/russian-hybrid-war-report-russia-retaliates-against-anti-war-celebrities-as-social-platforms-crack-down-on-russian-media/> accessed 27 December 2022

[23] Betcy Jose and Christoph H Stefes, 'Russian Norm Entrepreneurship in Crimea: Serious Contestation or Cheap Talk?' (2018) 311 GIGA WP, 10-11

[24] John Reid, 'Putin, Pretext, and the Dark Side of the "Responsibility To Protect"' (War on the Rocks, 27 May 2022) <https://warontherocks.com/2022/05/putin-pretext-and-the-dark-side-of-the-responsibility-to-protect/> accessed 27 December 2022

[25] Macarena Vidal Liy, 'Russia's propaganda campaign around the war in Ukraine reaches China' (EL PAIS, 14 March 2022) <https://english.elpais.com/international/2022-03-14/russias-propaganda-campaign-around-the-war-in-ukraine-reaches-china.html> accessed 27 December 2022

[26] 'Georgian PM: Ukraine failed to avoid war, Georgia will not join anti-Russian sanctions' (JAMnews, 25 February 2022) <https://jam-news.net/georgian-pm-ukraine-failed-to-avoid-war-georgia-will-not-join-anti-russian-sanctions/> accessed 27 December 2022

[27] G7 Foreign Ministers, 'Statement on Russia's War against Ukraine' (G7 Information Centre, 14 May 2022) <http://www.g7.utoronto.ca/foreign/220514-ukraine.html> accessed 27 December 2022

For example, many Ukrainian NGOs,[28] foreign embassies,[29] and the EU campaigns[30] have worked on rebutting Russian myths regarding the alleged genocide committed by Ukrainian authorities, chemical weapons, alleged maltreatment of Russian prisoners of war,[31] NATO impact over Ukraine and other outstanding issues.

However, debunking myths and delivering public statements are only one side of the response, predominantly a non-legal one. At the same time, legal responsibility on various levels is a necessary reaction not only to effectively combat such behaviour within the war in Ukraine but also to prevent similar activities in the future. Respectively, all types of disinformation delivered by Russian propagandists throughout the mentioned time-frames shall be reviewed under the applicable legal framework of the IHRL, PIL, and IHL with a focus on the potential legal responses to such activities and mechanisms for serving justice on international and domestic levels.

I. Disinformation and international human rights law: does a right to lie exist?

The notion of disinformation under the IHRL. Disinformation, deception and lies are well-known tactics employed for hundreds of years. At the same time, the legal concept of disinformation is rather a novel one and does not have a precise definition. It has not even been used consistently: as Baade rightfully notes, the EU prefers the term 'disinformation' over 'fake news'; some scholars prefer 'misinformation'.[32] This lacuna results in a lack of proper understanding of how to deal with this type of content under the current framework of the IHRL.

Amongst the international authorities dealing with disinformation, three have attempted to define the notion. The earliest attempt can be found in the Joint Declaration on freedom of expression and 'fake news', disinformation and propaganda issued by the Special Rapporteurs in 2017.[33] This definition looks rather insignificant and is almost hidden along the wording of the general prohibition for States and State

[28] DisinfoChronicle, 'Kremlin disinformation about the military offensive in Ukraine' (Детектор медіа, 16 may 2022) <https://detector.media/propahanda_vplyvy/article/196936/2022-02-25-disinfochronicle-kremlin-disinformation-about-the-military-offensive-in-ukraine/> accessed 27 December 2022

[29] 'Disinformation About Russia's invasion of Ukraine - Debunking Seven Myths spread by Russia' (Embassy and Consulates of Belgium in China, 18 March 2022) <https://china.diplomatie.belgium.be/en/news/disinformation-about-russias-invasion-ukraine-debunking-seven-myths-spread-russia> accessed 27 December 2022

[30] 'Disinformation About Russia's invasion of Ukraine - Debunking Seven Myths spread by Russia' (Delegation of the European Union to the People's Republic of China, 18 March 2022) <https://www.eeas.europa.eu/delegations/china/disinformation-about-russias-invasion-ukraine-debunking-seven-myths-spread-russia_en?s=166> accessed 27 December 2022

[31] 'Moscow Scrupulously Observes Geneva Conventions on Treatment of POWs' (EUvsDisinfo, 4 August 2022) <https://euvsdisinfo.eu/report/moscow-scrupulously-observes-geneva-conventions-on-treatment-of-pows> accessed 27 December 2022

[32] Björnstjern Baade, 'Don't Call a Spade a Shovel: Crucial Subtleties in the Definition of Fake News and Disinformation'' (VerfBlog, 14 April 2020) <https://verfassungsblog.de/dont-call-a-spade-a-shovel/> accessed 24 December 2022

[33] UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and ACHPR Special Rapporteur on Freedom of Expression and Access to Information, International Mechanisms for Promoting Freedom of Expression, Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda (3 March 2017)

actors to disseminate and sponsor the dissemination of such information. Therein, disinformation is defined as statements known or reasonably known to be false.[34] While being the first attempt to define the object of regulation, the notion misses on the fact that such type of information is being widely distributed by private actors who may frequently disseminate them unintentionally, simply lacking verification skills or the desire to check everything they shared on their social media.

A more nuanced approach was chosen by the EU European Commission during the drafting of the Code of Practice on Disinformation in 2018.[35] This self-regulatory (but essentially co-regulatory) effort imposed certain transparency obligations on online intermediaries, whose platforms became Petri dishes for spreading disinformation. The notion of disinformation therein targetedly excluded misleading advertising, reporting errors, satire and parody, or clearly identified partisan news and commentary to avoid excessive interpretation and protect legitimate types of expression, such as artistic and political speech. However, it still was designed in a relatively broad manner.

> Disinformation encompasses the **verifiably false or misleading information which is created, presented and disseminated for economic gain or to intentionally deceive the public and may cause harm** in the form of threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security.

In 2022, the EU Strengthened Code of Practice on Disinformation[36] expanded the definition even further, using it as an umbrella term for these four separate activities:

- **misinformation** – false or misleading content shared without harmful intent though the effects can still be harmful, *e.g.* when people share false information with friends and family in good faith;
- **disinformation** – false or misleading content that is spread with an intention to deceive or secure economic or political gain and which may cause public harm;
- **information influence operation** – coordinated efforts by either domestic or foreign actors to influence a target audience using a range of deceptive means, including suppressing independent information sources in combination with disinformation;
- **foreign interference in the information space** – coercive and deceptive efforts to disrupt the free formation and expression of individuals' political will by a foreign state actor or its agents, often carried out as part of a broader hybrid operation.

[34] Ibid, Principle 2a

[35] EU Code of Practice on Disinformation (September 2018) <https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation> accessed 25 December 2022

[36] EU Strengthened Code of Practice on Disinformation (16 June 2022) <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation> accessed 25 December 2022

Understanding the desire to simplify the terminology in the document, predominantly addressed to the very large online platforms, we consider the last two elements being covered by the rules on sovereignty and non-interference and will not discuss them at length in this section. The second element of the definition mirrors the Strengthened Code predecessor's provisions, while the first covers unintentionally disseminated speech – a distinction Special Rapporteurs failed to make in 2017. The Strengthened Code, thus, follows the approach of Wardle and Derakhshan's 2018 CoE report on information disorder.[37]

The third authority, the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan, devoted her 2021 annual report to the topic. She also noted the lack of clarity and agreement on what constitutes disinformation, including the frequent and interchangeable use of the term misinformation and the influence of this conundrum on the effectiveness of the legal response to the issues. She adopted an approach similar to Wardle and Derakhshan and the Strengthened Code, adopted after the report's publication. Thus, she agreed that the distinction between dis- and misinformation should be drawn on the basis of the existence of a specific intent to deceive the public and cause serious harm.[38] Irene Khan also underlined that some forms of disinformation can amount to incitement to hatred, discrimination and violence, which are prohibited under international law.[39]

> Accordingly, disinformation shall be understood as **false information disseminated intentionally to cause serious social harm to legitimate public interests**, such as democratic political processes, national security, public order, etc.

How does international human rights law treat false information and does it allow to punish those spreading it? Rather favourably, while untrue materials are considered as deserving some degree of protection from penal sanctions. The Joint Declaration insists that the general prohibition on disseminating false news or non-objective information is incompatible with international standards on freedom of expression and should be abolished.[40] Irene Khan views criminal sanctions as a last resort applicable solely to the very exceptional and most egregious circumstances of incitement to violence, hatred or discrimination and, akin to the Joint Declaration drafters, underlines the prohibition for the States to make, sponsor, encourage

[37] Claire Wardle, Hossein Derakhshan, Information Disorder: Toward an interdisciplinary framework for research and policy-making (CoE, 2017) 20 <https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77> accessed 22 December 2022

[38] Irene Khan, Disinformation and freedom of opinion and expression: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UNHRC A/HRC/47/25, 13 April 2021, para 15

[39] Ibid, para 10

[40] UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression & ACHPR Special Rapporteur on Freedom of Expression and Access to Information, International Mechanisms for Promoting Freedom of Expression, Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda (3 March 2017), Principle 2a

or disseminate disinformation.[41] Even the ECtHR, in its *Salov v Ukraine* judgment, stated that freedom of expression does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not be truthful.[42] The judgement concerned a very limited number of newspaper articles, also failing to address the intent – a crucial element of disinformation. Moreover, as mentioned before, disinformation might well serve as a constituent element of other prohibited types of speech.

Thus, to explore the possibilities of holding disinformation disseminators accountable, a direct prohibition on false information cannot be employed, but rather alternative options should be suggested. In this regard, the mentioned Joint Declaration uses the term disinformation along with propaganda.[43] The UN Special Rapporteur hints at the blurred lines between disinformation and other types of harmful speech clearly prohibited under the ICCPR, such as hate speech and propaganda for war.[44] It is time to have a closer look at Article 20 of the ICCPR.

ICCPR. The ICCPR contains two provisions on freedom of expression's limitation: Articles 19(3) and 20. While the first provision is a classic three-part-test clause (the restriction has to be provided by law, pursue a legitimate aim and be necessary), the second one delineates specific types of prohibited content: propaganda for war (Article 20(1)) and advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (Article 20(2)).[45]

Hate speech. There exists a rather clear understanding that disinformation may lead to violent consequences and further polarisation in society. Thus, judging by the consequences of harmful speech, one of the ways to restrict disinformation in line with international law is to implement hate speech legislation.

The ICCPR prohibits the advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. While some of the terms used are not entirely clear to understand, the UN High Commissioner on Human Rights and the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression[46] endorsed the Camden Principles on Freedom

[41] Irene Khan, Disinformation and freedom of opinion and expression: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UNHRC A/HRC/47/25, 13 April 2021, paras 88-89

[42] Salov v Ukraine, App no 65518/01(ECtHR, 6 September 2005), para 113

[43] UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression & ACHPR Special Rapporteur on Freedom of Expression and Access to Information, International Mechanisms for Promoting Freedom of Expression, Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda (3 March 2017), Principle 2a

[44] Irene Khan, Disinformation and freedom of opinion and expression: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UNHRC, A/HRC/47/25, 13 April 2021, para 10

[45] International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 ["ICCPR"], Article 20

[46] Frank La Rue, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UNHRC A/67/357, 7 September 2012, para 44

of Expression and Equality[47] which suggests the following definitions of 'advocacy' and 'incitement':

- **advocacy** is to be understood as requiring an intention to promote hatred publicly towards the target group;
- **incitement** refers to statements about national, racial or religious groups which create an imminent risk of discrimination, hostility or violence against persons belonging to those groups.[48]

The Rabat Plan of Action also suggests a set of criteria to assess the severity of hatred when adopting judicial decisions on whether a certain statement should be qualified as hate speech. These criteria are needed to determine whether, under the UN HRC guidelines, the prohibition of particular speech addressed in Article 20 is compliant with the three-part test requirements of Article 19(3) of the ICCPR,[49] and include:

- **context**, the analysis of which should place the speech act within the social and political context prevalent at the time the speech was disseminated;
- **status of the speaker** in society;
- **intent**, since negligence and recklessness are not sufficient for an act to be an offence;
- **content and form of the speech**, the analysis of which shall include the degree to which the speech was provocative and direct, as well as the form, style, nature of arguments deployed in the speech or the balance struck between arguments deployed;
- **the extent of the speech,** which such elements as the reach of the speech, its public nature, its magnitude and size of its audience, the means of dissemination used, the frequency, the quantity and the extent of the communications, etc.;
- **the likelihood, including imminence**, as the courts will have to determine that there was a reasonable probability that the speech would succeed in inciting actual action against the target group, recognising that such causation should be rather direct.[50]

The UN HRC lacks consistency in interpreting Article 20(1) of the ICCPR. In *JRT and WG Party v Canada*, the applicant's use of pre-recorded calls to "*warn the callers of the dangers of international finance and international Jewry leading the world into wars, unemployment and inflation and the collapse of world values and principles*" was considered by the HRC as unprotected by virtue of Canada's obligations under Article 20(1) of the Covenant.[51] At the same time, in *Malcolm Ross v Canada*, the UN HRC decided to analyse the case under Article 19(3) and the three-part test

[47] UNHRC, 'Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence' (2012), para 21 <http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf> accessed 25 December 2022

[48] ARTICLE 19. The Camden Principles on Freedom of Expression and Equality, para 12.1 <https://www.article19.org/resources/camden-principles-freedom-expression-equality/> accessed 27 December 2022

[49] General comment no 34, Article 19, Freedoms of opinion and expression (12 September 2011) UN Doc CCPR/C/GC/34, para 50

[50] UNHRC, 'Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence' (2012), para 29 <http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf> accessed 25 December 2022

[51] JRT and the WG Party v Canada Communication no 104/1981 UN Doc CCPR/C/OP/2 (1984)

despite the fact that it also involved the applicant's punishment for the dissemination of controversial views on Judaism.[52] It also did not subscribe to France's argument that the denial of gas chambers' existence should not be protected by its Article 20(1) obligations in *Faurisson v France*, although the UN HRC still decided in favour of the State.[53]

Propaganda for war. Propaganda for war remained overlooked in the disinformation debate for a long time. For instance, it was never mentioned along with the definition of disinformation in the research preceding the Russian full-scale invasion: Joint Declaration mentioned 'propaganda', whereas the UN Special Rapporteur in her 2021 report on disinformation mentioned that States *"should refrain from restricting freedom of expression online or offline except in accordance with the requirements of Articles 19(3) and 20(2) of the ICCPR, strictly and narrowly construed"*, somehow missing Article 20(1) of the ICCPR and generally mentioned the word 'propaganda' only once.[54] Only in her 2022 report Irene Khan had finally paid more attention to Article 20(1) of the ICCPR, nonetheless shifting the burden of defining the concept on the Office of the United Nations High Commissioner for Human Rights.[55]

What is the definition of 'propaganda for war', and how shall it be distinguished from legitimate propaganda? The UN HRC devoted its General Comment No 11 to the topic, though it leaves more questions than provides answers. It consists, however, of merely two paragraphs and indicates the following features of this notion. Such propaganda (a term left undefined by the HRC) should:

- threaten or result in the act of aggression or breach of the peace contrary to the UN Charter;
- not cover advocacy of the sovereign right of self-defence or the right of peoples to self-determination and independence in accordance with the UN Charter;
- be prohibited by internal legislation of Member States, which contains appropriate sanctions for its violation.[56]

Another body of the UN HRC work additionally indicates that the provisions of Article 20 are, in fact, non-derogable: a State party cannot engage itself in propaganda for war or in advocacy of national, racial or religious hatred that would constitute an incitement to discrimination, hostility or violence in times of emergency.[57] The UN HRC also noted that the acts addressed in Article 20 are all subject to restrictions pursuant to Article 19(3) – that is, the three-part test.[58] At the same time, commentators note that Article 20's relation to Article 19 is one related to Article 5 of the Covenant:

[52] Ross v Canada, Communication no 736/1997 UN Doc CCPR/C/70/D/736/1997 (2000)
[53] Faurisson v France, Communication no 550/1993 UN Doc CCPR/C/58/D/550/1993 (1996)
[54] Irene Khan, Disinformation and freedom of opinion and expression: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UNHRC A/HRC/47/25, 13 April 2021, para 88
[55] Irene Khan, Disinformation and freedom of opinion and expression during armed conflicts: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UNHRC A/77/288, 12 August 2022, para 105
[56] UNHRC, General Comment 11, Article 20 (19th session, 1983), UN Doc HRI/GEN/1/Rev.1 (1994), para 2
[57] UNHRC, General Comment 29, States of Emergency (Article 4), UN Doc CCPR/C/21/Rev.1/Add.11 (2001), para 13(e)
[58] UNHRC, General comment no 34, Article 19, Freedoms of opinion and expression, UN Doc CCPR/C/GC/34 (2011), para 50

Article 19 shall not protect propaganda for war as it constitutes an abuse of the right to freedom of expression.[59] Unfortunately, Article 20(1) has not been dealt with under the optional protocol, and no State party has been rebuked by the UN HRC for engaging in propaganda for war as of today.[60]

Such inattention by the chief interpretative body of the ICCPR can be linked to the complex negotiations surrounding the inclusion of Article 20 to the Covenant. This provision was predominantly advocated by the Third World countries.[61] Most Western democracies did not vote for the provision. They enacted the reservations to Article 20, specifically reserving the right not to enforce the prohibition on propaganda for war in their legislation.[62] One of the reasons was the fear of this prohibition's abuse by, for instance, the USSR, targeting its application at practically anything disseminated by the US press.[63]

If one looks at the term, it consists of two terms to be defined: 'propaganda' and 'war'. Both are called by McGonagle *"definitionally problematic terms"*,[64] the second meriting more suggestions from the international bodies. At the same time, it is also observed that the meaning of 'propaganda for war' is only as imprecise as states wish it to be.[65]

As to the notion of propaganda, the Joint Declaration defines it as statements which demonstrate a reckless disregard for verifiable information.[66] The UN Special Rapporteur, in her 2022 report, preferred to abstain from defining the term, underlining again that ordinary propaganda shall be protected,[67] though she indicated that it *"has a pejorative sense of disseminating information that may be true or false but is biased, partial, misleading and emotive"*.[68]

The UN General Assembly, in its earlier resolutions, elaborated on the overlapping concept of 'propaganda against peace', which includes not just incitement to conflicts or acts of aggression but also measures tending to isolate the peoples from any contact with the outside world, by preventing the press, radio and other media of communication from reporting international events, and thus hindering

[59] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) 17 <https://www.osce.org/fom/203926> accessed 27 December 2022

[60] Michael G Kearney, The Prohibition of Propaganda for War in International Law (OUP, 15 November 2007) ["Kearney"] 172

[61] Manfred Nowak, UN Covenant on Civil and Political Rights: CCPR Commentary (NP Engel, 2005) ["Nowak"] 470-471

[62] Nowak 478

[63] Kearney 100

[64] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) 31 <https://www.osce.org/fom/203926> accessed 27 December 2022

[65] Kearney 189

[66] UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression & ACHPR Special Rapporteur on Freedom of Expression and Access to Information, International Mechanisms for Promoting Freedom of Expression, Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda (3 March 2017), Principle 2a

[67] Irene Khan, Disinformation and freedom of opinion and expression during armed conflicts: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UNHRC A/77/288, 12 August 2022, para 39

[68] Ibid, para 11

mutual comprehension and understanding between peoples.[69] However, this norm is not the current law in this regard.

In his commentary to the ICCPR, Nowak suggests that propaganda shall be understood as *"intentional, well-aimed influencing of individuals by employing various channels of communication to disseminate, above all, incorrect or exaggerated allegations of fact. Also included ... are negative or simplistic value judgements whose intensity is at least comparable to that of provocation, instigation, or incitement"*.[70] He adds on the important element of intent, which has to create or reinforce the willingness to go to war.[71] These elements of the definition shall be considered in any further and more detailed interpretation of Article 20(1) of the ICCPR.

The UN HRC elaborated more on the notion of war, essentially excluding from the prohibition's scope the advocacy of the rights of self-determination and self-defence under the UN Charter.[72] Joseph and Castan argue that propaganda for wars sponsored by the UN Security Council under Chapter VII of the UN Charter, such as the Allied action against Iraq in defence of Kuwait in 1990, shall also fall outside the prohibition's scope.[73] Nowak adds that the prohibition was not meant to cover civil wars insofar as they do not reach the level of an international armed conflict, as well as the wars of liberation.[74] In Kearney's opinion, these other forms of propaganda inciting manifestations of violence such as civil war, rebellion against the government, or violence against the person were either prohibited under Article 20(2) or were grounds for permissible limitations on the right to freedom of expression in Article 19(3).[75]

Another question left undiscussed until recently is whether propaganda for war shall cover statements preceding the war or propaganda during the war. In 2022, the UN Special Rapporteur Irene Khan submitted that the prohibition should be strictly limited to incitement of war and not to propaganda during the war.[76] At the same time, the CJEU, interpreting Article 20(1) of the ICCPR in a case concerning the imposition of sanctions on Sputnik, opined that it *"includes not only incitement to a future war, but also continuous, repeated and concerted statements in favour of an ongoing war contrary to international law, in particular, if the statements emanate from a media controlled, directly or indirectly, by the aggressor State"*.[77] This interpretation is more persuasive as the propaganda for the war's continuation actually has the

[69] UN General Assembly Resolution 381(V) "Condemnation of propaganda against peace", UN Doc A/RES/381(V), 17 November 1950

[70] Nowak 472

[71] Nowak 473

[72] UNHRC, General Comment 11, Article 20 (19th session, 1983), UN Doc HRI/GEN/1/Rev.1 (1994), para 2

[73] Sarah Joseph and Melissa Castan, The International Covenant on Civil and Political Rights (3rd Edition): Cases, Materials, and Commentary (OUP, 2013) 672

[74] Nowak 473

[75] Kearney 148-149

[76] Irene Khan, Disinformation and freedom of opinion and expression during armed conflicts: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UNHRC A/77/288, 12 August 2022, para 39

[77] Case T-125/22 RT France v Council [2022] ECLI:EU:T:2022:483, para 211

intent of reinforcing the willingness of the aggressor State's population to continue the war of aggression.

The last important question is the nature of the obligations of states under Article 20(1) of the ICCPR. It follows from the wording of the provision itself that propaganda for war shall be prohibited by law. This law should most likely be criminal law,[78] though the final decision lies on the States, possessing certain discretion on the matter.[79] Despite that, the States should enforce this prohibition directly, not only in case the threat of war arises.[80] States are also obliged to withstand from engaging in official state propaganda and to prohibit any propaganda for war by private persons or semi-State media, thus enacting legislation prohibiting propaganda for war by state officials as well.[81]

CERD. An additional document on the UN level that deals with hate speech is the CERD. Under its Article 4, which was regarded as central to the struggle against racial discrimination,[82] State Parties shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin.[83] However, such a restriction should pay due regard to the principles embodied in the Universal Declaration of Human Rights – that is, the three-part test. In *Jewish Community of Oslo v Norway*, the CERD Committee found against Norway on Article 4 since it acquitted the leader of the Neo-Nazi organisation for the remarks such as "*people and country are being plundered and destroyed by Jews, who suck our country empty of wealth and replace it with immoral and un-Norwegian thoughts*" made during the public event.[84]

European human rights law. Up to date, three organisations centring around the European continent have adopted approaches to dealing with hate speech and propaganda for war: the CoE, the OSCE, and the EU. The CoE has the most developed framework centred on the interpretation of the ECHR.

CoE. Indeed, the ECHR does not have any separate provisions prohibiting propaganda for war in a fashion similar to Article 20(1) of the ICCPR. The limitations of freedom of expression are, thus, designed similarly to the provisions of Article 19(3) of the Covenant. However, the ECHR contains Article 17 – a norm prohibiting the abuse of rights, that is, the right to engage in any activity or perform any act

[78] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) 17 <https://www.osce.org/fom/203926> accessed 27 December 2022
[79] Nowak 474
[80] Ibid
[81] Kearney 142
[82] UN CERD Committee, General recommendation No 35: Combating racist hate speech, 26 September 2013, UN Doc CERD/C/GC/35, para 10
[83] CERD (21 December 1965) 660 UNTS 195 ["CERD"], Aricle 4
[84] Jewish Community of Oslo and Others v Norway, Merits, Communication No 30/2003, UN Doc CERD/C/67/D/30/2003

aimed at destroying any of the rights and freedoms outlined in the Convention. Its reach extends to any State, group and person.[85] This provision was also extensively interpreted by the ECtHR: probably, the most important and influential interpretative body in the machinery of human rights' scope interpretation around the world.

Hate speech. The practice of the ECtHR on hate speech remains quite extensive, although not addressing directly hateful messages based on false information. Notably, hate speech is explicitly outlawed even when it is deprived of the direct calls for violence,[86] since illegal incitements cover indirect attacks on identifiable groups.[87]

The examples of unlawful speech, in this respect, cover both harsh and, at first glance, relatively mild expressions, *e.g. "the day there are no longer 5 million but 25 million Muslims in France, they will be in charge"*,[88] *"Islam out of Britain"* [89] or *"Surinamers, Turks are not at all needed here"*,[90] *"they will start to burn, slaughter rape, rob and enslave"*,[91] and *"brazenness of this demonstrable Gypsy banditry"*.[92] The approach of the ECtHR is emblematic in recognising as impermissible the statements that may raise disgust,[93] negative stereotyping[94] or assimilate a protected group with the criminals.[95] Importantly, the ECtHR made no distinction with regard to the nature of the expression – false or based on actual facts. For instance, when the assimilation of the entire social group with criminals took place, apparently, no scientific or research data was available to prove that all individuals in that group committed criminal offences. Accordingly, the statement fell within the generally accepted notion of disinformation (although not labelled by the ECtHR as such), being false and intended to harm a protected interest.

Therefore, a conclusion can be made that the ECtHR has at least implicitly dealt with the cases of using false information for spreading hate, while in most cases, the Court recognised the incompatibility of such speech with freedom of expression. In some instances, the gravity of the expression has even reached Article 17, with the Court labelling it as an abuse of right. In this respect, many expressions distributed by Russians can be easily qualified as either raising disgust towards Ukrainians or negatively stereotyping them, which is all based on false or manipulated data. Thus, even if untrue information itself is not explicitly outlawed in the ECHR system,

[85] European Convention on Human Rights (adopted 4 November 1950, entered into force 3 September 1953) 213 UNTS 1932 ["ECHR"], Article 17
[86] Dmitriyevskiy v Russia, App no 42168/06 (ECtHR, 3 October 2017), para 99
[87] Pavel Ivanov v Russia, App no 35222/04 (ECtHR, 20 February 2007); Dmitriyevskiy v Russia, App no 42168/06 (ECtHR, 3 October 2017), para 100
[88] Le Pen v France, App no 18788/09 (ECtHR, 20 April 2010)
[89] Norwood v the United Kingdom, App no 23131/03 (ECtHR, 16 November 2004)
[90] Glimmerveen and Haqenbeek v the Netherlands, App nos 8348/78 and 8406/78 (EComHR, 11 October 1979)
[91] Atamanchuk v Russia, App no 4439/11 (ECtHR, 11 February 2020), paras 8, 62, 70
[92] Budinova and Chaprazov v Bulgaria, App no 12567/13 (ECtHR, 16 February 2021), para 65
[93] Savva Terentyev v Russia, App no 10692/09 (ECtHR, 28 August 2018), para 76
[94] Ibragim Ibragimov and Others v Russia, App nos 1413/08 and 28621/11 (ECtHR, 28 August 2018), para 94
[95] Budinova and Chaprazov v Bulgaria, App no 12567/13 (ECtHR, 16 February 2021), para 65

disinformation is *de facto* prohibited when it serves as a part of hate speech campaigns.

Propaganda for war. While not mentioning the concept of propaganda for war directly, the ECtHR had declared advocacy for jihad – a holy war in Islam – incompatible with the Convention's values under Article 17. In *Hizb Ut-Tahrir and Others v Germany*, an unincorporated association positioning itself as a global Islamic political party was banned in Germany. It agitated in a targeted fashion against the Islamic States and the governments and called for their overthrow, *inter alia*, by active Jihad. It also denied the right to exist to the State of Israel.[96] The Court reminded of its standards vis-à-vis stating that the general purpose of Article 17 is to prevent individuals or groups with totalitarian aims from exploiting in their own interests the principles enunciated in the Convention.[97] By analysing the brochures and leaflets of the organisation, the Court observed the repeated calls for the violent destruction of Israel and the banishment and killing of its inhabitants.[98] Therefore, the application was found inadmissible *ratione materiae* since the applicants employed their rights against the values of the Convention, such as the commitment to the peaceful settlement of international conflicts and the sanctity of human life.[99]

Another case where the ECtHR analysed the notion of propaganda was *Roj TV A/S v Denmark*. Therein, a broadcaster was deprived of its license by the Danish courts for the propaganda of PKK – the Kurdistan Workers' Party, classified previously as a terrorist organisation. The Strasbourg Court agreed with the national judiciary that the one-sided coverage with repetitive incitement to participate in fights and actions, incitement to join the organisation, and the portrayal of deceased guerrilla members as heroes, broadcasted by Roj TV, amounted to propaganda for the terrorist organisation, exceeding the mere declaration of sympathy.[100] Thus, the incitement to violence and support for terrorist activity were found to run contrary to Article 17 of the ECHR.

There is also a line of case law concerning the so-called 'separatist propaganda' legislation in Turkey, which was also predominantly applied in the Kurdish-Turkish conflict context. When the Court took up cases on that matter, it analysed them under the lens of the three-part-test, *de facto* sharing the approach of the HRC in distinguishing 'clearly unlawful' and other types of harmful speech. An illustrative example is *Sürek v Turkey (No 1)*, where the applicant, an owner of a newspaper, was penalised by a fine for allowing the publication of two reader's letters which condemned the military actions of the authorities in south-east Turkey and accused them of brutal suppression of the Kurdish people in their struggle for independence and freedom.[101]

[96] Hizb-Ut Tahrir and Others v Germany (dec), App no 31098/08 (ECtHR, 12 June 2012), para 6
[97] Ibid, para 72
[98] Ibid, para 73
[99] Ibid, paras 74-75
[100] Roj TV A/S v Denmark (dec), App no 24683/14 (ECtHR, 17 April 2018), para 46
[101] Sürek v Turkey (No 1), App no 26682/95 (ECtHR, 8 July 1999), paras 11, 60

The ECtHR called for the textual and contextual analysis of these letters to establish whether their publication by the applicant was protected under Article 10 of the Convention.[102] In the view of the Court, the impugned letters amount to an appeal to bloody revenge by stirring up base emotions and hardening already embedded prejudices which have manifested themselves in deadly violence.[103] In the context of the clashes in Kurdistan, which continued since 1985, these letters must be seen as capable of inciting further violence in the region by instilling a deep-seated and irrational hatred against those depicted as responsible for the alleged atrocities.[104] In light of the following, the ECtHR found Turkey's arguments persuasive and did not find a violation of the applicant's right to freedom of expression.

The OSCE and the EU. The other two pan-European organisations, the OSCE and the EU, deal with the prohibition of propaganda for war to a lesser extent. The Helsinki Final Act mentions the duty of participating states to abstain from propaganda for war, though this commitment was never described in great detail until the publication of the RFoM non-paper in 2015.[105] In the Joint Declaration, propaganda for war was also never mentioned.[106]

On the EU level, the Charter of Fundamental Rights, in line with the position of the European states when drafting the ICCPR and in line with the wording of ECHR, does not include propaganda for war or hate speech prohibitions. At the same time, it contains a provision on abuse of rights.[107] The EU Audiovisual Media Services Directive, though applicable only to a specific part of the media sector, prohibits the broadcasting of incitement to violence or hatred directed against a group of persons or a member of a group.[108] This notion was interpreted by the CJEU as one *"designed to forestall any ideology which fails to respect human values, in particular initiatives which attempt to justify violence by terrorist acts against a particular group of persons"*.[109]

As to the notion of propaganda for war, in the recent CJEU ruling in *RT France v Council*, the CJEU noted that *"international instruments for the protection of human rights to which the Member States had cooperated or acceded could also provide guidance which had to be taken into account in the context of EU law"*.[110]

[102] Ibid, para 62
[103] Ibid, para 62
[104] Ibid, para 62
[105] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) <https://www.osce.org/fom/203926> accessed 27 December 2022
[106] UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and ACHPR Special Rapporteur on Freedom of Expression and Access to Information, International Mechanisms for Promoting Freedom of Expression, Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda (3 March 2017)
[107] Charter of Fundamental Rights of the European Union [2012] OJ C 326/391, Article 54
[108] Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) (codified version) [2018] 02010L0013-20181218, Article 6(1)(a)
[109] Cases C-244/10 and 245/10 Mesopotamia Broadcast A/S METV and Roj TV A/S v Bundesrepublik Deutschland [2010] 2011 I-08777, para 42
[110] Case T-125/22 RT France v Council [2022] ECLI:EU:T:2022:483, para 207

This opened the avenue for the Court to interpret Article 20(1) of the ICCPR as a part of *acquis communutaire*. As noted above, the CJEU's interpretation of this prohibition is consistent with the HRC and scholarly position, with the only significant distinction being the recognition that propaganda for war covers *"also continuous, repeated and concerted statements in favour of an ongoing war contrary to international law"*.[111]

The notion of propaganda was separately mentioned in another ruling of the CJEU concerning Lithuania's ban on NTV Mir broadcasting. Therein, the Court recognised the actions of Lithuania as legitimate and justifiable in principle since the program:

> *"... was addressed ... in a targeted manner to the Russian-speaking minority in Lithuania and aimed, by the use of various propaganda techniques, to influence negatively and suggestively the opinion of that social group relating to the internal and external policies of the Republic of Lithuania, the Republic of Estonia and the Republic of Latvia, to accentuate the divisions and polarisation of society, and to emphasise the tension in the Eastern European region created by Western countries and the Russian Federation's role of the victim"*.[112]

Other regional human rights systems. Two more regional human rights protection instruments must be mentioned when assessing the prohibition of propaganda for war. The ACHR, in its Article 13(5), mentions that hate speech and propaganda for war shall constitute an offence punishable by law.[113] Similar to the UN HRC's lack of interpretation of this provision, neither the Inter-American Court nor the respective Commission clarified the scope of this provision.

The Inter-American Commission's Report on Terrorism and Human Rights stated that whilst propaganda for war is an offence punishable by law, legislation that broadly criminalises the public defence of terrorism or of persons who might have committed terrorist acts is incompatible with the right to freedom of expression if there is no incitement.[114] Similar stance is taken by the Declaration of Principles on Freedom of Expression issued by the OAS Special Rapporteur for Freedom of Expression.[115] Interestingly, the States did not make any reservations to Article 13(5), unlike under the ICCPR regime. In Kearney's words, the lack of this provision's interpretation remains a glaring void in the human rights discourse of the Inter-American human rights system.[116]

[111] Ibid, para 211
[112] Case C-622/17 Baltic Media Alliance Ltd v Lietuvos radijo ir televizijos komisija [2019] ECLI:EU:C:2019:566, para 79
[113] American Convention on Human Rights (adopted 22 November 1969, entered into force 18 July 1978) 1144 UNTS 123, Article 13(5)
[114] IAComHR, Report on Terrorism and Human Rights, OEA/Ser.L/V/II.116, Doc 5 rev 1 corr (2002), para 323
[115] OAS, Declaration of Principles on Freedom of Expression: Background and Interpretation of the Declaration of Principles (2000), para 27 <https://www.oas.org/en/iachr/expression/showarticle.asp?artID=132> accessed 27 December 2022
[116] Kearney 183

The African Charter on Human and People's Rights has no provisions similar to the ACHR. However, it contains a detailed provision on the duties of individuals. They include the duties:

- not to compromise the security of the State, whose national or resident he is;
- to preserve and strengthen social and national solidarity, particularly when the latter is strengthened;
- to preserve and strengthen the national independence and the territorial integrity of his country and to contribute to his defence in accordance with the law.[117]

In light of the provisions guaranteeing freedom of expression, such duties might be interpreted as imposing a duty to abstain from hate speech and propaganda for war. However, they were never interpreted in this manner by any human rights protection bodies in the African system. The Declaration of Principles on Freedom of Expression in Africa by the African Commission also remained silent on the matter.[118]

Liability for the IHRL violations. The development of the IHRL brought along the notion of State responsibility for human rights violations on an international scale. The individuals, who are always more vulnerable in front of the Leviathan-State, obtained the possibility to claim compensation from States. Human rights courts and institutions' interpretation of international conventions also contributed to filling the brief wording of the norms with content. Unfortunately, the possibilities for IHRL mechanisms to hold those responsible for propaganda for war and hate speech in Russia are scarce since IHRL primarily deals with the obligations of states with respect to individuals. Thus, apart from the interstate disputes, there might only be space to hold the States liable for breaches of their positive obligations to protect vulnerable groups from harmful content, directly mentioned by the ECtHR in *Aksu v Turkey*[119] and later *Behar and Gutman v Bulgaria*.[120]

Russia denounced the ECHR after it commenced the full-scale invasion and, therefore, excluded the ECtHR's jurisdiction over the violations it committed after 16 September 2022.[121] Moreover, it announced that it will not enforce any judgments of the Court delivered after 15 March 2022 and adopted the respective legislation.[122] Thus, the Strasbourg Court, the most influential and powerful of the human rights institutions, became unavailable both for inter-State complaints and individual complaints. This closed the most perspective avenue of litigation in the field of the IHRL.

[117] African Charter on Human and Peoples' Rights (adopted 27 June 1981, entered into force 21 October 1986) (1982) 21 ILM 58
[118] AComHPR, Declaration of Principles on Freedom of Expression in Africa (17-23 October 2002)
[119] Aksu v Turkey, App nos 4149/04 and 41029/04 (ECtHR, 15 March 2012), para 59
[120] Behar and Gutman v Bulgaria, App no 29335/13 (ECtHR, 16 February 2021), paras 98-101
[121] ECtHR, 'Russia ceases to be party to the European Convention on Human Rights' (16 September 2022) <https://www.coe.int/en/web/portal/-/russia-ceases-to-be-party-to-the-european-convention-on-human-rights> accessed 27 December 2022
[122] Госдума, 'Приняты законы о неисполнении Россией решений ЕСПЧ' (7 June 2022) <http://duma.gov.ru/news/54515/> accessed 27 December 2022

The UN HRC is one of the solutions that remain on the table, though its significance is doubtful. While Joseph and Castan rightfully note that though the HRC has not yet confirmed whether Article 20 is justiciable, nothing suggests the opposite. Russia accepted the competence of the UN HRC to receive and consider communications submitted by another State Party under Article 41 of the ICCPR.[123] It sets out a procedure for the resolution of disputes between States Parties over a State's fulfilment of its obligations under the ICCPR through the establishment of an ad hoc Conciliation Commission.[124] However, this Commission is only empowered to issue a non-binding report, which shall embody its findings on all questions of fact relevant to the issues between the States Parties concerned and its views on the possibilities of an amicable solution to the matter.[125] The States have the right to dismiss the reports issued under Article 42 of the ICCPR.

Russia has also ratified the Optional Protocol to the ICCPR, accepting the UN HRC jurisdiction over individual complaints from individuals whose rights were breached by the State.[126] To consider such complaints, the HRC has to be satisfied that the applicant has exhausted all the domestic remedies.[127] In practice, this would mean that someone who wishes to hold Russia accountable for human rights violations under the ICCPR has to at least attempt going to Russian courts – a highly unlikely avenue for any Ukrainian citizens or Russian citizens who might be repressed for submitting such claims. Unfortunately, even if a State or an individual receives a conclusion in its favour, there is no mechanism to force States to adopt the Commission's findings.

Similar procedures are prescribed by Articles 11-12 (regarding the CERD Committee's competence to review interstate claims) and 14 (regarding the individual complaint) of the CERD, both accepted by Russia.[128] Under the CERD, however, the states could go further to the ICJ if the mechanisms established by the Convention – the CERD Committee and the Conciliation Commission – proved to be insufficient to resolve the conflict between the parties.[129] This opens a potential opportunity for Ukraine to use these provisions against Russia and hold it liable for state-sponsored hate speech based on ethnic origin in the Hague Court.

If the pattern of human rights violations is consistent, individuals can also refer the situation to the UN Human Rights Council. These violations shall not be dealt with by a special procedure, a treaty body or other UN or similar regional complaints procedure in the field of human rights, and the applicants shall also exhaust all

[123] UNTC, ICCPR Ratification Status: <https://treaties.un.org/Pages/ViewDetails.aspx?chapter=4&clang=_en&mtdsg_no=IV-4&src=IND#EndDec> accessed 27 December 2022

[124] ICCPR, Articles 41-43

[125] ICCPR, Article 42

[126] UNHRTB, Ratification status for Russian Federation: <https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?CountryID=144&Lang=EN> accessed 27 December 2022

[127] Optional Protocol to the International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171, Article 2

[128] CERD, Articles 11-12, 14

[129] CERD, Article 22

the possible domestic remedies.[130] After reviewing the complaint, the UN Human Rights Council can, *inter alia*, appoint an independent and highly qualified expert to monitor the situation and report back to the Council (*de facto* create a Special Procedure) or recommend to OHCHR to provide technical cooperation, capacity building assistance or advisory services to the State concerned.[131] While this leeway beari mostly political nature, its use might be essential to keep track of violations committed during the Russian disinformation campaigns.

The IHRL legal framework for holding Russia liable for spreading disinformation may, thus, be summarised in the following manner:

I. Disinformation has no hard law regulation within the international human rights treaties, being addressed as a part of the propaganda for war prohibition, ban on hate speech, illegal incitements, or discriminatory speech.

II. The international and regional human rights bodies, including judicial and quasi-judicial ones, have repeatedly dealt with disinformation in practice, establishing lawfulness of restrictions when it constituted a part of the traditionally prohibited types of speech.

III. Since Russia has denounced the ECHR, the venue of inter-State application before the ECtHR is unavailable to establish the responsibility for violations of the Convention. Yet, the responsibility of Russia under the IHRL is possible via the invocation of:

- the ICCPR mechanisms through the reference to the UN HRC with a request for a review of the Russian conduct and its compliance with the ICCPR in the information sphere;
- the UN Human Rights Council complaints procedure by Ukrainian individuals and NGOs for monitoring of the situation and collection of evidence regarding the violations of human rights by Russia in the fields covered by the ICCPR, including its Article 20;
- the CERD mechanisms, which imply the application to the CERD Committee to recognize Russian disinformation as a form of discriminatory treatment.

[130] UN Human Rights Council, Institution-building of the United Nations Human Rights Council, UN Doc A/HRC/RES/5/1, 18 June 2007, para 87
[131] Ibid, para 109

II. Disinformation under public international law: do the States lie?

The 21st century is characterised by the individualisation of liability, shifting focus from the public actors to private individuals, groups of individuals and business companies. The discourse around the responsibility of social media, in this respect, serves as the brightest example of the changed attitude and restructured regulatory framework.[132] This approach might sound reasonable since modern technologies enable delivering avalanches of materials in a matter of seconds and without any physical border, while no one censors the contents or conducts an editorial review. However, it also created a powerful shield for the States and other 'formal' actors willing to intervene in the information environment and modify it. This, in its turn, created a perception of the States being in a 'helicopter position' since they cannot be held liable for any violation in the information sphere. However, this perception is quite wrongful. Following World War II, the international community developed a sequence of treaties addressing the issue of misleading information, propaganda for war and the State's inaction in tackling it, as well as numerous other concerns related to the circulation of malicious materials. Particularly, the existing international regulations include:

- **Bilateral treaties**. Agreements between two States, signed mostly before World War II, aimed at regulation of predominantly radio- and TV-broadcasting. For example, in 1931, Germany and Poland concluded an agreement obliging them *"to take reasonable steps to prevent broadcasts prejudicial to the spirit of cooperation and understanding"*, including the cessation of any hostile propaganda activities.[133] Similarly, the Soviet Union included certain provisions on regulation of propaganda in bilateral friendship and non-aggression treaties.[134] Yet, such bilateral treaties, in most cases, either resolved the problems short-term issues or were overly specific and inflexible (regulating a very particular issue of broadcasting). Rarely did they reach the regional level, transforming into multilateral treaties (which happened mainly in the Latin America region).[135] Unfortunately, no bilateral agreements on broadcasting or circulation of information exist between Ukraine and Russia, whereas the Budapest Memorandum, ensuring Ukrainian security after nuclear disarmament, contains no clauses on propaganda activities, while its binding character is highly disputable.[136] Thus, the responsibility under the bilateral

[132] Tetiana Avdieieva, 'Non-Star Wars: Social Media in the Times of Armed Conflict' (Centre for Democracy and Rule of Law, 27 September 2022) <https://cedem.org.ua/en/analytics/sotsmerezhi-zbroyni-konflikty/> accessed 27 December 2022

[133] Elizabeth A Downey, A Historical Survey of the International Regulation of Propaganda (1984) 5 (1) MJIL 341, 345

[134] Ibid

[135] Inter-American Radiocommunications Convention (entered into force 17 April 1939) UNTS 938; North American Regional Broadcasting Agreement (signed 13 December 1937, entered into force 29 March 1941)

[136] Роман Малко, 'Володимир Василенко: «Щоб Україна стала суверенною державою, було потрібно позбутись ядерної зброї»' (Український Тиждень, 17 April 2014) <https://tyzhden.ua/volodymyr-vasylenko-shchob-ukraina-stala-suverennoiu-derzhavoiu-bulo-potribno-pozbutys-iadernoi-zbroi/> accessed 27 December 2022

treaty can hardly be triggered to ensure responsibility for disinformation, propaganda for war or hate crimes.

- **International Convention concerning the Use of Broadcasting in the Cause of Peace (1936)**.[137] This document binds the States to *"restrict expression which constituted a threat to international peace and security"*, covering by this prohibition any incitement of the *"population of any territory to acts incompatible with the internal order or the security of a territory"*.[138] The restrictions on the free flow of information are allowed to ensure military security, which significantly limits the States' margin of appreciation. Although the OSCE RFoM called this agreement rather obsolete, with no distinction made between the State and private individuals,[139] it still remains in force for Russia as the legal successor of the Soviet Union.[140] At the time of ratification, many States doubted its enforceability, while others were concerned about the significant restrictions on sovereignty, thus making certain reservations.[141] For example, the Soviet Union made a reservation to Article 7, which requires negotiations and judicial settlement of any dispute under the Convention. Therefore, the ICJ would not have automatic jurisdiction over the dispute against Russia as its legal successor, under the Convention. In this respect, Ukraine has never ratified the Convention to even initiate the judicial proceeding. However, the UK objected to the Soviet Union's reservation; therefore, some scholars stress that **the UK might have had a legal standing before the ICJ**,[142] at least to contest the validity of the reservation, as well as ensure the Court's jurisdiction over the claims under the Convention. Others even go further, claiming that the very essence of reservation contravenes the object and purpose of the Convention since it is aimed at the peaceful settlement of emerging disputes. Unfortunately, the UK denounced this treaty in 1986, losing the right to invoke it before international courts. Nevertheless, many other States that are parties to the Convention have already condemned Russian disinformation activities, *e.g.* Norway, Finland, Estonia, Denmark, Luxembourg, Latvia, Hungary, and Bulgaria criticised the information warfare. Thus, they might well initiate the review of the case before the ICJ or non-judicial forum, such as the UN General Assembly, which has already dealt with this Convention in its work.[143]

  As regards the substantial breaches, **Russia has already violated Articles 2-4** of the Convention**,** which outlaw the use of broadcasting for propaganda for war, transmission which is likely to harm good international understanding (obliging the State to cease such activities as soon as it is aware of them),

[137] International Convention concerning the Use of Broadcasting in the Cause of Peace (signed 23 September 1936) 4319 UNTS 186

[138] Ibid, Article 1

[139] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) 12 <https://www.osce.org/fom/203926> accessed 27 December 2022

[140] Paul R Williams, 'The Treaty Obligations of the Successor States of the Former Soviet Union, Yogoslavia, and Czechoslovakia: Do The akia: Do They Continue y Continue in Force' (1994) 23(1) Denver Journal of International Law and Policy 1, 18-22

[141] Elizabeth A Downey, A Historical Survey of the International Regulation of Propaganda (1984) 5(1) MJIL 341, 344

[142] Talita de Souza Dias, 'Russia's "genocide disinformation" and war propaganda are breaches of the International Convention Concerning the Use of Broadcasting in the Cause of Peace and fall within the ICJ's jurisdiction' (EJIL Talk, 4 March 2022) <https://www.ejiltalk.org/russias-genocide-disinformation-and-war-propaganda-are-breaches-of-the-international-convention-concerning-the-use-of-broadcasting-in-the-cause-of-peace-and-fall-within-the/> accessed 27 December 2022

[143] International Convention concerning the Use of Broadcasting in the Cause of Peace, UN General Assembly (9th session: 1954)

as well as a prohibition of transmission of non-verified news. Since Russian false genocide allegations were used as a precondition for the illegal invasion and as an 'international law shield' against potential liability,[144] information was used in direct violation of the mentioned provisions. Accordingly, there are sufficient reasons to institute proceedings before competent international bodies.

- **Convention of the International Right of Correction (1952)**.[145] This treaty implies the right of the States to monitor and require rectification or removal of incorrect, inaccurate or misleading data about them delivered by foreign journalists. The historical preconditions were quite broadly described by Downey, who referred to Napoleon's demands to Great Britain to suppress hostile private propaganda and Belgium's passing of laws punishing private activities against foreign officials following the request of the French authorities in 1850s.[146] Both countries agreed on monitoring the information space to avoid the destruction of the foreign governments' reputation. However, little support was expressed towards the Convention itself outside Africa and South America since, in 1950s, most Western States stood on the position that the free flow of information could heal the unjustly damaged reputation of the foreign State.[147] Though, it is important to mention that in 1950s, any material reaching the general public was subjected to the editorial control of journalists and media. Nowadays, the applicability of the Convention is highly disputable given the absence of a specific actor to perform the obligations, especially when the regulation for social media, as the most popular tool for communication, is lacking or restricted. Moreover, it remains practically unfeasible for the State to monitor the information activities conducted online. Lastly, neither Ukraine nor Russia has ratified this Convention,[148] while the treaty implies no possibility to bind the third parties with any kind of obligation without their consent, as well as contains no *erga omnes* obligations.
- **Budapest Convention on Cybercrime (2001)**.[149] Although the Convention itself primarily addresses technical issues, containing only a ban on child porn and copyright infringements among the content restrictions, additional protocol to the Convention explicitly outlaws xenophobic and racist materials in cyberspace. In particular, it obliges to limit these types of speech on legislative and executive levels.[150] The same restriction applies to the denial and approval of genocide and crimes against humanity as the gravest possible speech violations. Although Russia has never been a party to this Convention,[151] the

[144] Reality Check team, 'Ukraine crisis: Vladimir Putin address fact-checked' (BBC News, 22 February 2022) <https://www.bbc.com/news/60477712> accessed 27 December 2022

[145] Convention on the International Right of Correction (entered into force 24 August 1962) 435 UNTS 191

[146] Elizabeth A Downey, A Historical Survey of the International Regulation of Propaganda (1984) 5(1) MJIL 341, 342

[147] Dale Stephens, 'Influence Operations & International Law' (2020) 19(4) JIW 1, 13

[148] Convention on the International Right of Correction, List of Parties <https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XVII-1&chapter=17&clang=_en> accessed 27 December 2022

[149] Convention on Cybercrime (signed 23 November 2001) ETS 185

[150] Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems (2003) ETS 189

[151] Convention on Cybercrime and its Protocols, List of Parties <https://www.coe.int/en/web/cybercrime/the-budapest-convention> accessed 27 December 2022

Explanatory Note on the first additional protocol explicitly refers to the case law of the ICTY, the ICTR, and the ICC, while addressing the jurisdictional matters.[152] Since Ukraine is a party to this Convention, the commission of at least some elements of the crime on its territory enables the State to introduce additional limitations on the free flow of information, particularly to safeguard the population against approval of crimes against humanity. Accordingly, this legal instrument provides Ukraine with an additional framework for content restrictions.

- **International human rights treaties**.[153] Some scholars mention a potential for interpretation of the human rights conventions that imply the possibility for inter-State disputes as outlawing State-backed malicious propaganda.[154] Among the courts established under such treaties, the ECtHR fairly is considered one of the most progressive courts, yet having dealt only with the issues of illegal propaganda being spread by private individuals. Thus, the Court never addressed the potential breach of human rights following the State propaganda or illegal incitements. As was mentioned in the previous Chapter of this research, in the context of aggression against Ukraine, the ECHR can hardly be applicable given the Russian denunciation of this convention since September 2022, while the ICCPR and the CERD have never been interpreted in the mentioned way. Yet, importantly, the UN, the AComHPR, the IAComHR, and the OSCE Representative on Freedom of the Media have already qualified Russian propaganda for war and disinformation campaigns as contravening Article 20 of the ICCPR, delivering this statement in the context of the State responsibility rather than an individual criminal liability.[155] This, in turn, proves the **potential for a dispute before the ICJ or the CERD Committee** to be launched, specifically addressing the abuse of free speech regulations by the Russian side during the aggressive war against Ukraine.
- International custom. Customary law reflects the general practice accepted by the States as binding upon them, which, however, has no formal manifestation, such as codification in the provisions of the international treaty or other hard law documents. For example, scholars consider the amount of State practice and opinio juris as sufficient to evidence that "incitement to conflict constitutes a grave menace to peace", being customarily outlawed.[156] In practice, false claims and disinformation might fall within the ambit of this prohibition as the elements of illegal incitements. An approximate period of the emergence of the customary norm is dated by the end of the French Revolution.[157] Nowadays, the validity of the norm can be proven by numerous documents reflecting the

[152] Explanatory Report to the Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems (2003) ETS 189, 7

[153] Vernon Van Dyke, 'The Responsibility of States for International Propaganda' (1940) 34(1) AJIL 58

[154] Elizabeth Willmott-Harrop, 'Iraq: propaganda's war on human rights' (Liberty & Humanity, June 2005) <https://libertyandhumanity.com/themes/international-human-rights-law/iraq-propagandas-war-on-human-rights/> accessed 27 December 2022

[155] 'Ukraine: Joint statement on Russia's invasion and importance of freedom of expression and information' (UN Human Rights Office, 4 May 2022) <https://www.ohchr.org/en/statements-and-speeches/2022/05/ukraine-joint-statement-russias-invasion-and-importance-freedom> accessed 27 December 2022

[156] John B Whitton, 'Hostile International Propaganda and International Law' (1971) 398 SPI 14, 21

[157] Elizabeth A Downey, A Historical Survey of the International Regulation of Propaganda (1984) 5 (1) MJIL 341, 342

same position, e.g. the 1947 UN General Assembly Resolution 127(III) inviting the States to combat the diffusion of false or distorted reports, which may hinder friendly relations between them,[158] the 1978 Declaration on Fundamental Principles Concerning the Contribution of the Mass Media to Strengthening Peace and International Understanding (addressing apartheid and racism issues),[159] the 1987 Declaration on the Enhancement of the Effectiveness of the Principle of Refraining from the Threat or Use of Force in International Relations (stipulating the States' obligation to "refrain from propaganda for wars of aggression")[160] and many other soft law documents.[161] Another proof of the customary nature of the norm is that States resorting to such activities usually look for plenty of excuses to justify their actions, thus considering subversive propaganda to be illegal.[162] The scientific research also proves that the domestic practice of various States is emblematic of municipal laws and judicial decisions supporting the customary nature of the prohibition on subversive misleading foreign propaganda.[163]

In this respect, the existence of the customary prohibition switches the discourse regarding the responsibility from the applicable standard to the venue where the dispute can be adjudicated. For example, Russia did not recognise ipso facto jurisdiction of the ICJ regarding any question of international law,[164] thus complicating the process of instituting proceedings for violation of customary norms. Namely, Russia can serve as a respondent party only where international treaties explicitly refer the dispute to a particular judicial body, as happened in Ukraine v Russia dispute under the CERD.[165] In other cases, there will be, at least, a possibility to object to the jurisdiction of the Court, as now happens in Ukraine v Russia dispute under the Genocide Convention.[166] Nevertheless, if a special arbitration is created for reviewing the dispute over the Russian aggression in Ukraine with its statute comprising the breach of international custom, Ukraine can bring a claim on the dissemination of disinformation as a part of the propaganda for war campaign.

As numerous historical examples demonstrate, illegal propaganda is reviewed with the closest scrutiny under the law on State responsibility since State-affiliated media usually have more coverage and are less balanced in information delivery. In this

[158] Resolution on False or Distorted Reports 127(II) (UN General Assembly, 1947)

[159] Declaration on Fundamental Principles Concerning the Contribution of the Mass Media to Strengthening Peace and International Understanding (UNESCO, 28 November 1978, 20th session)

[160] Declaration on the Enhancement of the Effectiveness of the Principle of Refraining from the Threat or Use of Force in International Relations : resolution / adopted by the General Assembly. (UN General Assembly 1988, 42nd session)

[161] Elizabeth Willmott-Harrop, 'Iraq: propaganda's war on human rights' (Liberty & Humanity, June 2005) <https://libertyandhumanity.com/themes/international-human-rights-law/iraq-propagandas-war-on-human-rights/> accessed 27 December 2022

[162] John B Whitton, 'Hostile International Propaganda and International Law' (1971) 398 SPI, 14-25, 18

[163] Vernon Van Dyke, 'The Responsibility of States for International Propaganda' (CUP, 12 April 2017) <https://www.cambridge.org/core/journals/american-journal-of-international-law/article/abs/responsibility-of-states-for-international-propaganda/7A16191D0CE3E7875071C3163E4D8D0E> accessed 27 December 2022

[164] 'Declarations recognizing the jurisdiction of the Court as compulsory' (ICJ) <https://www.icj-cij.org/en/declarations> accessed 27 December 2022

[165] Application of the International Convention for the Suppression of the Financing of Terrorism and of the International Convention on the Elimination of All Forms of Racial Discrimination (Ukraine v. Russian Federation) (2017) ICJ Rep 104

[166] Allegations of Genocide under the Convention on the Prevention and Punishment of the Crime of Genocide (Ukraine v Russian Federation) (2022) ICJ Rep 182

regard, the OSCE RFoM stressed that State-owned and State-run media outlets should be viewed as a corrupt practice that is only "throwing gasoline on an open fire" when distributing propaganda in times of armed conflict.[167] Particularly, disinformation not only escalates the aggressive action but also misleads as to the course of armed activities, their outcome and accompanying dangers for civilians. Although the State's participation in propaganda activities is highly criticised, it is also very difficult to be proven. Especially in the issues of causation between a delivered material and the repercussions, as well as the attributability of such actions to the State.

> It is not enough to prove that a misleading message contains propaganda for war. One shall also prove that **the State itself is responsible** for this message by proclaiming it through its organs, instructing, controlling or directing the media, or *post factum* adopting the expression as its own.

Attribution. According to Article 2 of the ARSIWA, an internationally wrongful act occurs when the conduct in the form of an action or omission is attributable to the State and constitutes a breach of international obligations.[168] The discourse around the breach concerns the relevant conventions, rules of customary law or the general principles of law. The rules on attribution, however, are exact and limited in their scope to the specific instances when the conduct can be ascribed to a State. For the purposes of breaches in the information sphere, the relevant rules on attribution are:

- Conduct of a **State organ** (ARSIWA, Article 4). A good example in this regard can be found in the speech of King Saud against the Israeli people, whom he described as cancer for the Arab world , which needed to be wiped out, in 1954.[169] Since the Head of the State, together with the Minister for Foreign Affairs and the Prime Minister, is empowered to make the declarations binding for the State,[170] such a statement can be fairly considered as one proclaimed on behalf of the State (or by the State organ). Therefore, the issue will primarily concern State responsibility for violation of the relevant norms of international law, and only then an individual criminal liability of the persons concerned.
  An example from the context of Russian aggression against Ukraine implies the wide-scale disinformation campaigns launched by Russian embassies regarding the atrocities and mass human rights violations committed in Mariupol.[171] Since the embassies are acting on behalf of the State, dissemination of malicious and misleading narratives is attributable to the State itself. Accordingly, Russia can be brought to justice for violating its

[167] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) 11 <https://www.osce.org/fom/203926> accessed 27 December 2022
[168] Articles on Responsibility of States for Internationally Wrongful Acts (2001) UN Doc A/RES/56/83 ["ARSIWA"], Article 2
[169] John B Whitton, 'Hostile International Propaganda and International Law' (1971) 398 SPI 14, 21
[170] Arrest Warrant of 11 April 2000 (Democratic Republic of the Congo v Belgium) (2002) ICJ Rep 3 ["Arrest Warrant"], para 51
[171] Katie Jennings, 'How Technology Might Bring War Criminals To Justice In Ukraine' (Forbes, 21 May 2022) <https://www.forbes.com/sites/katiejennings/2022/03/21/the-world-watched-russia-bomb-ukraines-hospitals-what-happens-next/?sh=465d860a5763> accessed 27 December 2022

international obligations. The same applies to the public statement of the Russian President, Prime Minister and official releases on the websites of State organs.

- Conduct of persons **empowered to exercise the governmental authority** (ARSIWA, Article 5). To trigger Article 5 of the ARSIWA, the authorisation shall concern specific elements of the governmental authority. The decisive criterion is the functional one, namely, an entity should discharge state functions, both public and regulatory.[172] A State may empower a private entity, even in a limited or specific context, to perform governmental powers either by laws or other legal mechanisms, *i.e.* contracts.[173] The lack of express empowerment is not decisive for establishing State responsibility.[174] For example, in *Yeager v Iran*, the mere tolerance of Iran towards the activity of private entities, which *de facto* exercised governmental powers, caused the applicability of Article 5.[175]
  Russia sporadically provides State support or financing based on public contracts, in parallel empowering the media outlets with depicting the current state of affairs or performing the role of the State information agencies. Depending on the nature of the contracts and the scope of the powers conferred upon such media, the attribution links under Article 5 can be built.
- Conduct of the State organ or a person empowered with governmental authority if such an organ or person **violates the instructions or exceeds the authority** (ARSIWA, Article 7). This rule prescribes that campaigns coordinated by the State are attributable to it even when hyperbolisation on behalf of individual propagandists might take place. For example, in Myanmar civilian government and military were commenting on the current State of affairs in the media and on their Facebook pages, acting in their official capacity.[176] Although they were not each time authorised to make certain comments, their conduct directly falls within the notion of excess of authority. Thus, it is directly attributable to the State.
  The same situation is now happening in the case of Russian aggression: numerous Russian parliamentarians, representatives of different Ministries and other State organs are delivering public statements in the media, as well as on their private social media profiles, where they are mentioning their public officials status and acting as the State representatives. Accordingly, such behaviour shall be qualified as the activity performed on behalf of the State, triggering the norms on the State's responsibility.
- Conduct of a person based on **instructions, directions or under the control of the State** (ARSIWA, Article 8). If instructions and directions usually imply relatively clearcut guidance for a non-State actor on how to behave, the PIL

[172] Commentary to the Articles on Responsibility of States for Internationally Wrongful Acts (YILC 2001/II(2)) UN Doc. A/CN.4/SER.A/2001 ["ARSIWA Cmt"] 43; Maffezini v Spain, ICSID Case no ARB/97/7, Award, 13 November 2000, para 52; Jan de Nul N.V. and Dredging International N.V. v Arab Republic of Egypt, ICSID Case no ARB/04/13, Award, 6 November 2008, para 168

[173] ARSIWA Cmt to Article 5, 43; Crawford J., State Responsibility: The General Part (CUP, 2013) 130-131

[174] Brownlie I., Crawford J., Brownlie's Principles of Public International Law (OUP, 8th edn., 2012) 549; Helnan International Hotels A/S v Arab Republic of Egypt, ICSID Case no ARB/05/19, Award, 3 July 2008, para 93

[175] Yeager v Islamic Republic of Iran, Award no 324-10199-1, 17 Iran-US CTR 92 (1987), para 45

[176] Jenny Domino, 'Crime as Cognitive Constraint: Facebook's Role in Myanmar's Incitement Landscape and the Promise of International Tort Incitement Landscape and the Promise of International Tort Liability Liability' (2020) 52(1) Case Western Reserve University School of Law 143, 155

discourse often struggles with the threshold for the "control test".[177] Namely, the main issue remains whether the specific control of each step undertaken by the individual or an entity is required or whether the overall coordination suffices for the Article 8 rule to apply. For the purposes of State responsibility, the ICJ in *Nicaragua* and *Bosnian Genocide* cases has explicitly outlined a strict threshold for control over the non-State actor.[178] Though the mentioned cases concerned the activities of paramilitary groups, the requirements for non-military entities, such as media, might be milder.

In this regard, a relevant example is the State-sponsored information campaigns on Internet platforms, which intermediaries usually qualify as activities directly attributable to the States. For instance, Russia paid for the advertising of materials developed by the State-controlled media outlets, which contained massive avalanches of disinformation, manipulations or propaganda, thus forcing the intermediary platforms to disable the advertising function within Russian territory.[179] The main reason behind such a decision was that the media got the instructions, directions or were under the control of Russian State authorities, therefore, the activities of such outlets constitute part of the aggressive war against Ukraine. Another example is Russia-24, which usually distributes pro-governmental narratives, such as accusations of Ukrainian of creating deepfakes, allegations of the breach of the Geneva Convention etc.[180] Since the media is controlled by the State, its actions are attributable to Russia in the dimension of State responsibility.

- Conduct **acknowledged and adopted by the State as its own** (ARSIWA, Article 11). In the 1956 *Lighthouses* case, the arbitral tribunal found Greece liable for Crete's breach of agreement with the French entity, partly because Greece endorsed such breach and continued it, *i.e.* treated it as if it had been a proper act.[181] Similarly, in *Tehran Hostages*, the ICJ found approval of Iranian authorities towards the occupation of the US Embassy as being sufficient to trigger Article 11. In the dimension of media, if considering disinformation to be an action in breach of international law, any endorsement of such activity will amount to acknowledgement and adoption of the act as the State's own. As POLITICO research shows, a significant part of the Russian disinformation campaign is led by anonymous or *prima facie* private sources, whose link with the State is difficult to be proven.[182] However, the State often refers to the so-

[177] Christiane Rexilius, 'Scenario 18: Legal status of cyber operators during armed conflict' (Cyber Law Toolkit, 3 March 2022 (edited) <https://cyberlaw.ccdcoe.org/wiki/Scenario_18:_Legal_status_of_cyber_operators_during_armed_conflict> accessed 27 December 2022

[178] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America) (1986) ICJ Rep 14; Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v Serbia and Montenegro) (2007) ICJ Rep 43

[179] Brad Smith, 'Digital technology and the war in Ukraine' (Microsoft, 28 February 2022) <https://blogs.microsoft.com/on-the-issues/2022/02/28/ukraine-russia-digital-war-cyberattacks/> accessed 27 December 2022

[180] Digital Forensic Research Lab, 'Russian Hybrid War Report: Russia retaliates against anti-war celebrities as social platforms crack down on Russian media' (Atlantic Council, 28 February 2022) <https://www.atlanticcouncil.org/blogs/new-atlanticist/russian-hybrid-war-report-russia-retaliates-against-anti-war-celebrities-as-social-platforms-crack-down-on-russian-media/> accessed 27 December 2022

[181] Affaire relative à la concession des phares de l'Empire ottoman, UNRIAA (Vol. 12, pp. 155-269) (1956) 198

[182] Mark Scott, 'As war in Ukraine evolves, so do disinformation tactics' (POLITICO, 10 March 2022) <https://www.politico.eu/article/ukraine-russia-disinformation-propaganda/> accessed 27 December 2022

called *"debunking of myths around Russian atrocities in Ukraine"*,[183] which is done by private actors. By doing this, the State *de facto* acknowledges the existence of disinformative materials, while positive feedback on their substance can serve as approval of malicious content. Accordingly, the State responsibility mechanism is triggered.

Non-State actors. The situations when the States act through the so-called proxies (non-State actors) become widespread, especially in view of modern technologies emerging. This, in turn, makes the mechanism of State responsibility much more sophisticated. Sometimes establishment of an attribution link becomes impossible, at first glance, excluding State responsibility. However, everything is more complex than it seems. In this respect, one shall bear in mind that the State also is obliged to exercise due diligence[184] while protecting human rights. Namely, it implies a **duty to effectively prevent and expeditiously investigate and punish the violations** committed within its territory. The jurisdiction of the State over the impugned activities is a crucial element. Where it is present, and the State failed to diligently prevent the malicious action or react upon the violation by ceasing it – there is a breach on behalf of the State authorities in the form of an omission. Respectively, the attribution links will be built in accordance with the Article 4 rule. Although a discourse regarding the jurisdiction concerning international crimes, such as war crimes or crimes against humanity, is quite sophisticated, commission of any element of the crime within the State's territory is enough to trigger its jurisdiction.[185] For example, Facebook activities, or rather inaction, in Myanmar are emblematic of the State's lack of regulation of the media sphere, as well as ineffective investigation into intermediaries' activities, *i.e.* contribution of tech giants to crime commission.[186] Thus, even the absence of any link of control or directions towards the non-State actors does not always exclude State responsibility.

Another problematic issue, which is high on the agenda in the armed conflict launched by Russia against Ukraine, is the application of AI technologies, in particular for the purposes of creating deepfakes. For example, a video with Ukrainian President Zelenskyi allegedly announcing the capitulation,[187] and a deepfake of Kyiv mayor used to mislead the mayors of other European cities in the course of Zoom-calls[188] are only a couple of examples, when AI-driven technologies were used. In this regard, scholars indicate the difficulty in the attribution of the conduct of the artificially

[183] Digital Forensic Research Lab, 'Russian Hybrid War Report: Russia retaliates against anti-war celebrities as social platforms crack down on Russian media' (Atlantic Council, 28 February 2022) <https://www.atlanticcouncil.org/blogs/new-atlanticist/russian-hybrid-war-report-russia-retaliates-against-anti-war-celebrities-as-social-platforms-crack-down-on-russian-media/> accessed 27 December 2022

[184] Tim Stephens and Duncan French, Second Report of ILA Study Group on Due Diligence in International Law (ILA, July 2016)

[185] Tommi Aromäki, 'The International Criminal Court's Jurisdiction Over Incitement to Genocide in the Internet Era - Some special situations' (University of Helsinki, 2021) 23-29

[186] Jenny Domino, 'Crime as Cognitive Constraint: Facebook's Role in Myanmar's Incitement Landscape and the Promise of International Tort Incitement Landscape and the Promise of International Tort Liability Liability' (2020) 52(1) Case Western Reserve University School of Law 143, 173-178

[187] Jane Wakefield, 'Deepfake presidents used in Russia-Ukraine war' (BBC News, 18 March 2022) <https://www.bbc.com/news/technology-60780142> accessed 27 December 2022

[188] 'Зловмисник за допомогою DeepFake видав себе за Кличка і поговорив з мерами Берліна і Мадрида' (Ukrinform, 25 June 2022) <https://www.ukrinform.ua/rubric-kyiv/3514835-zlovmisnik-za-dopomogou-deepfake-vidav-sebe-za-klichka-i-pogovoriv-z-merami-berlina-i-madrida.html> accessed 27 December 2022

created images and videos to a State since they are created by the AI.[189] On the one hand, the problem implies the proof of the control over the activity of the AI tool, which cannot be considered as being fully dependent upon the State (because the outcome of the AI work might be unpredictable, especially in real-time deepfakes). On the other hand, the high quality of videos might mislead the audience[190] to such an extent one believes it was an actual speech of the State official. In this respect, the issue of the burden of proof remains a critical point for the discourse. Nevertheless, when the use of AI-driven tools to mislead the counterparty to the conflict by distributing deepfakes has been proven, the absence of effective or overall control is not decisive since the same due diligence rule applies. Namely, the State shall prevent any malicious activity stemming from its territory if it contravenes the IHL or any other international obligation.

Accordingly, the positive obligations of the State to prevent malicious incitements, propaganda for war, and other prohibited types of speech oblige it to interfere with the work of media, social media platforms and other actors in the information environment. When the State neglects such an obligation or even tacitly accepts the illegal behaviour, providing a silent permit for the continuation of such actions – the responsibility will occur. Some scholars also argue that such an obligation extends to the States where the headquarters of the platforms are located.[191] For instance, in the case of Myanmar, the alleged breach of due diligence was attributed to the US as a domicile state for Facebook. The expected action from the US government, in turn, comprised the investigation into the use of social media as a venue for the dissemination of incitements to genocide. In the case of the Russian aggression against Ukraine, this model of liability applies as well. Namely, media outlets and social media platforms are used for spreading disinformation, propaganda for war and hateful messages, while the State, even if not directly encouraging it, still remains reluctant towards numerous breaches occurring under its jurisdiction. Among the social media used for such actions, in particular, Vkontakte is fairly labelled as creating a hostile environment, discriminating against Ukrainians and being favourable towards malicious propaganda and disinformation. Hence, Russian State authorities fail to exercise due diligence with regard to the third parties acting from Russian territory.

> The State's choices are intended to be *"political and circumstantial, rather than exclusively or constantly of a legal character"*.[192] Thus, foreign interference with such choices amounts to **a breach of the State's international obligations**, triggering State responsibility.

[189] David Allen, 'Deepfake Fight: AI-Powered Disinformation and Perfidy Under the Geneva Conventions' (2021) SSRN 1, 32-36
[190] Tetiana Avdieieva, 'Who Is Hiding Behind Digital Avatars?' (Centre for Democracy and Rule of Law, 29 October 2021) <https://cedem.org.ua/en/analytics/tsyfrovi-avatary/> accessed 27 December 2022
[191] Kyle Rapp, 'Social media and genocide: The case for home state responsibility' (Taylor & Francis Online, 8 September 2021) <https://www.tandfonline.com/doi/abs/10.1080/14754835.2021.1947208?journalCode=cjhr20> accessed 27 December 2022
[192] Rajan M S, United Nations and World Politics: essays from a nonaligned perspective (New Delhi: Har-Anand Publ, 1995) 147

Breach of sovereignty. Although the armed conflict primarily implies the applicability of the IHL legal regime, information operations might likewise violate the sovereignty of the State even prior to the war beginning. When the State's activities of delivering malicious content do not reach the threshold of an armed attack, they still significantly affect the choices of the target State in a cultural, economic or information sphere. In this respect, in *Lotus* and *Nicaragua*, the PCIJ and ICJ respectively explicitly stated that the States *"may not exercise [their] power in any form in the territory of another State"*.[193] While analysing the impact of foreign disinformation on the election process, scholars reached a conclusion that disinformation can be qualified as a violation of sovereignty and the principle of non-intervention if it affects the freedom of the State's choice in formulation of its policies of any kind.[194]

Malicious propaganda, including disinformation, can be ascribed to different actors, being diversified in its target groups and intensity.[195] Yet, neither the aim nor the form themselves are decisive, but the overall effect is important for State responsibility to arise. According to the Declaration on Principles of International Law concerning Friendly Relations and Cooperation Among States, a breach emerges in case of **unconsented coercive interference of any form or attempted threats against the personality of the State**, including the support of subversive activities or civil strife.[196] For interference to become a breach of the non-intervention principle, a certain degree of coercion is required. Importantly, in this paradigm the information operation shall extend beyond the mere persuasion or expression of external vision on the current state of affairs.[197] For example, comments on the foreign economy or policy of other States apparently would not constitute unlawful interference even if a certain degree of criticism is present. Foreign propaganda aimed at the Soviet Union population but not encouraging it to rebel or change the regime in the State was not in itself coercive or radically subversive.[198] In contrast, fraud or threats to the State officials, based on data exchange,[199] naturally qualifies as a coercive practice, thus breaching the rule of non-intervention.

Despite some critical voices, disinformation is still considered manipulative enough to be perceived as coercive for the purposes of State responsibility.[200] Lauterpacht once claimed that sovereignty is undermined if a government sends disruptive

[193] Lotus case (France v Turkey) (1927) PCIJ Ser A No 10, para 45; Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America) (1986) ICJ Rep 14, para 263

[194] Tetiana Avdieieva, 'Disinformation as an Instrument of Foreign Intervention into Electoral Process: a Manifestation of Lawfare' (Centre for Democracy and Rule of Law, 23 June 2021) < https://cedem.org.ua/en/analytics/internet-regulation/> accessed 27 December 2022

[195] Arthur Larson, 'The present status of propaganda in international law' (1966) 31 Law and Contemporary Problems 439, 448

[196] Declaration on Principles of International Law concerning Friendly Relations and Cooperation Among States, UNGA Res 2625, UN Doc A/1883 (1970)

[197] Dale Stephens, 'Influence Operations & International Law' (2020) 19(4) Journal of Information Warfare 1, 4

[198] Natalie David McDonald, 'Radio Free Europe and the Right to be Informed: National Sovereignty and Freedom of Information During the Cold War' (2017) 27(4) Claremont Colleges Library Undergraduate Research Award 1, 10

[199] Jens David Ohlin, 'Did Russian Cyber Interference in the 2016 Election Violate International Law?' (2017) 95 Texas Law Review 1579, 1594

[200] Lando Kirchmair, 'It's Not Propaganda If It's True' (Verfassungsblog., 2 March 2022) <https://verfassungsblog.de/its-not-propaganda-if-its-true/> accessed 27 December 2022

appeals to foreign people, thus doubting the State's right to existence.[201] For example, designing a visual image of a failed State in the eyes of the nation or stimulating the failure of the economy by creating an atmosphere of distrust in the bank system directly falls within the category of prohibited information operations. Since in the context preceding the armed conflict, the information operations might themselves trigger physical violence, shooting or lead to civil unrest, further transforming into an armed conflict, it may well suffice to be qualified as a breach of the State's sovereignty.

In the context of Russian aggression against Ukraine, it is of paramount importance to analyse and appropriately qualify the Russian information operations both before the invasion (before 2014) and within the course of the ongoing armed conflict. First and foremost, such disinformation campaigns may themselves serve as the precondition for launching armed aggression, making the population highly receptive towards misleading and malicious messages. Although the Dutch Ministry of Foreign Affairs underlined that a cyber intervention shall aim at impacting the behaviour of the target State,[202] such influence shall not necessarily be momentaneous or straightforward. The process of weaponization of information usually take years, if not decades, to successfully cause the expected practical consequences. In this respect, when the issue concerns the international crimes' commission, both the ICTR and ICTY have unanimously stated that it is enough to prove that instigation or incitement served as a contributing factor to a crime, not necessarily being the primary cause for its commission.[203] Accordingly, international community has an unrebuttable duty to prevent illegal information activities since the approach primarily shall remain proactive rather than reactive, while the absence of immediate repercussions does not exclude their probability in the future.

The Russian disinformation campaigns conducted before the launch of the aggressive war against Ukraine shall be likewise analysed, with the legal mechanisms of responsibility being adjusted for considering such cases. Given the current circumstances, a judicial body adjudicating the claims of Ukraine shall take into account the violation of Ukrainian sovereignty occurring long before the physical military invasion. Moreover, disinformative propaganda influences not only the target State and its population but also the international perception of the State's sovereignty by other actors. Russian propaganda eventually led to the recognition of the so-called 'DPR' and 'LPR' created in Eastern Ukraine by other States, including Syria, which treated these non-State formations as exercising their right to self-determination. Consequently, propaganda for war, subversive propaganda and certain types of illegal incitements, particularly those based on untrue facts, violate the State's sovereignty when distributed by foreign actors.

[201] Hersch Lauterpacht, International Law: A Treaties (Lindon, Green, 1955) 259

[202] Schmitt M N, 'Foreign Cyber Interference in Elections' (2021) Vol 92 Int'LL.Stud. 793

[203] Prosecutor v Kvočka, Trials Chamber (2001) IT-98-30/I-T, para 252; Prosecutor v Nahimana, Barayagwiza and Ngeze, Appeals Chamber (2007) ICTR-99-52-A ["The Media case"], para 501

Incitement to genocide. Ukraine has recently filed an application before the ICJ for the breach of the Genocide Convention by Russia.[204] Particularly, the application contains parts on the falsity of Russian allegations in the genocide commission by the Ukrainian side, with a request to the Court to clean Ukraine's reputation. *De facto*, this claim also amounts to the recognition of the Russian campaign as containing disinformation. Genocide may well exist in the context of the armed conflict or beyond it,[205] because the presence of the elements of the crime still remains the decisive element. In this respect, incitement to genocide may as well be delivered by the State or its proxies during the armed conflict or even before it is launched.

Although incitement to genocide has been reviewed mostly by the tribunals dealing with individual criminal liability, responsibility for such acts can also be attributable to the State. This was established by the ICJ in *Bosnian Genocide*, where the Court stressed that the States might as well become accomplices in the commission of genocide, although failing to address specifically the gap in the application of the Genocide Convention to such tricky circumstances as in Serbia.[206] Nevertheless, one might speak about the violation of the Genocide Convention, where incitement to genocide is mentioned as a separate violation in line with the complicity in genocide and attempt to commit it.[207] The difference between these acts will be reviewed in more detail in the dimension of international criminal law, whereas for the State responsibility, an attribution issue is important. In this respect, the same rules prescribed in ARSIWA are applicable – namely, pronouncing incitements by the State organs or persons controlled by the State may well lead to the State's responsibility for incitement to genocide. As well, the State's adoption of an act as its own, support of incitements delivered in media or via any other public source shall be qualified as attributable to that State.

Another obligation implies the duty to prevent. Scholars usually refer to the example of the speech of the Congolese Minister Ndombasi, which, as the Belgian warrant described, was *"inciting racial hatred"*.[208] Namely, if the DRC was accused of non-preventing incitement to racial hatred, failing to exercise due diligence in this regard, the accusations most probably would result in outcome unfavourable for the DRC. The same model would work with Rwanda[209] and Côte d'Ivoire,[210] where the State violated its obligations to stop malicious, often misleading, propaganda of illegal activities. Yet, the State responsibility was not triggered for the reasons that no third State actually considered this option, given the involvement of only Rwandan

[204] Allegations of Genocide under the Convention on the Prevention and Punishment of the Crime of Genocide (Ukraine v Russian Federation) (2022) ICJ Rep 182, Application of Ukraine

[205] Etienne Ruvebana and Marcel Brus, 'Before It's Too Late: Preventing Genocide by Holding the Territorial State Responsible for Not Taking Preventive Action' (2015) 62 Netherlands International Law Review 25, 27-29

[206] David Scheffer, 'The World Court's Fractured Ruling on Genocide' (2007) 2(2) Genocide Studies and Prevention: An International Journal 122, 129

[207] Convention on the Prevention and Punishment of the Crime of Genocide (signed 9 December 1948, effective 12 January 1951) 78 UNTS 277, Article 3

[208] Etienne Ruvebana and Marcel Brus, 'Before It's Too Late: Preventing Genocide by Holding the Territorial State Responsible for Not Taking Preventive Action' (2015) 62 Netherlands International Law Review 25, 36

[209] Mugesera v Canada (Minister of Citizenship and Immigration) [2005] 2 SCR 100, 2005 SCC 40

[210] UN Security Council, Resolution 1572 (2004, 5078th meeting)

officials in the genocide commission. To compare, if the RTLM was located in the neighbouring DRC, operating there with the knowledge of the DRC authorities, Rwanda could require the withdrawal of its license. If the DRC did not agree to comply with such a request, Rwanda would have a legitimate ground to institute the proceedings under the Genocide Convention.

Numerous scholars emphasised that the main difficulty of bringing States to justice for commission of genocide implies the necessity to prove intent to commit such an internationally wrongful act.[211] The same relates to incitement, where a specific *mens rea* is also required. Namely, one shall prove that the highest State officials or State-controlled media, or persons whose actions the State adopted had a specific intention. In this regard, the Russian full-scale invasion is emblematic of numerous public statements generalising, dehumanising and diminishing the independence and the right to existence of the Ukrainian nation.[212] Indeed, most disinformation narratives rather relate to an attempt to justify Russian war crimes in the territory of Ukraine since a direct knowledge of the illegality of such attacks is present.[213] However, some of such public expressions label Ukrainians as Nazis, encouraging to resolve 'the Ukrainian question', which apparently is of a different level of severity.

Russian State-controlled media even have equated de-Nazification with de-Ukrainisation,[214] which in its nature means if not ethnic cleansing, then, at least, total destruction of culture and language, denying the right of the nation to existence. Russian media also frequently resorted to the polarisation of Russian society, sharing radical views, which is quite similar to the tactics employed during the Rwandan genocide.[215] In the report, the US New Lines Institute for Strategy and Policy and Raoul Wallenberg Centre for Human Rights also reached a conclusion that there is a high risk that genocide of Ukrainians might be committed by Russia,[216] which was based primarily on the statements of Russian officials and media supporting and promulgating this idea. For instance, the report stresses that Russian soldiers might echo the State propaganda while committing crimes in Ukraine, which proves the imminence of consequences and makes the speech of Russian officials closer to illegal incitements outlawed by the Genocide Convention.[217] This research also provides extensive evidence and examples of phrases used and the context for the

[211] Sergiy Sydorenko, 'Not Only Putin Behind Genocide. Entire Russia Should Be also Tried for Utmost Crime' (European Pravda, 25 October 2022) <https://www.eurointegration.com.ua/eng/articles/2022/10/25/7149327/> accessed 27 December 2022

[212] Sara L Ochs, 'Russia has committed atrocities. But is genocide one of them?' (THINK, 7 April 2022) <https://www.nbcnews.com/think/opinion/ukraine-accuses-russia-war-crimes-genocide-bucha-just-one-problem-rcna23307> accessed 27 December 2022

[213] Yonah Diamond et al., 'An Independent Legal Analysis of the Russian Federation's Breaches of the Genocide Convention in Ukraine and the Duty to Prevent' (2022) US New Lines Institute for Strategy and Policy and Raoul Wallenberg Centre for Human Rights 19

[214] Ashish Kumar Sen, 'Is Russia Committing Genocide in Ukraine?' (US Institute of Peace, 21 September 2022) <https://www.usip.org/publications/2022/09/russia-committing-genocide-ukraine> accessed 27 December 2022

[215] Mohi Kumari, 'Genocide Under International Criminal Law' (Genocide Under International Criminal Law) <https://www.legalserviceindia.com/article/l433-Genocide-Under-International-Criminal-Law.html> accessed 27 December 2022

[216] Yonah Diamond et al., 'An Independent Legal Analysis of the Russian Federation's Breaches of the Genocide Convention in Ukraine and the Duty to Prevent' (2022) US New Lines Institute for Strategy and Policy and Raoul Wallenberg Centre for Human Rights

[217] Ibid

conflict, as well as elaborates on the deliberate nature of most State-coordinated disinformation campaigns.

Although the case under the Genocide Convention, instituted by Ukraine against Russia, is already pending, there is still a possibility to initiate a separate set of proceedings regarding the incitement to genocide or at least attach additional evidence to the existing one. Interestingly, another pending case before the ICJ – *the Gambia v Myanmar* – is also paying specific attention to the incitement to genocide, *i.e.* use of Facebook for spreading hate.[218] Although this case would probably be more interesting from the perspective of the corporate liability of intermediaries,[219] the responsibility of the States for platforms' activities in their territories would also be high on the agenda. As was mentioned, Russia is actively using Vkontakte to deliver hateful messages and create a hostile environment towards Ukrainians. And since most Russian soldiers use this social media, it is expected that incitements are qualified as likely to lead to imminent harm. Thus, State responsibility for the misuse of platforms for the sake of conducting disinformation campaigns shall also be analysed. Since the case against Myanmar will apparently be reviewed more rapidly, Ukraine will, at least, have an understanding of the ICJ approach towards qualifying online speech as a potential breach of the Genocide Convention.

IHL perspective. The issues around the violation of sovereignty are within the pre-armed conflict legal framework, while disinformation can also be shared after the war starts. When the issue concerns the IHL application, the first thing to be addressed is the existence of the armed conflict and its nature – either international or non-international one. Since in Ukraine, no doubts are raised as to the presence of an international armed conflict, this research will focus on the role of disinformation in the armed conflict itself and its qualification under the IHL.

Previously, non-physical activities hardly were viewed under the IHL, being not intensive enough to trigger its application.[220] However, technological development changed the approach towards stricter limitations on information activities. For the purposes of legal regulation two important issues are: whether disinformation campaigns amount to direct participation in hostilities, and, respectively, whether those, distributing such content become the legitimate targets. In this respect, many possible scenarios exist. For example, a certain group might use social media for civilian journalism or for dissemination of military information, which can be further

[218] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v Myanmar) (2022) ICJ Rep 78

[219] Jenny Domino, 'How Myanmar's Incitement Landscape Can Inform Platform Regulation in Situations of Mass Atrocity' (OpinioJuris, 2 January 2020) <http://opiniojuris.org/2020/01/02/how-myanmars-incitement-landscape-can-inform-platform-regulation-in-situations-of-mass-atrocity/> accessed 27 December 2022

[220] Robin Geiß and Henning Lahmann, 'Protecting the global information space in times of armed conflict' (International Review of the Red Cross, January 2022) <https://international-review.icrc.org/articles/protecting-the-global-information-space-in-times-of-armed-conflict-915> accessed 27 December 2022

used for planning military operations.[221] In the second case, such information operations would amount to direct participation in hostilities, impacting the course of the armed conflict. Accordingly, the persons responsible for their organisation and sharing will become the **legitimate targets of the attack**.

The historical examples in this regard include the operation of the RTLM in Rwanda, publications on Facebook instigating the violent clashes between the government and opposition forces in Ethiopia, and threats of ethnic violence spread through TikTok in Kenya.[222] Another example was the bombing of the Serbian State radio and television by NATO with a justification of countering military propaganda.[223] Similar cases can also be found throughout the World War I history, when undersea cables were destroyed to interrupt communications,[224] which was, among others, reasoned by a need to prevent propaganda and instigations from spreading by the belligerents. The real-life example, though, is the recent assassination of Russian propagandist Dugina, around which there was an extensive IHL debate regarding the journalists' participation in hostilities by virtue of delivering incitements to conduct warfare, commit war crimes of crimes against humanity. The researchers, in this respect, reached a conclusion that a journalist might become a legitimate target of an armed attack only when *"such person … is producing substantial and direct incitement to commit specific, serious, and unlawful 'violence' that is expected to occur relatively close in time"*.[225] It is needless to say that some of the Russian propagandists have already said enough to fall under this requirement.

Respectively, the physical facilities delivering propagandistic materials also might be destroyed when there is sufficient proof that it assists in conducting military operations or, for instance, is used for providing military orders.[226] Yet, if this threshold is not reached, any attack on broadcasting infrastructure will amount to a violation of the IHL since it will constitute an attack on a civilian object. Importantly, if within the IHL regime disinformation campaign is qualified as a part of an armed attack – it shall necessarily comply with all the standards applicable to armed attacks,[227] including the distinction, proportionality, precaution, military necessity and other principles. Finally, if individuals who share disinformation or control propaganda campaigns are taking a direct part in hostilities, the issue of their responsibility for the commission of crimes in this field also arises.

[221] Christiane Rexilius, 'Scenario 18: Legal status of cyber operators during armed conflict' (Cyber Law Toolkit, 3 March 2022 (edited) <https://cyberlaw.ccdcoe.org/wiki/Scenario_18:_Legal_status_of_cyber_operators_during_armed_conflict> accessed 27 December 2022

[222] ARTICLE 19, Response to the consultation of the UN Special Rapporteur on Freedom of Expression on her report on challenges to freedom of opinion and expression in times of conflicts and disturbances (19 July 2022) 8

[223] Alexandre Balguygallois, 'Protection des journalistes et des médias en période de conflit armé' (2004) 86(853) IRRC 52

[224] Lawrence T Greenberg, Seymour E Goodman and Kevin J Soo Hoo, Information Warfare and International Law (National Defense University Press, 1998) 9

[225] Charlie Dunlap, 'Law and the killing of a Russian propagandist: Some Q&A' (Lawfire, 9 October 2020) <https://sites.duke.edu/lawfire/2022/10/09/law-and-the-killing-of-a-russian-propagandist-some-q-a/> accessed 27 December 2022

[226] 'Russia, Ukraine & International Law: On Occupation, Armed Conflict and Human Rights' (HRW, 23 February 2022) <https://www.hrw.org/news/2022/02/23/russia-ukraine-international-law-occupation-armed-conflict-and-human-rights> accessed 27 December 2022

[227] Robin Geiß and Henning Lahmann, 'Protecting the global information space in times of armed conflict' (International Review of the Red Cross, January 2022) <https://international-review.icrc.org/articles/protecting-the-global-information-space-in-times-of-armed-conflict-915> accessed 27 December 2022

However, the qualification of the disinformation campaign as direct participation in hostilities does not automatically imply its lawfulness under the IHL. The information operation might (and usually does) breach the regulations prescribed under the Geneva Conventions and customary IHL. Qualifying disinformation as an armed attack is a highly disputable issue since such psychological reactions as *"[i]nconvenience, irritation, stress, fear are outside of the scope of the proportionality principle"*.[228] Thus, not every information operation will reach this threshold. The general rule, in this regard, is that journalists are protected as civilians under Article 79 of AP I to the Geneva Conventions. However, to violate the IHL standards, the party to the conflict shall not conduct armed attacks only. Some activities assisting in the conduct of a military operation may well constitute a breach, excluding the respective objects from the scope of protection provided to the civilian objects. The most outstanding breaches are described below:

- **Perfidy and ruses of war**. One of the rules directly applicable to false and manipulative behaviour in the course of warfare is the prohibition of perfidy, which is outlawed by Article 37(2) of AP I to the Geneva Conventions,[229] namely, the acts, misleading the counterparty, which have a consequence of the death, injury or capture of a person belonging to the adversary party.[230] The intent here is to betray confidence that another subject participates in hostilities and mislead as to the legal qualification of the circumstances or the status of a protected person.[231] In this respect, the perfidy was manifested in disinformation regarding the opening of green corridors from Mariupol, which further turned out to be false, while individuals willing to evacuate got under heavy shelling. A theoretical example might also involve a deepfake of the ICRC inviting the party to the peace talk, which in fact, is a manipulation with the intent to gain a military advantage.[232] Also, it may take the form of using false flags under the news on certain military operations, presenting the old news as new ones or using coordinated bot campaigns to create a visual image of numerous supporters of the occupying powers or their advantage.[233]

  A different standard implies acting under the ruses of war. The threshold for ruses of war is much lower, since such acts shall intend only to *"mislead the enemy or to induce enemy forces to act recklessly"*.[234] Accordingly, if a deepfake video of a President leads to soldiers leaving their positions and being shot to death or captured, such an activity shall be qualified as perfidy,

[228] Michael N Schmitt, Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (CUP, 2017) ["Tallin Manual 2.0"], Rule 113, para 5

[229] 'Article 37 - Prohibition of perfidy' (International Humanitarian Law Databases) <https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-37> accessed 27 December 2022

[230] Tallinn Manual 2.0, Rule 122, para 5

[231] Lawrence T Greenberg, Seymour E Goodman and Kevin J Soo Hoo, Information Warfare and International Law (National Defense University Press, 1998) 13

[232] Lisa M Cohen, 'The new era of disinformation wars: Does international humanitarian law sufficiently regulate the use of deep-fakes?' (Völkerrechtsblog, 30 November 2020) <https://voelkerrechtsblog.org/de/the-new-era-of-disinformation-wars/> accessed 27 December 2022

[233] Sarah Morris, 'Blog | Disinformation, Propaganda, and the War in Ukraine' (The Carter Center, 21 March 2022) <https://www.cartercenter.org/news/features/blogs/2022/disinformation-propaganda-and-war-in-ukraine.html> accessed 27 December 2022

[234] Tallinn Manual 2.0, Rule 123, para 2

running contrary to the IHL rules. For example, Russian attempts to depict Ukrainian President Zelensky calling its nation to surrender directly fall within the prohibited type of behaviour since it could potentially lead to the capture of the Ukrainian soldiers. In contrast, if no such consequences occur or might be anticipated, the ruses of war standards apply, remaining in the grey zone of the IHL.

Another potential misuse of information might be compromising and extorting civilians through the spreading of disinformation, *i.e.* amounting to blackmailing or manipulation by one's behaviour to gain military advantage, affecting the adversaries' will to resist.[235] The Russian side, for example, created fake accounts on Telegram, which resembled ones designed to report Russian military vehicles.[236] Since individuals sometimes misperceived Ukrainian units as Russian ones, the fake bot also gained information about the location of Ukrainian armed forces. Hence, fake bots assisted Russia in gaining a military advantage. Similarly, the dissemination of disinformation on the occupied territories may transform the attitude of the local population to occupying powers, which may lead to the assistance of locals, including providing the proviant and other necessary non-military supplies. As a result, a positive attitude of civilians will assist in conducting military operations, thus indirectly leading to a military advantage by virtue of spreading false information. Examples of such a technique can be found in Russian information strategies employed in the occupied Donbas region and the Crimean Peninsula.[237]

- **Terror of civilian population**. Disinformation can violate Article 51(2) of AP I to the Geneva Conventions, which outlaws the terror of civilian population, including *"[a]cts or threats of violence the primary purpose of which is to spread terror among the civilian population"* .[238] In this respect, scholars stress that this prohibition covers a real threat, which has psychological consequences for the population.[239] For instance, it may cover messages about future nuclear or biological attacks, forcing people to leave the territories chaotically and intervening in the course of the ongoing military operation. The deepfakes can depict the untrue dialogues between the President and the highest military commandership, discussing the possible attack (which, in fact, will never happen), thus making the population frustrated and frightened.[240] As a result,

[235] Robin Geiß and Henning Lahmann, 'Protecting the global information space in times of armed conflict' (International Review of the Red Cross, January 2022) <https://international-review.icrc.org/articles/protecting-the-global-information-space-in-times-of-armed-conflict-915> accessed 27 December 2022

[236] 'СБУ попереджає про створення фейкового чат-боту про переміщення військ' (НашіГрошіЛьвів, 26 April 2022) <https://lviv.nashigroshi.org/2022/04/26/sbu-poperedzhaye-pro-stvorennya-fejkovogo-chat-botu-pro-peremishhennya-vijsk/> accessed 27 December 2022

[237] Report of Rapporteur Boriss Cilevičs, Legal challenges related to hybrid war and human rights obligations (CoE Committee on Legal Affairs and Human Rights, 6 April 2018) <https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=24547&lang=en> accessed 27 December 2022

[238] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1125 UNTS 3, Article 51(2)

[239] Robin Geiß and Henning Lahmann, 'Protecting the global information space in times of armed conflict' (International Review of the Red Cross, January 2022) <https://international-review.icrc.org/articles/protecting-the-global-information-space-in-times-of-armed-conflict-915> accessed 27 December 2022

[240] Lisa M Cohen, 'The new era of disinformation wars: Does international humanitarian law sufficiently regulate the use of deepfakes?' (Völkerrechtsblog, 30 November 2020) <https://voelkerrechtsblog.org/de/the-new-era-of-disinformation-wars/> accessed 27 December 2022

another party will gain a military advantage.

At the same time, certain African States very persistently insisted on the extension of the notion of terror of civilians to various information activities, which are not limited purely by propaganda.[241] In this regard, an example might be disinformation regarding the inaction or failure of the administrative authorities in the territories, which are intended to be occupied, which may also force individuals to leave the area. In such a case, although not being explicitly dangerous to the life of a civilian, it still creates an atmosphere of fear, affecting the psychological dimension, to which the advocates of extensive interpretation of Article 51(2) usually refer. However, the threat shall be at least anticipated and relatively imminent for the speech to fall under this prohibition. Namely, it must be realistic and expected, as well as connected with a delivered material, otherwise, this norm remains inapplicable.

In this regard, Katz stresses that terror, although needing an attack or a threat of attack, shall lead to an injury, which also includes severe mental suffering.[242] From this perspective, the deepfake may well cause intense fear and intimidation, misleading as to a potential surrender or loss in the war. The important element is still the causation between the threat and feelings of fear among civilians.

- **Incitement to violation of the IHL**. Common Article 1 to the Geneva Conventions in line with Article 1(1) of the AP I to them proscribes the incitement to violation of the IHL. Particularly, these provisions oblige the State to *"respect and ensure respect"* to the IHL,[243] postulating that the States shall not encourage the third parties to resort to prohibited means of warfare, as well as shall not approve them as lawful. In short, this rule outlaws the illegal incitements, but in the dimension of the IHL. Importantly, the self-defence speech is not covered by the prohibition of propaganda for war,[244] yet the ban on incitement to violation of the IHL applies regardless of the party to the conflict. Similarly, in this regard, the UN stressed that propaganda for war, including the violations of the rules of war, usually goes in line with silencing speech on the occupied territories.[245] Accordingly, suppression of alternative voices and information isolation under certain circumstances might also be qualified as a breach of this prohibition, creating a desperate state of mind in the audience and posing it in an environment of hostility and aggression. The artificial videos showing the tortures of prisoners of war or civilian population by a counterparty, produced by the belligerent to justify its own violations and encourage similar

[241] Michael G Kearney, 'Propaganda in the Jurisprudence of the International Criminal Tribunal for the Former Yugoslavia' (2011) Routledge 1, 13

[242] Eian Katz, 'Liar's war: Protecting civilians from disinformation during armed conflict' (International Review of the Red Cross, December 2021) <https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914> accessed 27 December 2022

[243] Geneva Convention relating to the Protection of Civilian Persons in Time of War (entered into force 21 October 1950) 75 UNTS 287, Article 1; Zhu Wenqi, 'On Co-operation by States not Party to the International Criminal Court' (2006) 88 IRRC 87, 92

[244] Vivek Krishnamurthy, 'War Propaganda and International Law: A Conversation with Vivek Krishnamurthy' (Tech Policy Press, 18 March 2022) <https://techpolicy.press/war-propaganda-and-international-law-a-conversation-with-vivek-krishnamurthy/> accessed 27 December 2022

[245] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) 19 <https://www.osce.org/fom/203926> accessed 27 December 2022; UN General Assembly, Resolution 381 on Condemnation of propaganda against peace (15 December 1950)

treatment of prisoners and civilians among its armed forces[246] might as well be qualified as an incitement to violate the IHL. Moreover, faking the violations might trigger even graver crime commission,[247] up to genocide since the population becomes more and more receptive towards the harmful speech.

Another example of terror might include the encouragement of the dehumanizing attitude towards civilians among the armed forces of the aggressor State, which will further lead to hatred towards the local population. In this regard, scholars note that such an approach is a relatively common practice among governments,[248] willing to occupy the territories or orchestrate ethnic cleansing. As a result, the ordinary soldiers perceive the population in a demonized manner, committing more atrocities against it.

In the context of the Russian aggression, the Russian delegates try to justify aggression and atrocities by allegations of discriminatory treatment on the Ukrainian side.[249] Moreover, the research of the Ukrainian Public Broadcaster shows numerous speeches when State-affiliated media were allowing direct incitements to kill civilian population (for example, drowning children in the river etc).[250] Accordingly, since the media did not cease the violations, while the State never initiated investigations into such crimes, Russia is responsible for such breaches.

- **The principle of humanity**. Being enshrined in Article 27 of the IV Geneva Convention,[251] it requires the States to respect protected persons' honour, dignity, and human rights, treating them humanely. Any leaked data about the individuals participating in warfare violates the principle of human dignity, while posing such individuals in a false light endangers their life and health, making them the potential targets for the attacks.[252] For instance, in Syria, the personal data of soldiers has been stolen to further spread disinformation against them, threatening with violence to their relatives.[253] At the very last, there is even an idea of creating a Digital Geneva Convention, addressing all the issues of warfare in the digital age including the protection of vulnerable groups and rules on conducting warfare with a resort to the modern technologies.[254]

[246] 'In violation of the Geneva Convention relative to the Treatment of Prisoners of War, the occupiers used the labor of the Ukrainian military' (Ukrainian Parliament Commissioner for Human Rights, 20 May 2022) <https://www.ombudsman.gov.ua/en/news_details/upovnovazhenij-okupanti-vikoristovuyut-pracyu-ukrayinskih-vijskovih-ta-znimayut-z-nimi-propagandistski-video> accessed 27 December 2022

[247] Lawrence T Greenberg, Seymour E Goodman and Kevin J Soo Hoo, Information Warfare and International Law (National Defense University Press, 1998) 12-13

[248] Elizabeth Willmott-Harrop, 'Iraq: propaganda's war on human rights' (Liberty & Humanity, June 2005) <https://libertyandhumanity.com/themes/international-human-rights-law/iraq-propagandas-war-on-human-rights/> accessed 27 December 2022

[249] 'Warning incitement of racial, religious hatred can trigger atrocity crimes, Special Adviser stresses states' legal obligation to prevent genocide' (ReliefWeb, 21 June 2022) <https://reliefweb.int/report/ukraine/warning-incitement-racial-religious-hatred-can-trigger-atrocity-crimes-special-adviser-stresses-states-legal-obligation-prevent-genocide> accessed 27 December 2022

[250] Suspilne News, 'Російська пропаганда: як вона вбиває і чи можливо судити пропагандистів Кремля' (YouTube, 23 December 2022) <https://www.youtube.com/watch?v=O4WuEJu-36k> accessed 27 December 2022

[251] Geneva Convention relating to the Protection of Civilian Persons in Time of War (Fourth Geneva Convention) (adopted 22 August 1864, entered into force 22 July 1865) 75 UNTS 287, Article 27

[252] 'Disinformation endangering Red Cross work in Ukraine: ICRC' (RFI, 29 March 2022) <https://www.rfi.fr/en/disinformation-endangering-red-cross-work-in-ukraine-icrc> accessed 27 December 2022

[253] Patrick Howell O'Neill, 'Why the Syrian uprising is the first social media war' (Daily dot, 18 September 2013) <https://www.dailydot.com/debug/syria-civil-social-media-war-youtube/> accessed 27 December 2022

[254] Joseph Guay and Lisa Rudnick, 'What the Digital Geneva Convention means for the future of humanitarian action' (The Policy Lab, 25 June 2017) <https://www.unhcr.org/innovation/digital-geneva-convention-mean-future-humanitarian-action/> accessed 27 December 2022

With regard to the specific types of violations when disinformation is employed, the depiction of victims in a false light or neglecting their suffering directly breaches the principle of humanity.[255] In this regard, the Russian depiction of civilian victims after the shelling of Mariupol as the *"puppet theatre organised by Ukraine" diminishes* the victims' dignity, thus being a clearcut violation of the IHL. Similarly, disinformation regarding the availability and quality of medical care also is considered to be a violation of the humanity standard.[256]

Accordingly, there is a set of norms under the IHL which indicates that dissemination of disinformation might serve as a violation of international law if it reaches a certain threshold of gravity, with the grave consequences being highly anticipated or imminently turning into reality. Having both conventional and customary manifestations, these rules provide the international community with a vast variety of options to bring the responsible States to justice. It also provides the international organisations and individual States, which do not participate in the armed conflict, with an opportunity to develop due responses and prevent the violations of the fundamental principles.

[255] Eric Jensen and Sean Watts, 'Ukraine Symposium – Doxing Enemy Soldiers and the Law of War' (Articles of War, 31 October 2022) <https://lieber.westpoint.edu/doxing-enemy-soldiers-law-of-war/> accessed 27 December 2022

[256] Eian Katz, 'Liar's war: Protecting civilians from disinformation during armed conflict' (International Review of the Red Cross, December 2021) <https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914> accessed 27 December 2022

Conditions for bringing Russia as an aggressor State in the international armed conflict against Ukraine to justice for spreading disinformation:

I. A line of international treaties to which Russia is a party can be invoked by the relevant actors before the competent international tribunals:

- The International Convention concerning the Use of Broadcasting in the Cause of Peace can be brought up by one of its Parties before the ICJ by contesting the USSR's reservation to jurisdiction and claiming the breach of Articles 2-4 of the Convention;
- The CERD can be invoked by Ukraine before the ICJ if the proceedings in the CERD Committee will be unsatisfactory regarding discriminatory speech and illegal incitements based on ethnic and national origin and to additionally determine the number of reparations;

- The Genocide Convention can be invoked by Ukraine before the ICJ regarding the incitement to genocide committed by Russian State authorities and organs whose actions are attributable to Russia as a State.

II. Disinformation campaigns are attributable to Russia under Articles 4, 5, 7 and 11 of the ARSIWA, all of which, having customary nature, serve as the widely recognised rules on attribution. Particularly, Russian State organs both directly incited violence, war and genocide in Ukraine and failed to exercise due diligence in preventing and punishing such activities by other actors. The same approach was maintained concerning the behaviour of the State organs and officials who exceeded their powers. Furthermore, Russian authorities frequently acknowledged and adopted disinformation, which breaches international law, as its own, promulgating and popularising such speech.

III. Russian disinformation constitutes a breach of sovereignty, and a cause of the armed conflict as well amounts to the incitement to genocide and numerous violations of the IHL (perfidy, terror of civilian population, incitement to the IHL violations, and a breach of humanity principle).

III. International criminal law: who pulls the trigger of disinformation?

Following the end of hostilities, the State responsibility mechanisms are best suited for further reparation of damages, redressing the harm experienced by individuals and compensating for the losses of the defending State. However, according to the IMT, they cannot meaningfully perform the preventive and penalising functions since *"crimes against international law are committed by men, not abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced"* .[257] Subsequently, the history of legal regulation of criminal liability for illegal propaganda comprised various documents – starting from the 1954 Draft Code of Offences Against Peace and Security of Mankind, the 1996 Draft Code of Crimes Against Peace and Security of Mankind,[258] and, finally, the Rome Statute. And this is even without mentioning the statutes of the specialised criminal tribunals established for the specific cases of atrocities, which, in some instances, also addressed the issues of propaganda.

Nevertheless, most of the mentioned documents and their preparatory works explicitly mentioned that propaganda for war constitutes the preparation for the war and thus is punishable in the dimension of the ICL.[259] Importantly, even back in 1954, there were proposals to likewise outlaw the incitements to any criminally punishable actions.[260] In the 1984 report of the Special Rapporteur of the working group for the Draft Code of Crimes Against Peace and Security of Mankind, though, there even was an attempt to criminalise false information deliberately distributed to harm international relations (although this attempt remained unsuccessful at that time).[261] However, all the efforts were mostly rejected in view of the lack of the State consensus on the adoption of the document, which provides for such an extensive and all-encompassing liability framework.

Nevertheless, the first steps towards criminalising the speech (apart from the incitement to genocide) were made in the middle of 20th century. Even where the voices supporting such novelties were not strong enough to end up with adding specific formulations to the statutes and codes, they still significantly impacted the judicial practice. Further analysis of the legacy of criminal tribunals will show that propaganda, including by virtue of disinformation, has been repeatedly addressed by the international courts dealing with post-conflict situations. Before moving forward to the application of criminal punishment by the specialised tribunals,

[257] Ameer F Gopalani, 'The International Standard of Direct and Public Incitement to Commit Genocide: An Obstacle to U.S. Ratification of the International Criminal Court Statute?' (2001) 32(1) California Western International Law Journal 87, 93

[258] Kearney 194-205

[259] Ibid, 196

[260] Draft Code of Offences Against the Peace and Security of Mankind [1951] 2 YB ILC 112

[261] Second report on the Draft Code of Offences against the Peace and Security of Mankind by Mr Doudou Thiam, Special Rapporteur [1984] UN Doc A/CN.4/377, para 74

one shall think of how to reach that tribunal and what issues might serve as a legal obstacle for criminal liability to occur.

Immunities. One of the main challenges to institute criminal proceedings against the incumbent high-ranking State officials on the international level is their functional immunity from criminal jurisdiction and inviolability.[262] Namely, they cannot be prosecuted for any activity done in the official capacity, while they are occupying their official post. Akin to that, certain highest State officials, including the President, the Minister for Foreign Affairs and the Prime Minister, enjoy immunity *ratione personae*, which extends to acts performed in their private capacity before occupying a post.[263] Yet, personal immunity is applicable only and exclusively when the person serves in these roles. Subsequently, after leaving a post, this person can still be criminally penalised for a number of grave crimes. In *Arrest Warrant*, the ICJ clarified that such type of immunity relates to cases involving allegations of war crimes or crimes against humanity.[264]

The waiver of the immunity by the State is the only instance when proceedings over high-ranking State officials for their private acts becomes possible.[265] Otherwise, the States are forced to wait for the resignation of the person from the office, which, however, happens very rarely in authoritarian States. Obligation to respect the immunities, in turn, burdens the States with a duty to prevent any attack on the foreign State officials, especially in the form of arrest or detention.[266] As a result, any action brought against an individual who enjoys immunity should be dismissed.[267] Although these rules were primarily designed to shield individuals from criminal prosecution before the foreign domestic courts, a dispute regarding their applicability before the international courts is still ongoing.

> In *Arrest Warrant,* the ICJ explicitly stated that highest State officials *"may be subject to criminal proceedings before certain international criminal courts,* ***where they have jurisdiction****"*.[268]

The only international court, dealing with the ICL and being unlimited to a specific situation, is the ICC, which bases its work on the provisions of the Rome Statute. Article 27 of this treaty *"applies equally to all persons without any distinction based*

[262] Arrest Warrant, para 51; Certain Questions of Mutual Assistance in Criminal Matters (Djibouti v France) (2008) ICJ Rep 177 ["Djibouti v France"], para 170; Robert Cryer et al., An Introduction to International Criminal Law and Procedure (2nd edn, CUP, 2010) 534

[263] Dapo Akande and Sangeeta Shah, 'Immunities of State Officials, International Crimes, and Foreign Domestic Courts' (2011) 21 EJIL, 819; Preliminary Report on Immunity of State Officials from Foreign Criminal Jurisdiction by Special Rapporteur Kolodkin (60th session, 2008) UN Doc A/CN.4/601, para 79

[264] Arrest Warrant, para 58

[265] Ibid, para 61; Third Report on Immunity of State Officials from Foreign Criminal Jurisdiction by Special Rapporteur Kolodkin (63rd session, 2011) UN Doc A/CN.4/646, para 32

[266] Djibouti v France, para 174

[267] US v Khobragade (2014) 15 F.Supp.3d 383, 387

[268] Arrest Warrant, para 61

on official capacity", insofar as the States Parties are concerned.[269] However, the Prosecutor may initiate investigation into any situation and bring the case to the ICC, while only the issue of the arrest warrant's enforcement will remain. In this regard, Article 89 of the Rome Statute provides that the ICC may transmit a request for arrest and surrender together with the supporting material *"to any State on the territory of which that person may be found"*.[270] The State Party must immediately comply with it,[271] and promptly bring the arrested person before its domestic judicial authority in accordance with the proper process.[272] Article 98 of the Rome Statute, obliging the ICC to abstain from requests for surrender which would require the State to act inconsistently with its obligations under law on immunities, is called a *"procedural device to avoid a conflict of duties"*,[273] which does not affect the State's obligations to comply with the ICC request in practice.

In this respect, there is a legal presumption that immunity cannot shield the State representatives from the jurisdiction of the international tribunal, in contrast to the national jurisdictions. This, *inter alia*, was proven by the position of the ICJ in *Arrest Warrant*, where the Court made a distinction between absolute immunity from foreign jurisdiction and the situation when an international criminal tribunal exercises its jurisdiction over the Minister.[274] Analogous finding was made by the SCSL in *Taylor* case.[275] Even more, scholars prove that such a rule possesses a customary nature:

- **State practice**. There are numerous examples when the States behaved in a manner as if the immunity from the jurisdiction of the international courts does not exist unless it is expressly recognised in a specific treaty. It is evidenced, *inter alia*, by Liberia's acceptance of proceedings over its president before the SCSL, which, similar to the ICC, is a treaty-based tribunal.[276] Likewise, the US, being a non-party to the Rome Statute, has concluded at least 45 bilateral 'non-surrender' agreements with the States Parties to the ICC, extending immunities to its officials before the court,[277] showing its belief that the Rome Statute would operate against its officials should they be brought before the ICC. Another exemplary case is the prosecution of Al-Bashir. Namely, lots of the ICC States Parties, including Botswana,[278] Central African Republic,[279] France,[280]

[269] Rome Statute of the International Criminal Court, 2187 UNTS 38544 ["Rome Statute"], Article 27
[270] Rome Statute, Article 89
[271] Otto Triffterer, The Rome Statute of the International Court: A Commentary (3rd edn, Beck/Hart, 2016) 1462
[272] Rome Statute, Article 59; Prosecutor v Jean-Bosco Barayagwiza, AC Decision (1999) ICTR-97-19-AR72, para 70
[273] Claus Kreß, 'The International Criminal Court and Immunities under International Law for States Not Party to the Court's Statute' (2012) 15 FICHL SSICL 223, 233
[274] Arrest Warrant, para 61
[275] Prosecutor v Charles Ghankay Taylor, Decision on immunity from jurisdiction (2004) SCSL-2003-01-1, para 52
[276] Hugh King, 'Immunities and Bilateral Immunity Agreements: Issues Arising from Articles 27 and 98 of The Rome Statute' (2006) 4 NZJPIL 269, 280
[277] Boucher R., 'US Signs 100th Article 98 Agreement' (2005) <https://2001-2009.state.gov/r/pa/prs/ps/2005/45573.htm> accessed 27 December 2022
[278] 'Bashir Travel Map: Botswana' <http://bashirwatch.org/#botswana> accessed 27 December 2022
[279] 'Bashir Travel Map: Central African Republic' <http://bashirwatch.org/#central-african-republic> accessed 27 December 2022
[280] 'Bashir Travel Map: France' <http://bashirwatch.org/#france> accessed 27 December 2022

Kenya,[281] Malawi,[282] and the EU[283] expressed their commitment to arrest Al-Bashir if he visited their territories. Non-compliance of certain States with the ICC's request to arrest and surrender Al-Bashir does not affect the crystallisation of a customary exception.[284] For instance, South Africa, which refused to arrest Al-Bashir, was one of a few States which, regardless of consequences, did not sign a bilateral agreement with the US extending immunities to its officials before the ICC.[285] Nevertheless, in other cases, it acted as if the customary rule existed.

- ***Opinio juris***. It is inferred from the States' acts or omissions insofar as they are done following *"a belief that this practice is rendered obligatory by the existence of a rule of law"*.[286] The mentioned behaviour of Liberia, the US, Botswana, the Central African Republic, France, Kenya, Malawi and the EU are indicative of a firm belief that there is an exception from immunity *ratione personae* before the ICC. Similarly, the ICC consistently confirmed that Al-Bashir did not enjoy immunities *vis-à-vis* the ICC under customary international law,[287] and explained that by performing the ICC's requests, the State Party exercises jurisdiction of the ICC,[288] acting *"on behalf of the international community as a whole"*.[289] Accordingly, immunity *ratione personae* is not available before the international tribunal, such as the ICC, when it is validly seized with the matter under Articles 12-14 of the Rome Statute,[290] *i.e.* if a crime was committed at the territory of the State Party or by its national.

Another option implies either a waiver of immunity or the cases when non-party to the Rome Statute is ordered by the UN Security Council to cooperate fully with the ICC under Chapter VII of the UN Charter.[291] This is hardly possible in cases against the permanent members of the UN Security Council who can apply the veto right to any decision which contravenes their interests. It is also hardly possible until the authoritarian regime changes following the inner transformations inside the

[281] Claus Kreß, 'The International Criminal Court and Immunities under International Law for States Not Party to the Court's Statute' (2012) 15 FICHL SSICL 223, 259

[282] Ibid

[283] 'Declaration by the Presidency on behalf of the European Union following the ICC Decision concerning the Arrest Warrant for President Al-Bashir' (2009) <https://ec.europa.eu/commission/presscorner/detail/en/PESC_09_29> accessed 27 December 2022

[284] Claus Kreß, 'The International Criminal Court and Immunities under International Law for States Not Party to the Court's Statute' (2012) 15 FICHL SSICL 223, 259

[285] Max Du Plessis, Stephen Pete, 'Who Guards the Guards?: The International Criminal Court and Serious Crimes Committed by Peace Keepers in Africa' (2006) ISSMS No 121, 32

[286] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America) (1986) ICJ Rep 14, para 207

[287] Al-Bashir, Judgment in the Jordan Referral re Al-Bashir Appeal (2019) ICC-02/05-01/09-397 ["Jordan Appeal"], paras 1, 50; Al-Bashir, Decision pursuant to article 87(7) of the Rome Statute on the refusal of the Republic of Chad to comply with the cooperation requests issued by the Court with respect to the arrest and surrender of Omar Hassan Ahmad Al-Bashir (2011) ICC-02/05-01/09-140, para 43; Al-Bashir, Corrigendum to the Decision pursuant to article 87(7) of the Rome Statute on the failure by the Republic of Malawi to comply with the cooperation requests issued by the Court with respect to the arrest and surrender of Omar Hassan Ahmad Al-Bashir (2011) ICC-02/05-01/09-139-Corr, para 43

[288] Al-Bashir, Corrigendum to the Decision pursuant to article 87(7) of the Rome Statute on the failure by the Republic of Malawi to comply with the cooperation requests issued by the Court with respect to the arrest and surrender of Omar Hassan Ahmad Al-Bashir (2011) ICC-02/05-01/09-139-Corr, para 46; Mark Klamberg, Commentary on the Law of the International Criminal Court (TOAE, 2017) 667

[289] Jordan Appeal, para 115

[290] Al-Bashir, Corrigendum to the Decision pursuant to article 87(7) of the Rome Statute on the failure by the Republic of Malawi to comply with the cooperation requests issued by the Court with respect to the arrest and surrender of Omar Hassan Ahmad Al-Bashir (2011) ICC-02/05-01/09-139-Corr, para 36; Michiel Blommestijn and Cedric Ryngaert, 'Exploring the Obligations for States to Act upon the ICC's Arrest Warrant for Omar al-Bashir' (WP No 48, 2010) 13

[291] Jordan Appeal, para 7

State concerned. Although Ukraine has not yet ratified the Rome Statute, it has an agreement on the *ad hoc* jurisdiction of the Court over the Maidan events, as well as over the crimes committed in the course of armed conflict.[292] In this regard, the Office of the Prosecutor General has started an official investigation in the atrocities committed by Russian armed forces in the territory of Ukraine,[293] collecting evidence for further trials. Thus, the jurisdiction of the court can be established based on the principle that crimes have been committed in a territory of the State recognising the jurisdiction of the Court.

Even if the ICC as a judicial forum is unavailable, the possibilities of bringing responsible individuals to justice remain on the table. Unlike the ICC, which is a treaty-based judicial body, the ICTY and ICTR removed the head of State immunity for all UN members.[294] For example, the statutes of the special criminal tribunals do not incorporate the clause on immunities or rather incorporate a provision on their automatic waiver before the court. As a result, the responsible individuals can be legitimately brought to justice. Moreover, the Rome Statute still does not fully cover disinformation issues as a separate crime, viewing them rather as an element of the other more 'physical' crimes. Therefore, for the purposes of prosecuting Russian illegal incitements, instigation to aggression and war crimes, a special criminal court looks like a more suitable and legally appropriate venue.

Lastly, common Article 1 of the Geneva Conventions contains an obligation to *"respect and ensure respect"* for the IHL,[295] which reflects customary international law.[296] In this regard, Article 49 of the I Geneva Convention provides that the parties thereto should bring to justice persons alleged to have committed grave breaches of the IHL,[297] which, according to the Rome Statute, is covered by a notion of war crimes.[298] Hence, the parties to the Geneva Conventions should at least *"make an effort not to block actions taken by the ICC to punish or prevent serious violations of the Geneva Conventions"*.[299] However, this breach would be put in the dimension of the State responsibility given the applicability of the Geneva Conventions to the States rather than private individuals. Hence, although Russia might be in a potential violation of it, if the ICC provides an arrest warrant against its officials, it will not be the matter of the ICL exclusively.

[292] Олена Мошенець, 'Міжнародний кримінальний суд – шлях до Гааги та репарацій' (LB.ua, 26 August 2022) <https://lb.ua/blog/olena_moshenets/527492_mizhnarodniy_kriminalniy_sud-shlyah.html> accessed 27 December 2022

[293] Ірина Венедіктова, 'Надія на Гаагу: чому Міжнародний кримінальний суд життєво важливий для України і світу' (Українська правда, 1 July 2022) <https://www.pravda.com.ua/columns/2022/07/1/7355729/> accessed 27 December 2022

[294] Hugh King, 'Immunities and Bilateral Immunity Agreements: Issues Arising from Articles 27 and 98 of the Rome Statute' (2006) 4 NZJPIL 269, 279

[295] Geneva Convention relating to the Protection of Civilian Persons in Time of War, 75 UNTS 287, Article 1; Zhu Wenqi, 'On Co-operation by States not Party to the International Criminal Court' (2006) 88 IRRC 87, 92

[296] Zhu Wenqi, 'On Co-operation by States not Party to the International Criminal Court' (2006) 88 IRRC 87, 93

[297] Geneva Convention relating to the Protection of Civilian Persons in Time of War, 75 UNTS 287, Article 49

[298] Rome Statute, Articles 8(2)(c), 8(2)(e)

[299] Geneva Convention relating to the Protection of Civilian Persons in Time of War, 75 UNTS 287, Article 1; Zhu Wenqi, 'On Co-operation by States not Party to the International Criminal Court' (2006) 88 IRRC 87, 94

Post-World War II trials. The military tribunals following World War II were emblematic of military propaganda and propaganda for war prosecution, as well as developing the key approaches to an individual criminal liability for the gravest crimes. The international courts, namely the IMT and the IMT for the Far East, were specifically designed to review the cases of German and Japanese leaders with the statutes being drafted by the international community. Some scholars also consider them to be the first ever trials dealing with ideological aggression.[300] Akin to that, some individuals trying to escape responsibility were apprehended in other countries, being prosecuted by the domestic courts of the US, the UK and several other States. Despite the ICL being in a rather inchoate state, these criminal processes addressed the speech crimes in parallel with the crimes related to physical violence and other atrocities:

- **IMT**. The Statute of this tribunal explicitly mentions that the "*leaders, organizers, instigators and accomplices participating in the formulation or execution of a common plan or conspiracy to commit any or the foregoing crimes are responsible*".[301] Thus, the criminal responsibility extended even to individuals who were inciting the commission of the crime. Meanwhile, the liability ranged depending on the gravity of the offences and the status of the individuals or, more specifically, one's impact on the decision-making processes. Thus, it basically did not exclude the possibility for ordinary perpetrators to be held liable with a pure difference in the severity of the punishment received. Within the Nuremberg process, two cases are symbolic with regard to the propaganda issues: trials over Streicher and Fritzsche. While considering their cases, the IMT also paid attention to the fact that they targeted not only German, but also foreign audiences,[302] which contributed to the gravity of the crimes.

  Julius Streicher ran an anti-Semitic weekly newspaper called 'Der Stürmer', where he has been *"infect[ing] the German mind with the virus of anti-Semitism"* for 25 years.[303] During the trial 23 articles were presented as evidence of the calls to the extermination of Jews, dehumanising speech and illegal propaganda for aggressive war, which continued even when the mass extermination was ongoing.[304] As a result, the IMT established that Streicher's expressions amounted to persecution based on ethnic origin,[305] convicting him for crimes against humanity. Scholars note that the assessment was relatively unproblematic for a tribunal,[306] since the expressions were the clearcut *prima facie* calls for extermination, while the printed evidence did not give any opportunity for mistakes in the specific wording.

[300] John B Whitton, 'Hostile International Propaganda and International Law' (1971) 398 SPI, 14-25

[301] Charter of the International Military Tribunal (established 8 August 1945), Article 6

[302] John B Whitton, 'Hostile International Propaganda and International Law' (1971) 398 SPI 14, 20

[303] Tommi Aromäki, 'The International Criminal Court's Jurisdiction Over Incitement to Genocide in the Internet Era - Some special situations' (University of Helsinki, 2021) 12

[304] Ibid

[305] Prosecutor v Julius Streicher (Trial Transcript, 29 April 1946) <https://avalon.law.yale.edu/imt/04-29-46.asp> accessed 27 December 2022

[306] N R Okany and J Hoffmann, Taking the Prevention of Genocide Seriously: Media Incitement to Genocide Viewed in the Light of the Responsibility to Protect in J Hoffmann and A Nollkaemper (eds), Responsibility to Protect: From Principle to Practice (Amsterdam University Press, 2012) 325

The case of **Hans Fritzsche** was slightly different, with him serving as a well-known radio commentator and head of the Home Press Division of the Ministry of Popular Enlightenment and Propaganda. In his official capacity, he supervised the German press of that time and instructed the publishers to highlight *"the Jewish problem"*, as well as share Nazi propaganda. Although he sometimes made propagandistic statements in his broadcasts, in contrast to Streicher, he was not convicted of any crime by the IMT.[307] Specifically, the tribunal found no intent to incite Germans to commit the atrocities but only to express excitement with and sentiment in support of Hitler, a popular politician of the time. Thus, falsifying news to arouse in the German population a certain passion to commit war crimes could not be established with the same clarity as in Streicher's case.[308] Here, the Dissenting Opinion of the Soviet Union Judge Nikitchenko is particularly interesting, calling the *"dissemination of provocative lies and the systematic deception of public opinion … as necessary to the Hitlerites for the realisation of their plans as … the production of armaments"* .[309] He also stressed that without propaganda, it would be impossible to realise the aggressive intentions of German Fascism. Akin to Streicher and Fritzsche, **Rudolf Hess, Wilhelm Keitel,** and **Alfred Rosenberg** were convicted of spreading Nazi propaganda with Rosenberg being recognised as a chief ideologist of the Nazi Party.[310] Yet, the guilt in their cases was established based on the commission of the other substantive crimes, while propagandistic activities were assessed as supplementary evidence and rather contributing factors. Nevertheless, these examples are important for understanding that the highest State officials may well be charged with propaganda even if they committed graver crimes. Thus, it shall not be neglected during the qualification of the criminal activities committed.

Another set of proceedings, which remains important from the legal perspective, although not directly related to the media or free speech topic, is **I.G. Farben case**. The IMT, in particular, found that within the notion of complicity, the attribution can be established between the *"parts of what each individual defendant knew [and] a unified whole on the part of the company"*, which can be traced through *"corporate minutes, transactional records, and aggressive cross-examination"* .[311] In this respect, scholars propose to establish a link between the contribution of the large platforms and media to the armed conflict and the knowledge of the criminal nature of acts on behalf of the individuals participating in such companies.[312] In the

[307] Prosecutor v Hans Fritzsche (Trial Transcript, 29 April 1946) <https://avalon.law.yale.edu/imt/06-28-46.asp> accessed 27 December 2022

[308] N R Okany and J Hoffmann, Taking the Prevention of Genocide Seriously: Media Incitement to Genocide Viewed in the Light of the Responsibility to Protect in J Hoffmann and A Nollkaemper (eds), Responsibility to Protect: From Principle to Practice (Amsterdam University Press, 2012) 325

[309] Predrag Dojcinovic, 'Propaganda in the Jurisprudence of the International Criminal Tribunal for the Former Yugoslavia' (2011) Routledge 1, 4

[310] Richard B Collins, 'Propaganda for War and Transparency' (2010) 87(4) Denver Law Review 819, 821

[311] Michael J Kelly, Atrocities by Corporate Actors: A Historical Perspective, (2018) 50 W.RES.J.INT'L L. 49, 76-77

[312] Jenny Domino, 'Crime as Cognitive Constraint: Facebook's Role in Myanmar's Incitement Landscape and the Promise of International Tort Incitement Landscape and the Promise of International Tort Liability Liability' (2020) 52(1) Case Western Reserve University School of Law 143, 177-180

context of Russian aggression, it might be relevant for the case of TV channels, spreading propagandist materials, especially those directly controlled by the State. Applying the rule from *I.G. Farben* case, one might reach a conclusion that at least the administration of the channel shall be held liable for instigation and aiding to international crimes commission.

- **IMT for the Far East**. Its Statute, similar to the IMT Statute, covers any type of participation in the crime, including the instigation, complicity, planning and conspiracy akin to the actual commission of a crime.[313] As well as in Nuremberg processes, the crimes varied depending on the gravity and the impact of an individual. One of the examples, in this respect, was a **process over Shūmei Ōkawa**, who advocated the occupation of the islands of the East Indies, and *"had predicted a war between East and West in which Japan would emerge as the champion of the East"*.[314] Particularly, Ōkawa served as an ideologist of Japanese militarism and was accused of conspiracy in the breach of laws and customs of war, as well as preventing the Japanese authorities from taking steps to cease the violations. In fact, he prepared the Japanese society for war and inspired it for aggressive expansion.[315] At the end of the day, he was found unfit to stand his trial and be brought to justice,[316] thus charges against Ōkawa were dropped. However, two factors are important with regard to this trial. Firstly, among 29 individuals being charged at the trial,[317] at least one person was charged purely for spreading propagandistic, hateful and aggressive messages. Secondly, this person was ascribed to the criminals of the Class A,[318] which covered the gravest crimes committed throughout World War II.

  In contrast to Ōkawa, five other individuals were indeed found guilty of conspiracy based predominantly, but not exclusively, on their propaganda activities. For instance, **Sadao Araki** was recognised as a chief propagandist encouraging Japanese people to aggressive war from 1928, while **Koichi Kido** was prosecuted for his activities as the Minister for Education.[319] Although in these cases no speech assessments were made, they are interesting from the perspective of criminal liability for coordination of the propagandistic campaigns.
- **Domestic proceedings**. Subsequent to the cessation of hostilities and the end of World War II, the US authorities held several trials over the individuals who were not prosecuted by the IMT or the IMT for the Far East. One of such persons was American poet **Ezra Pound**, who openly supported the ideas of Benito Mussolini and encouraged the aggressive policies of Italy and Germany.[320] In this

[313] Statute IMT for the Far East, Art 5

[314] International Military Tribunal for the Far East (Judgement of 4 November 1948) 67 <https://www.legal-tools.org/doc/8bef6f/pdf/> accessed 27 December 2022

[315] Ibid, 75

[316] 'Tokyo War Crimes Trial' (WWII) <https://www.nationalww2museum.org/war/topics/tokyo-war-crimes-trial> accessed 27 December 2022

[317] Dariia Lytoshenko, 'What awaits Putin: History of international trials of war criminals' (Ukrainer, 20 September 2022) <https://ukrainer.net/what-awaits-putin/> accessed 27 December 2022

[318] 'Ōkawa Shūmei' (Britannica, 20 December 2022) <https://www.britannica.com/biography/Okawa-Shumei> accessed 27 December 2022

[319] Richard B Collins, 'Propaganda for War and Transparency' (2010) 87(4) Denver Law Review 819, 822

[320] Conrad L Rushing, '"Mere Words": The Trial of Ezra Pound' (1987) 14(1) University of Chicago Press Journal 111, 111

respect, it is interesting to observe that even a claimed defence of an artistic speech could not lift the charges given the expressly propagandistic nature of his works, praising and justifying the warfare launched in Europe. However, this case, similar to Ōkawa trial, ended up with charges being dropped due to Pound's insanity and inability to stand a trial.[321] Further, upon the numerous requests of his artistic fellows, he was even released from the hospital, being let to freely move to Italy.

Another example is a trial over American citizen **Iva Toguri D'Aquino**, a famous propagandist acting under the name of 'Tokyo Rose'. Particularly, she tried to demoralise the American soldiers by highlighting the US losses during the war, as well as delivered propaganda on the Radio Tokyo on the Zero Hour program.[322] Though, in this situation, she was convicted not of the propaganda for war or related crimes, but became one of seven American citizens who were labelled as committing treason. Yet, almost 20 years after the trial, it was discovered that many witnesses resorted to perjury following the threats from the US side, which led to granting of the presidential pardon.[323] Moreover, as it turned out, the nickname of 'Tokyo Rose' was used by at least 11 other women.[324] Nevertheless, the case still remains emblematic of how the broadcasts and propagandistic materials during wartime are treated under the criminal law and, what is even more important, how much attention the States pay to the prosecution of the individuals spreading propaganda.

A similar conviction for treason was also attributed to **William Joyce**, more known as 'Lord Haw-Haw', who actively broadcasted Nazi propaganda. In contrast to Iva Toguri D'Aquino, his activities indeed constituted illegal incitements and radically discriminatory speech towards the Jewish population, thus providing substantial grounds for a criminal conviction.[325]

Lastly, the trial over **Otto Dietrich** (known as ***the Ministries* case**) is worth mentioning in the context of the coordination of illegal propaganda activities. Although being not as popular from the legal perspective as the trials over Streicher and Fritzsche, it is emblematic of how even vague and indirect incitements might lead to criminal prosecution. Basically, Dietrich's role in the propaganda machine implied the control and editorial review of the Nazi Party's Reich Press Office.[326] Finally, he also was appointed to the position of the State Secretary at the Propaganda Ministry, thus coordinating the propaganda campaigns,[327] including those aimed at the extermination of Jews. He was prosecuted after the IMT ended, appearing before the US-organised court

[321] 'Federal court decides to release poet Ezra Pound from hospital for criminally insane' (History, 15 April 2021) <https://www.history.com/this-day-in-history/federal-court-decides-to-release-ezra-pound> accessed 27 December 2022

[322] 'Iva Toguri D'Aquino and "Tokyo Rose"' (FBI Records) <https://www.fbi.gov/history/famous-cases/iva-toguri-daquino-and-tokyo-rose> accessed 27 December 2022

[323] Evan Andrews, 'How 'Tokyo Rose' Became WWII's Most Notorious Propagandist' (History, 20 January 2017) <https://www.history.com/news/how-tokyo-rose-became-wwiis-most-notorious-propagandist> accessed 27 December 2022

[324] Richard B Collins, 'Propaganda for War and Transparency' (2010) 87(4) Denver Law Review 819, 822

[325] Gemma R Birnbaum, 'The Capture and Execution of William Joyce' (WWII, 1 January 2021) <https://www.nationalww2museum.org/war/articles/william-joyce-capture-and-execution> accessed 27 December 2022

[326] Alexander G Hardy, Hitler's Secret Weapon: The "Managed Press" And Propaganda Machine of Nazi Germany (Vantage Press, 1968) 50

[327] Ibid, 66

in line with twenty other ministerial defendants.[328] Following the "self-made" trial, Dietrich was convicted with participation in the criminal organisation and crimes against humanity.[329] Interestingly, the conviction required that each defendant knew Hitler's plans for aggressive war and the commission of specific crimes against humanity and peace.[330] Lack of evidence regarding such knowledge in relation to some crimes led to dropping of some charges, although the judgements assumed their validity.

The ICTY. Following mass atrocities of World War II and subsequent trials over the perpetrators, the ICL was filled with a framework for prosecution of the gravest crimes. This framework, however, was imperfect, which can be seen from the scope and detalisation of the IMT Statute (a vague and overly generalistic document). Thus, when the Yugoslav Wars took place, a need to develop a better mechanism arose from both procedural and substantive points of view. Firstly, the ICTY Statute was based on the UN Security Council Resolutions,[331] which ensured the respect towards sovereignty in accordance with the UN legal framework. Accordingly, no more issues of the "trial of victors" could be claimed. Secondly, the Statute itself provided a relatively clear and precise definitions of crimes, prohibited conduct and the applicable penalties. To clarify, Articles 2 to 5 addressed the grave breaches of the Geneva Conventions, violations of laws of war, genocide and crimes against humanity. As regards genocide, a direct and public incitement was outlawed in line with the conspiracy to commit this crime. Yet, no incitement to other grave crimes was addressed as a separate prohibition. Article 7, though, was also devoted to a scope of liability, clarifying that the ICTY could try individuals, who *"planned, instigated, ordered, committed or otherwise aided and abetted in the planning, preparation or execution of a crime"*.[332] As the Tribunal's practice shows, these modes of liability were employed in all cases concerning the speech crimes, since a separate qualification for propaganda or manipulation was lacking.

- One of the most related to propaganda cases was the **trial over Vojislav Šešelj**. He was a radical Serbian nationalist, who was prosecuted for public speeches during the Yugoslav Wars, including ones distributed through the popular media. Particularly, he repeatedly shared public threats targeting Bosnia and its population, *e.g.* that Bosnia would flow with *"rivers of blood"*, Serbs had to defend themselves from *"Ustasha and pan-Islamist hordes"*, and that Serbs should *"clean the left bank of the river Drina"*.[333] Practically speaking, the followers of Šešelj were then accused of inhumane treatment, unlawful detention of Croats, Muslim and other individuals.[334] The indictment

[328] United States v Ernst van Weizsaecker (Ministries Case), in Xii Trials Of War Criminals Before The Nuremberg Military Tribunals Under Control Council Law No 10 (1951), 350

[329] Kevin Jon Heller, The Nuremberg Military Tribunals and The Origins Of International Criminal Law 457 (2011) 290-291

[330] Richard B Collins, 'Propaganda for War and Transparency' (2010) 87(4) Denver Law Review 819, 822

[331] Statute of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia since 1991 (adopted 25 May 1993) UN Security Council Resolution 827 (1993)

[332] Ibid, Article 7

[333] Jordan Kiper, 'Remembering the Causes of Collective Violence and the Role of Propaganda in the Yugoslav Wars' (2022) Nationalities Papers 1, 2

[334] Prosecutor v Vojislav Seselj, Indictment (2003) IT-03-67, paras 25-26

also pointed at his capacity to aid and abet in preparation and commission of the specific acts of violence, such as extermination of non-Serb civilians,[335] extending the charges to the war propaganda.[336] The Chamber itself firstly lifted all the charges, being unable to find a connection between the speech and actual consequences.[337] Yet, after a detailed analysis of the content of Šešelj's speeches and the context for their sharing, the Appeal Chamber has reached a conclusion that *"such statements [were] undoubtedly capable of creating fear and emboldening perpetrators of crimes against the non-Serbian population"*.[338] Since the ICTY Statute contained no direct prohibition on the speech crimes, Šešelj was prosecuted for instigation of the crimes against humanity. In this respect, his calls were qualified as instigation to persecution, deportation and other inhumane acts. Importantly, he was the only person, who was tried purely for speech crimes, while other cases addressed below viewed propagandistic activities rather in-between the lines.

- The case of **Brdjanin** is another instance where propaganda was analysed by the ICTY as the part of the accused's conduct rather than the general background in Serbian society. In this respect, the Chamber noted that his expressions aimed at *"creating mutual fear and hatred and particularly inciting the Bosnian Serb population against the other ethnicities"*.[339] Further, the ICTY even noted that propaganda has achieved its goal of polarizing society and creating an atmosphere of terror, which served as a precondition for the mass atrocities in the region.[340] Importantly, in this case the resort to media has been taken into account,[341] as a factor influencing the scope of dissemination of inflammatory speech and increasing its impact and trustworthiness.
- The prosecution of **Krajišnik** likewise was emblematic of the consequences for the accused individual participating in *"providing misleading information to the public as well as to the international community and nongovernmental organizations about crimes committed by Bosnian-Serbs"*.[342] Basically, this was almost an exceptional instance, when the ICTY directly dealt with the analogue of disinformation reaching a conclusion that manipulations of information for criminal ends, propaganda and hate speech shall lead to criminal liability.
- Another interesting case was ***Kordić and Čerkez***, where Dario Kordić was accused of crimes against humanity, *i.e.* persecution of individuals via spreading the hate propaganda, which *was "encouraging, instigating and promoting hatred, distrust and strife on political, racial, ethnic or religious grounds, by propaganda, speeches and otherwise"*.[343] The ICTY, however, faced a dilemma since the described crime was absent in the Statute. Respectively, its decision was that Kordić could not be prosecuted for the

[335] Kearney 217
[336] Kearney 216-217
[337] Judgement Summary, Trial Judgement Summary for Vojislav Šešelj (ICTY, 31 March 2016) 12-13
[338] Judgement Summary, Appel Judgement Summary for Vojislav Šešelj (MICT, 11 April 2018), para 26
[339] Prosecutor v Radoslav Brdjanin, TC Judgement (1999) IT-99-36-T, para 80
[340] Ibid, para 83
[341] Ibid, paras 322-323
[342] Prosecutor v Momcilo Krajisnik, TC Judgement (2006) IT-00-39-T, para 7
[343] Prosecutor v Dario Kordic and Mario Cerkez, Amended Indictment (1998) IT-95-14/2, paras 37(c) and 39(c)

hate speech itself in absence of the customary prohibition, yet his behaviour was qualified as instigation to persecutions.[344] And this case directly became the instance, where a legal lacuna in formulation of the Statute was filled via the reference to the modes of liability to cover the illegal actions. Yet, in this regard, it is important to remember that liability for the instigation, aiding and abetting is possible only where the crime was actually committed, which was found by the Chamber in **Popović et al**.[345]

- The representatives of the Federal Security Service testified for the Prosecution in **Slobodan Milošević's trial** about their involvement in the extensive and well-coordinated propaganda campaign.[346] Among the numerous examples of disinformation spread during the conflict, they mentioned the instances of TV broadcasts with corpses, described as Serb civilians being massacred by Croats. Importantly, Milošević and another accused coordinator of the atrocities Martić were trying to resort to disinformation even during the trial. For example, they have both referred to the quote of the Croatian president Franjo Tuđman, which allegedly praised the idea of aggressive war for Croatian independence. In fact, when the prosecutors obtained the original recording of the speech, the words of the Croatian president appeared to be absolutely the opposite[347] – he supported a peaceful declaration of the Croatian independence and *de facto* condemned the aggression.
- Finally, there was a line of cases in which illegal propaganda was reviewed on its surface. For instance, in **Banović case**, the accused tried to shield himself by stating that he was impacted by war propaganda, which should lift his criminal liability. The ICTY, however, explicitly mentioned that being subjected to propaganda cannot serve as a defence since *"the role of the war propaganda, clearly does not affect the gravity of the criminal conduct of the Accused and is more appropriately considered in relation to mitigating factors"*.[348] A similar approach was maintained in **Babić judgement**, where the Chamber noted that the accused was *"strongly influenced and misled by Serbian propaganda, which repeatedly referred to an imminent threat by the Croatian regime"*,[349] but nevertheless charged him for participation in the campaigns of persecution. Lastly, in the famous **Tadić case**, the Tribunal underlined the overall extensive impact of propaganda, media campaigns and deeply hostile messages, which circulated in Serbian society during the periods preceding the escalation.[350] By stating this, the ICTY referred to the necessity of the detailed context-assessment for determining the role of each individual in commission of the international crimes, as well as contribution to their widespread and systematic character. This finding can be supplemented

[344] Kearney 216

[345] Prosecutor v Popovic et al., TC Judgement (2010) IT-05-88-T, paras 1009, 1015

[346] Prosecutor v Slobodan Milosevic (Trial Transcript, 11 November 2002) <https://www.icty.org/x/cases/slobodan_milosevic/trans/en/021111ED.htm> accessed 27 December 2022

[347] Prosecutor v Slobodan Milosevic (Trial Transcript, 23 October 2006) <https://www.icty.org/x/cases/martic/trans/en/061023IT.htm> accessed 27 December 2022

[348] Prosecutor v Predrag Banovic, TC Sentencing Judgment (2003) IT-02-65/1-S, para 48

[349] Prosecutor v Milan Babic, TC Sentencing Judgment (2007) IT-03-72-S, para 24(g)

[350] Prosecutor v Dusko Tadic, TC Opinion and Judgment (1997) IT-94-1, paras 87-89

by **Galić judgement**, where the ICTY suggested that propaganda itself may constitute an act or threat of violence, which enables the prosecution for such activity as for the war crime.[351]

Although there were few instances when individuals were in fact held criminally liable specifically for spreading disinformation, or rather any kind of malicious information, there were lots of practical examples of how disinformation campaigns were conducted, impacting the course of warfare and contributing to the human rights violations. The disinformation topics varied from the depiction of the false flags, under which the military operations have being held (perfidy outlawed by the Geneva Conventions, to which the ICTY Statute refers),[352] lies about the course of the military operations and the quantity of victims, negative stereotyping of Bosnian Muslims and many others,[353] part of which cannot even be verified nowadays. The lack of attention to illegal propaganda and disinformation during the ICTY proceedings can be partly reasoned by the imperfection of the Tribunal's Statute or its focus on the more substantial crimes. Nevertheless, as practice shows, it was disinformation, which was a solid ground from development of a conflict, as well as unsuccessful attempts of its justification by the main perpetrators.

The ICTR. The Statute of the ICTR is quite similar in wording to what has been done in post-Yugoslav context from both procedural and substantive point of view. Namely, the Statute was likewise approved by the Resolutions of the UN Security Council,[354] therefore effectively lifting the immunities and resolving the jurisdictional issues. Akin to that, the crimes included in the Statute were narrowly tailored to the Rwandan context, reflecting the societal events and peculiarities of the atrocities' history. In this regard, the Statute mainly focused on the crime of genocide, including the incitement, attempts to and complicity in its commission, also containing inexhaustive list of the crimes against humanity, violation of the Article 3 of the Geneva Conventions and Additional Protocol II thereto.[355] Article 6 of the ICTR Statute reflected Article 7 of the ICTY Statute, postulating that it is applicable to individuals, who "*planned, instigated, ordered, committed or otherwise aided and abetted in the planning, preparation or execution of a crime*" .[356] The factual difference, however, was that the ICTR mostly dealt with the high-ranking individuals. The rest of the processes over the ordinary perpetrators were held on the domestic level via the 'gacaca' courts,[357] which were composed of the publicly selected non-professional

[351] Predrag Dojcinovic, 'Propaganda in the Jurisprudence of the International Criminal Tribunal for the Former Yugoslavia' (2011) Routledge 1, 12

[352] 'Operation Labrador' (Military Wiki) <https://military-history.fandom.com/wiki/Operation_Labrador> accessed 27 December 2022

[353] 'Propaganda during the Yugoslav Wars' (Military Wiki) <https://military-history.fandom.com/wiki/Propaganda_during_the_Yugoslav_Wars#Serbian_propaganda_cases> accessed 27 December 2022

[354] Statute of the International Criminal Tribunal for the Prosecution of Persons Responsible for Genocide and Other Serious Violations of International Humanitarian Law Committed in the Territory of Rwanda and Rwandan Citizens Responsible for Genocide and Other Such Violations Committed in the Territory of Neighbouring States, between 1 January 1994 and 31 December 1994 (adopted 8 November 1994) UN Security Council resolution 955 (1994)

[355] Ibid, Articles 2-4

[356] Ibid, Article 6

[357] 'Rwanda: Justice After Genocide — 20 Years On' (HRW, 28 March 2014) <https://www.hrw.org/news/2014/03/28/rwanda-justice-after-genocide-20-years> accessed 27 December 2022

judges (and were actually criticized for the quality of the decisions). According to the estimates of the Institute for Democracy and Electoral Assistance, the ordinary courts in Rwanda have dealt with 77,269 cases, the non-professional ones – with almost 800,000 (including 433,557 against the participants of massacres and 308,738 of looters).[358] The system of domestic justice in Rwanda, though, has been subjected to a harsh criticism. In contrast to the IMT trial – a tribunal of victors, the Rwandan courts amounted to the trial of victims over the perpetrators. Respectively, lots of bias and prejudice distorted the notion of justice,[359] making the judicial process and all the verdicts an additional ground for the internal hatred. In contrast, the ICTR decisions were balanced, produced by the independently appointed judges, part of which were unrelated to the Rwandan context.

- The first person ever prosecuted for a direct and public incitement to genocide within the ICTR jurisprudence was **Jean-Paul Akayesu**, the bourgmestre of Taba commune and a popular figure in a local community *"treated with great respect and deference"*.[360] In contrast to Streicher, convicted with the incitement to crimes against humanity, Akayesu's expressions were viewed as deliberate and conscious call to kill the whole Tutsi group. Particularly, the speech instigating to *"unite and eliminate the sole enemy"* [361] was construed in a manner to generalize the Tutsi population, depicting it as a threat and a target for extermination, while repeated calls to rape Tutsi women were a part of a coordinated propaganda campaign.[362] Besides that, the ICTR made an important finding that the incitement shall not necessarily be absolutely successful, but rather have a chance of transforming into the actual violence.[363] In this respect, significant attention has been paid to mass media, as a mean for dissemination of illegal speech, which only aggravates the potential consequences.[364] Akin to that, the ICTR drew a line between aiding and abetting in commission of genocide and delivering a speech as a separate crime of incitement.[365] Incitement, in particular, shall *"assume a direct form and specifically provoke another to engage in a criminal act"*, being aimed at causing a specific offence, not merely being related to it.[366] Lastly, the socio-political context was considered to be an important element for interpretation of the speech, including the cultural and linguistic peculiarities of it.
- One of the most emblematic cases regarding the freedom of expression as a tool for commission of international crimes apparently is a trial over ***Nahimana, Barayagwiza and Ngeze***, more known as ***the Media* case**. It is fairly considered by Kearney to be *"the most comprehensive analysis*

[358] Luc Huyse et al., Traditional Justice and Reconciliation after Violent Conflict (IDEA, 2008) 42

[359] Gerald Gahima, 'Alternatives to prosecution: The case of Rwanda' in Edel Hughes, William A Schabas and Ramesh Thakur (eds) Atrocities and International Accountability: Beyond Transitional Justice (CUP, 2007) 167-168

[360] Prosecutor v Akayesu, TC Judgement (1998) ICTR 96-4-T, para 54

[361] Ibid, para 361

[362] Kearney 221

[363] Thomas E Davies, 'How the Rome Statute Weakens the International Prohibition on Incitement to Genocide' (2009) 2 Harvard Human Rights Journal 246, 253

[364] Prosecutor v Akayesu, TC Judgement (1998) ICTR 96-4-T, para 559

[365] Johan D van der Vyver, 'Prosecution and Punishment of the Crime of Genocide' (1999) 23(2) Fordham International Law Journal 286, 314-315

[366] Prosecutor v Akayesu, TC Judgement (1998) ICTR 96-4-T, para 557

undertaken by an international criminal tribunal of 'incitement to crimes of an international dimension'".[367] These three individuals were convicted for a direct and public incitement to genocide for their participation in the RTLM, production of a weekly newspaper 'Kangura' and contribution to the activity of the political party Coalition for the Defence of the Republic. Nahimana, specifically, was found liable for programming the RTLM, thus exercising editorial control over it and having a direct knowledge of the incitements being spread. In this judgement, the ICTR made important findings about the all-encompassing nature of the intent to commit genocide, which cover the intent to persecute, exterminate and destroy a group in whole.[368] The Appeals Chamber clarified that incitement becomes a completed crime from the moment it is delivered to the general public, not requiring the performance of the incited actions.[369] Namely, its potential to cause genocide is decisive for qualification as an international crime.[370] The forms of the illegal incitement are also unlimited, varying from speech and writing to images and audiovisual works.[371] Importantly, the ICTR also tried to distinguish the incitement from a legitimate use of media,[372] basically referring to the criteria resembling the Rabat Plan of Action.

- Within the **Bikindi trial**, the ICTR reached a conclusion that such statements as *"rise up and look everywhere possible"* not to miss any *"snakes"* amount to a hateful message by its wording designed to destroy Tutsi as an ethnic group.[373] Yet, the ICTR, although establishing the malicious and dangerous nature of the songs, was unable to prove the intent of Bikindi to exterminate Tutsi, which is a necessary element to establish an incitement to genocide.[374] In this respect, Bikindi himself was a quite famous composer and signer, who contributed with his 'artistic projects' to escalation of genocidal practices,[375] sharing his songs within the existing atmosphere of political and ethnic tension in Rwanda.[376] This case serves another instance proving that artistic speech does not enjoy absolute immunity under freedom of expression, being able under some circumstances to bring even more harm. Namely, Bikindi's song *Nanga Abahutu* ("I Hate These Hutu") was easily learnt by thousands of Hutu, reaching much broader audiences than an ordinary speech ever could. A comparison in this regard can be drawn with the poetic expressions distributed by the Russian social media users, aimed at humiliation of Ukrainians and calls for violence against them. Since even one expression suffices for criminal liability

[367] Kearney 222

[368] The Media case, para 1035

[369] The Media case, para 723

[370] Prosecutor v Nahimana, Barayagwiza and Ngeze, TC Judgement (2003) ICTR-99-52-T, paras 1015, 1029

[371] The Media case, para 148

[372] 'HRW Report on Genocide' (HRW, February 2004) <https://www.hrw.org/reports/2004/ij/ictr/3.htm> accessed 27 December 2022

[373] 'Facebook blacklists Myanmar hardline Buddhist group' (Frontier Myanmar, 7 June 2018) <https://frontiermyanmar.net/en/facebook-blacklists-myanmar-hardline-buddhist-group> accessed 27 December 2022

[374] Prosecutor v Bikindi, TC Judgement (2008) ICTR-01-72-4-T, paras 247-253

[375] United States Holocaust Memorial Museum, Washington, DC, 'Incitement to Genocide in International Law' (Holocoust Encyclopedia, 21 October 2021) <https://encyclopedia.ushmm.org/content/en/article/incitement-to-genocide-in-international-law> accessed 27 December 2022

[376] N R Okany and J Hoffmann, Taking the Prevention of Genocide Seriously: Media Incitement to Genocide Viewed in the Light of the Responsibility to Protect in J Hoffmann and A Nollkaemper (eds), Responsibility to Protect: From Principle to Practice (Amsterdam University Press, 2012) 327

to occur, while artistic speech cannot be used as a defence, there is a space for considering the potential liability of the particularly popular publishers.

- Another prosecuted perpetrator **Kambanda**, the then Rwandan Prime Minister, delivered a speech reading as *"you refuse to give your blood to your country and the dogs drink it for nothing"*,[377] which despite the vague wording was considered to be a direct incitement to commit genocide. In a plea of guilty to the prosecutor, Kambanda mentioned the large-scale plans regarding the use of the RTLM to mobilize and incite Hutu to commit massacres against Tutsi population.[378] Particularly, he incited the RTLM to proceed with propaganda viewing it as *"an indispensable weapon in the fight against the enemy"*.[379] As a result, the coordinating activities were considered to be sufficient contribution to the incitement to genocide, though few direct speeches of Kambanda were actually assessed by the ICTR compared to other perpetrators.
- **Eliézer Niyitegeka,** a journalist and a news presenter on the Radio Rwanda and the Minister of Information of the Interim Government was also charged with a direct and public incitement to genocide. Scholars, in this respect, consider his speeches to be rather vague and implicit.[380] Nevertheless, the Chamber found that Niyitegeka by *"urging the attackers to work, thanking, encouraging and commending them for the "work" they had done"* was using the word 'work' as a reference to mass massacres of Tutsi.[381] Even though the Tribunal has found that some of the incitements apparently have not reached their goal, it clarified that unsuccessful acts of incitement are still punishable.[382] Moreover, in contrast to many other accused individuals, the ICTR has established an intent of Niyitegeka to incite other to commit atrocities*, i.e. "inciting attackers to cause the death and serious bodily and mental harm of Tutsi refugees"*.[383] Interestingly, the formulation of his conviction did not sound as incitement to genocide, though Niyitegeka's conduct was reviewed as a separate crime, not as a mode of liability within the scope of other prohibited activity.

- The **Ruggiu proceedings** before the ICTR were important in view of the qualification of hate speech as an integral part of the persecution campaign. Namely, the Tribunal established that hateful and hostile *"radio broadcasts all aimed at singling out and attacking the Tutsi ethnic group and Belgians on discriminatory grounds … [having] as its aim the death and removal of those persons from the society in which they live alongside the perpetrators,*

[377] Prosecutor v Jean Kambanda, AC Judgment and Sentence (1998) ICTR 97-23-S, para 39(x); Tommi Aromäki, 'The International Criminal Court's Jurisdiction Over Incitement to Genocide in the Internet Era - Some special situations' (University of Helsinki, 2021) 14

[378] Kearney 222

[379] Prosecutor v Jean Kambanda, AC Judgment and Sentence (1998) ICTR 97-23-S, para 39(vii)

[380] Tommi Aromäki, 'The International Criminal Court's Jurisdiction Over Incitement to Genocide in the Internet Era - Some special situations' (University of Helsinki, 2021) 14

[381] Prosecutor v Eliézer Niyitegeka, TC Judgement and Sentence (2003) ICTR-96-14-T, paras 436-437

[382] Ibid, para 431

[383] Ibid, paras 436-437

or eventually even from humanity itself".[384] In this regard, inability to prove the actual link with the occurred violence was found to be irrelevant,[385] which is interesting given the fact that hate speech itself, in contrast to the incitement to genocide, is not viewed as a separate crime. Accordingly, it usually requires the actual negative repercussions following the expression and a direct causation with it. However, in *Ruggiu*, the ICTR stepped aside of such interpretation, stressing that "unsuccessful acts of incitement can be punished".[386] The Tribunal once again stressed the contribution of the RTLM to the development of hostilities.

Meanwhile, the trials did not end up with the ICTR processes purely, having reached some foreign domestic jurisdictions, and even created another court in the Hague. For example, a famous ***Mugesera case*** concerned the prosecution of a Rwandan Hutu extremist, who called upon massacres of Tutsis, in Canada, where he claimed refuge and obtained a permanent resident status.[387] Mugesera could not be tried by the ICTR since his speech occurred long before the January 1, 1994[388] – a date from which the Tribunal had jurisdiction. Nevertheless, both the first instance court and the appeal court have found his expressions to constitute an incitement to genocide. In particular, the Canadian courts have paid specific attention to the phrase *"we will send you by the Nyabarongo"*, which was qualified as a clear suggestion that the corpses of murdered Tutsis would be sent back to Ethiopia via the Nyabarongo river.[389] One statement, in this respect, was sufficient to establish a direct and public incitement, given the context in which it was delivered as well as the receptiveness of the audience.

Finally, the process over **Kabuga** promises to echo many findings of the ICTR trials with regard to incitements and complicity. Kabuga, the founder and the main sponsor of the RTLM, has been hiding for over 23 years, being apprehended in France in late 2020.[390] To address the remaining cases of the situation in Rwanda, the UN Security Council has specifically established the International Residual Mechanism for Criminal Tribunals located in the Hague.[391] The indictment includes the participation in genocide, incitements to it and complicity in commission of genocide and crimes against humanity.[392] Most probably, the tribunal will view his actions as a direct

[384] Prosecutor v Georges Ruggiu, Trial Chamber (2000) ICTR-97-32-I

[385] Mohamed Elewa Badar and Polona Florijančič, 'The Prosecutor v. Vojislav Šešelj: A Symptom of the Fragmented International Criminalisation of Hate and Fear Propaganda' (Brill, 28 May 2020) <https://brill.com/view/journals/icla/20/3/article-p405_405.xml?language=en> accessed 27 December 2022

[386] Prosecutor v Georges Ruggiu, Trial Chamber (2000) ICTR-97-32-I, para 16

[387] Justus Reid Weiner et al., Referral of Iranian President Ahmadinejad on the Charge of Incitement to Commit Genocide (The Jerusalem Center for Public Affairs, 2006) 30

[388] Ibid

[389] UN Human Rights Council, Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar, UN Doc A/HRC/39/CRP.2 (17 September, 2018), para 1319

[390] Marlise Simons, 'Rwanda Genocide Tribunal's Most Wanted Man Finally Faces Trial' (The New York Times, 29 September 2022) <https://www.nytimes.com/2022/09/29/world/europe/rwanda-genocide-trial-felicien-kabuga.html> accessed 27 December 2022

[391] 'Start of trial in Prosecutor v. Félicien Kabuga at the Mechanism's Hague branch on Thursday, 29 September 2022' (UN IRMCT, 26 August 2022) <https://www.irmct.org/en/news/start-trial-prosecutor-v-felicien-kabuga-mechanisms-hague-branch-thursday-29-september-2022> accessed 27 December 2022

[392] 'KABUGA, Félicien (MICT-13-38)' (UN IRMCT) <https://www.irmct.org/en/cases/mict-13-38> accessed 27 December 2022

participation in genocide via its financing rather than any kind of speech crimes. Especially, since for an incitement a direct and public call is needed, which according to the available evidence Kabuga never did. However, this case is quite important for Ukrainian future criminal liability mechanism against Russian propagandists since many individuals, who have not publicly incited or instigated commission of international crimes, still participated in their commission via financing of media outlets.

Other criminal tribunals. If one thinks that the history of the special criminal tribunals ends up with the ICTY and the ICTR, this opinion is rather wrongful. The special courts were also created to review the cases emerging following the atrocities in Sierra Leone, Lebanon, Cambodia, and East Timor. Findings of few of them, however, are relevant for prosecuting individuals for dissemination of disinformation, malicious propaganda and illegal incitements. For example, in East Timor no cases were launched with regard to the illegal incitements or propaganda, but a significant role was ascribed to Indonesian propaganda in war escalation.[393] And this might serve as a counterexample, when non-prosecution of propagandistic activities *de facto* leaves the victims without a proper remedy since the conflict is redressed based on its consequences with an absolute ignorance towards the causes.

However, there is an interesting case in jurisprudence of the Special Tribunal for Lebanon, which recognized that the legal persons are not exempt from the prosecution under the ICL.[394] Particularly, the Tribunal established that it has jurisdiction to consider a **case against New TV S.A.L.**, which employed the issue of the corporate liability. In this respect, it is important to note that media can be held liable for dissemination of propaganda and illegal incitements, as well as false information contained therein. This precedent, thus, is of particular relevance for qualification of the actions of the Russian media outlets, which constantly delivered disinformation not only on the level of particular individuals, but also as a part of their corporate policies. The same approach shall be developed towards the State-controlled online intermediary platforms.

The ICC. It is a treaty-based permanent international criminal court, which can decide the cases referred to it by the Parties to the Rome Statute or based on the initiative of the Prosecutor's office.[395] The second instance, however, is still limited to ratification of the Statute since Article 34 of the VCLT, which reflects international custom,[396] provides that *"a treaty does not create either obligations or rights for a third State without its consent"*.[397] Hence, the territorial limits are as well established

[393] Antonio Barbedo de Magalhaes, 'East Timor: A People Shattered by Lies and Silence' (CourseHero, 5 March 2021) <https://www.coursehero.com/file/91499791/East-Timor-a-People-Shattered-by-Lies-and-Silencepdf/> accessed 27 December 2022

[394] Prosecutor v New TV S.A.L. and Karma Mohamed Tahsin al Khayat, Decision on Interlocutory Appeal Concerning Personal Jurisdiction in Contempt Proceedings (2014) STL- 14-05/PT/AP/ARI26.1, paras 39-40

[395] Rome Statute, Articles 14, 15

[396] Dörr O., Schmalenbach K., Vienna Convention on the Law of Treaties: A Commentary (Springer, 2012) 607

[397] Vienna Convention on the Law of Treaties, 1155 UNTS 332, Article 34

for exercising the ICC jurisdiction. Yet, the only exceptional case was reviewed in *Myanmar-Bangladesh* situation, where the ICC found that jurisdiction is present if at least part of the crime was committed in the territory of the State Party.[398] Notably, this finding might be relevant for cases when crimes are committed from abroad, *e.g.* sharing of propaganda or disinformation. Akin to that, additional restrictions are provided for jurisdiction over the crime of aggression,[399] implying a need of a State Party to the Rome Statute to accept the ICC jurisdiction with regard to the crime of aggression.

The Rome Statute covers four categories of crimes, *i.e.* genocide, war crimes, crimes against humanity and the crime of aggression.[400] Among those four categories the speech crime is absent, except for a very small potential of being covered by *"other inhumane acts of a similar character intentionally causing great suffering, or serious injury to body or to mental or physical health"* as a part of the crimes against humanity.[401] Although many scholars conducted the large-scale advocacy campaigns in favour of inclusion of the propaganda for war into the Rome Statute,[402] the drafters of the Rome Statute had fear of the objections from the States' side and, thus, abstained. The same fears were expressed towards the inclusion of a direct and public incitement to aggression as a separate crime, which neither became a part of the 1996 Draft Code of Offences, nor a part of the Rome Statute.[403] Hence, even the general prohibition of incitements in the IHRL and the PIL[404] was insufficient for the drafters to include this in the list of prohibited actions. Accordingly, the speech crimes are viewed by the ICC as the modes of liability rather than inchoate crimes, as was repeatedly done by its numerous predecessors regarding the incitement to genocide.

The liability occurs not only for the direct perpetrators, but also for those facilitating the crime commission: ordering or organising it, in any way assisting in its commission, or contributing to it. The Rome Statute incorporates the various modes of liability in Article 25.[405] In this respect, sharing of propaganda or disinformation can be qualified as prompting an international crime under Article 25(3) of the Rome Statute in the form of inducing, as 'other assistance' or 'other contribution' to a certain crime. Nevertheless, the most classical example of the liability for speech would be incitement to genocide, covered by Article 25(3)(e).[406] Notably, a kind of 'reservation' regarding an incitement to genocide was made, *i.e.* a person can

[398] Tommi Aromäki, 'The International Criminal Court's Jurisdiction Over Incitement to Genocide in the Internet Era - Some special situations' (University of Helsinki, 2021) 23-29

[399] Rome Statute, Arts 15 bis, 15 ter

[400] Rome Statute, Art 5

[401] Rome Statute, Art 7(1)(k)

[402] Propaganda and Freedom of the Media: Non-paper of the OSCE Office of the Representative on Freedom of the Media (Vienna, 26 November 2015) 32 <https://www.osce.org/fom/203926> accessed 27 December 2022

[403] Kearney 194-210

[404] United States Holocaust Memorial Museum, Washington, DC, 'Incitement to Genocide in International Law' (Holocoust Encyclopedia, 21 October 2021) <https://encyclopedia.ushmm.org/content/en/article/incitement-to-genocide-in-international-law> accessed 27 December 2022

[405] Rome Statute, Article 25(3)

[406] Rome Statute, Article 25(3)(e)

be charged under Article 25(3)(e) even if genocide was not in fact committed.[407] As regard other crimes, their commission is required for liability to occur. For example, disinformation can be addressed as solicitation, inducement, aiding and abetting, or contribution to a group crime,[408] which is never a separate crime.

At the same time, Kearney stressed that the ICC followed the approach of other criminal tribunals in prosecuting primarily those, who *"ordered, organized, planned, and incited genocide, crimes against humanity, war crimes, and aggression"* .[409] Yet, it also made a step forward, since the Rome Statute finally addresses indirect commission of international crimes through the so-called proxy-individuals (by virtue of making others to commit a crime).[410] In this respect, dissemination of disinformation might well be considered as *"instrumentalising another person to commit a crime, be it the use of force or the exploitation of an error or any other handicap on the tool's side"* .[411] Accordingly, a window of interpretation of contribution to the crimes under different modes of liability is broader than in the practice of the other international tribunals.

Another issue of the heated debate among the scholars and the policy community implies the proof of the leadership of an individual to trigger the applicability of Article 8 *bis*.[412] In this regard, the 2010 Kampala Conference clarified that any element of the crime of aggression relates only to those, shaping the State policies, controlling and directing the performance of the criminal activities.[413] Even if it is broad enough to cover individuals, who are not affiliated with the State, such as industrialists,[414] the circle of potential perpetrators is a limited one. This, in turn, makes the prosecution even more sophisticated, where the issue concerns the propaganda for war or aggression. In such a case, incitement or provocation shall be directed against individuals, who fall within this legal criterion. Not surprisingly, the ICC have never dealt with such cases in its practice up to this day.

Nevertheless, the jurisprudence of the ICC is quite diversified since it is designed to decide any dispute referred to it by the Parties. It implies the mixture of contexts and situations to be addressed by the Court. Nevertheless, the ICC still had

[407] Robin Geiß and Henning Lahmann, 'Protecting the global information space in times of armed conflict' (International Review of the Red Cross, January 2022) <https://international-review.icrc.org/articles/protecting-the-global-information-space-in-times-of-armed-conflict-915> accessed 27 December 2022

[408] Eian Katz, 'Liar's war: Protecting civilians from disinformation during armed conflict' (International Review of the Red Cross, December 2021) <https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914> accessed 27 December 2022

[409] Kearney 235

[410] Kai Ambos, 'Article 25: Individual Criminal Responsibility' in Otto Triffterer (ed.), Commentary on the Rome Statute of the International Criminal Court (Baden-Baden: Nomos Verlagsgesellschaft, 1999, 2d ed), 743–770, 749

[411] Albin Eser, 'Individual Criminal Responsibility', in Antonio Cassese, Paolo Gaeta, and John RWD Jones (eds), The Rome Statute of the International Criminal Court: A Commentary (OUP, 2002) 767–822, 794

[412] Nikola Hajdin, 'Complicity in a War of Aggression: Private Individuals' Criminal Responsibility' (JustSecurity, 1 April 2022) <https://www.justsecurity.org/80937/complicity-in-a-war-of-aggression-private-individuals-criminal-responsibility/> accessed 27 December 2022

[413] Review Conference of the Rome Statue (Kampala, Uganda 31 May – 11 June 2010) <https://asp.icc-cpi.int/reviewconference> accessed 27 December 2022

[414] Report of the Special Working Group on the Crime of Aggression (2009) ICC-ASP/7/20/Add.1, para 25

an opportunity to deal with the information and propaganda crimes in a couple of its judgements:

- One of a few cases dealing with the media issues was **Ruto and Sang case.** Joshua Arap Sang was a radio DJ, accused of solicitation or inducement and aiding and abetting in commission of the crimes against humanity.[415] His actions, in particular included broadcasting of false news regarding alleged murders of Kalenjin people to inflame the atmosphere in the days preceding the elections.[416] However, the prosecutors failed to present any exemplary of his speeches, thus depriving the Court of possibility to analyse the instigating or inflammatory nature of his speech.[417] Accordingly, the charges regarding the radio broadcasts were dropped due to the lack of evidence and lack of reasonable grounds to believe that Sang's actions have in fact caused any illegal activity.

  William Samoei Ruto was charged with co-perpetration of the crimes against humanity, which requires the proof of an essential contribution to the crime. Namely, The ICC stressed on the requirement for a causal nexus between the speech and the criminal acts by saying that *"it still has to be established that this message was actually heeded by the physical perpetrators or that his speeches had a direct effect on their behaviour"*.[418] Nevertheless, the case was terminated with a possibility to bring new charges against both defendants.[419]

  Another emblematic case, which apart from that was highly disputable in terms of the ICC findings with some dissenting opinions emerging was a trial over **Callixte Mbarushimana**. Being the Executive Secretary of the Forces Démocratiques pour la Libération du Rwanda, based in Paris, he *"issued several press releases on behalf of the organisation in the aftermath of operations, systematically denying any responsibility of the group"*, while also being engaged *"in international peace talks and negotiations, shrewdly portraying the FDLR as an actor seeking peace and stability in the Kivu area"*.[420] His charges were brought up under Article 25(3)(d), implying the 'other contribution' to the crimes committed. The majority found lack of sufficient evidence to believe that his denial or concealing of atrocity crimes already committed has contributed to future crimes.[421] Namely, the interpretation of aiding and abetting concept was proposed to be done

[415] Prosecutor v William Ruto and Joshua Sang, Decision on Defence Applications for Judgments of Acquittal (2016) ICC-01/09-01/11, para 139

[416] Jenny Domino, 'Crime as Cognitive Constraint: Facebook's Role in Myanmar's Incitement Landscape and the Promise of International Tort Incitement Landscape and the Promise of International Tort Liability Liability' (2020) 52(1) Case Western Reserve University School of Law 143, 180

[417] Mohamed Elewa Badar and Polona Florijančič, 'The Prosecutor v. Vojislav Šešelj: A Symptom of the Fragmented International Criminalisation of Hate and Fear Propaganda' (Brill, 28 May 2020) <https://brill.com/view/journals/icla/20/3/article-p405_405.xml?language=en> accessed 27 December 2022

[418] Prosecutor v William Ruto and Joshua Sang, Decision on Defence Applications for Judgments of Acquittal (2016) ICC-01/09-01/11, para 135

[419] 'Ruto and Sang Case' (ICC, 5 April 2016) <https://www.icc-cpi.int/kenya/rutosang> accessed 27 December 2022

[420] Prosecutor v Callixte Mbarushimana, Decision on the Confirmation of Charges (2011) ICC-01/04-01/10-465-Red, para 8

[421] Ibid, paras 295-340

in a restrictive manner.[422] Nevertheless, there was a disagreement with such a finding in a separate opinion.[423] Similarly, scholars stressed that the ICC main concern implied a distance between Mbarushimana and the location of the crimes since he contributed to their commission only by press releases and broadcasts.[424] Yet, the distance and absence of physical participation in a crime cannot exclude or diminish the impact of the contribution to the commission of such crime.

As can be seen from the outcome of two key cases, dealing with the incitement issue is a relatively sophisticated task for the ICC since its non-inclusion with the list of crimes complicates the work of the prosecution. Or, being fair, makes it hardly possible to prove the causation between an expression and the actual crime committed, as well as invalidating the charges were the crime itself was not completed or was unfinished, or exceeded the request of the speaker. Likewise, the problem might relate to the lack of cases dealing specifically with incitements and speech crimes, as happened in the mentioned *Media case* or the trial over Streicher.

The special criminal tribunal in Ukraine. The European Parliament has expressed the full support of the ICC jurisdiction and investigation into the crimes committed by Russia in Ukraine, prosecuting the perpetrators, including those, who *"assisted such crimes by way of propaganda"*.[425] However, since the ICC jurisdiction and the legal framework in general is rather unsuitable for prosecution of speech crimes in the context of the Russian aggression in Ukraine, many scholars and public officials are now discussing the idea of establishing a special criminal tribunal. In particular, it will enable to resolve the issues of immunities, as well as to cover the crimes, which are specific for Ukrainian context and are not addressed in the Rome Statute due to their novelty for the ICL. Moreover, the jurisdiction over the crime of aggression can be established only based on the consent of the State,[426] which likewise makes bringing the responsible individuals to justice hardly possible.

In this respect, the procedure followed by the international community in cases of Rwanda and Yugoslavia is hardly applicable. The UN Security Council is now frozen with its hands tied by the Russian right to veto any decision, including ones on creation of the special tribunal. However, the representative of Lichtenstein to the UN proposed to establish a tribunal under the Resolution of the UN General Assembly,[427] which is empowered to take a decision in cases of veto by the permanent

[422] Talita de Souza Dias, 'Propaganda and Accountability for International Crimes in the Age of Social Media: Revisiting Accomplice Liability in International Criminal Law' (OpinioJuris, 4 April 2018) <http://opiniojuris.org/2018/04/04/propaganda-and-accountability-for-international-crimes-in-the-age-of-social-media-revisiting-accomplice-liability-in-international-criminal-law/> accessed 27 December 2022

[423] Prosecutor v Callixte Mbarushimana, AC Decision (2012) ICC-01/04-01/10 OA 4, Separate Opinion of Judge Silvia Fernandez de Gurmendi

[424] Vernon Van Dyke, 'International Criminal Liability for Spreading Disinformation in the Context of Mass Atrocity' (2022) 20(1) Journal of International Criminal Justice 223, 240-242

[425] Marika Lerch and Sara Mateos del, 'Russia's war on Ukraine in international law and human rights bodies: Bringing institutions back in' (2018) European Parliament 8

[426] Rome Statute, Articles 15 bis, 16 ter

[427] 'В ООН обсудили правовые аспекты наказания за агрессию против Украины' (United Nations, 27 October 2022) <https://news.un.org/ru/story/2022/10/1433987> accessed 27 December 2022

members of the UN Security Council when the issues concern international peace and security matters. The same idea was supported on the EU level, where the European Commission proposes a launch of the hybrid tribunal with the authorisation of the UN.[428] It is still early to talk about the launch of the tribunal itself, a need to consider the potential wording for Articles in its Statute, as well as the evidentiary basis for prosecution still exists.

Accordingly, the formulation of the relevant Articles in the Statute of the criminal tribunal over the Russian propagandists can be designed in the following manner:

Article XX. Crimes against Humanity
For the purpose of this Statute, "crime against humanity" means any of the following acts when committed as part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack:
(a) direct and public incitement to commit crimes against humanity;
For the purpose of paragraph 1:
(a) "direct and public incitement to commit crimes against humanity" means expression publicly distributed in any form and by any means, including printed press, audiovisual media and online intermediary platforms, which aims at provoking a person or group of persons to commit crimes against humanity listed in this Article.

Article XX. War Crimes
For the purpose of this Statute, «war crimes» means:
(a) direct and public incitement to commit war crimes;
For the purpose of paragraph 1:
(a) "direct and public incitement to commit war crimes" means expression publicly distributed in any form and by any means, including printed press, audiovisual media and online intermediary platforms, which aims at provoking a person or group of persons to commit war crimes listed in this Article.

Article XX. Genocide
(a) direct and public incitement to commit genocide;

Article XX. Crime of Aggression
Act of aggression means the use of armed force by a State against the sovereignty, territorial integrity or political independence of another State or in any other manner inconsistent with the Charter of the United Nations. Any of the following acts, regardless of a declaration of war, shall qualify as an act of aggression:
(a) direct and public incitement to aggression;*
* For the purposes of Article on Grounds for Excluding Criminal Responsibility, the expressions pronounced on behalf of the population in the defending State, designed as a speech in self-defence, shall not be punishable under Article on Crime of Aggression.

[428] Alexandra Brzozowski, 'EU seeks special court to try Russia's war crimes in Ukraine' (Euractiv, 30 November 2022) <https://www.euractiv.com/section/europe-s-east/news/eu-seeks-special-court-to-try-russias-war-crimes-in-ukraine/> accessed 27 December 2022

As regards the evidence for the proceedings before the international criminal tribunal, numerous organisations are now working on collection of data pertaining to the calls to violence, incitements to genocide and propaganda for war, including through dissemination of disinformation and manipulative narratives. For instance, the mentioned report of the US New Lines Institute for Strategy and Policy provides a broad analysis of the inciting expressions, *e.g. "soldiers have explicitly threatened to rape "every Nazi whore"*, *"we've got to kill these fuckers"*, calls for *"hunting Nazis"* and many more similar statements, which in the context of armed aggression might even reach a threshold for incitement to genocide.[429] The same research also extensively elaborates on the evidentiary basis for proving the intent to destroy the Ukrainian population in whole or in part.[430]

Even the Russian media Meduza clarifies that there were lots of instances of direct and public incitement if not to genocide, then at least to the commission of crimes against humanity. To exemplify, the head of RT Russia, Anton Krasovskyi, proposed to drown children in a river for any statement regarding the Russian occupation, further adding that children can also be burnt.[431] Other examples of the hateful expressions of the high-ranking State officials can be found in various other media broadcasts and news articles.[432] Accordingly, the future tribunal will have a significant amount of cases to deal with within the dimension of speech crimes, which is another reason for not limiting it to the highest State officials purely.

Another type of violation includes propaganda, containing the breach of the III Geneva Convention in relation to the depiction of the prisoners of war. For instance, there is a widely discussed case involving the British citizen Aiden Aslin, who served in the Ukrainian armed forces, was captured by Russians and interviewed by Graham Phillips, another British citizen living in the occupied territories of Ukraine and delivering pro-Russian propaganda.[433] The latter one depicted Aslin handcuffed, physically injured and under duress to crated the propaganda materials.[434] The discourse in this case also concerns the level of involvement of Phillips in the commission of international crimes and instigation to such crimes. Nevertheless, the approach from the post-World War II trials can be upheld with regard to the

[429] Yonah Diamond et al., 'An Independent Legal Analysis of the Russian Federation's Breaches of the Genocide Convention in Ukraine and the Duty to Prevent' (2022) US New Lines Institute for Strategy and Policy and Raoul Wallenberg Centre for Human Rights 19

[430] Ibid, 37

[431] Петр Сапожников, 'Трибунал в Гааге судит главного пропагандиста Руанды, который сыграл ключевую роль в геноциде. Его российских «коллег» ждет то же самое?' (Meduza, 7 November 2022) <https://meduza.io/cards/tribunal-v-gaage-sudit-glavnogo-propagandista-ruandy-kotoryy-sygral-klyuchevuyu-rol-v-genotside-ego-rossiyskih-kolleg-zhdet-to-zhe-samoe> accessed 27 December 2022

[432] Тимофей Сергейцев, 'Что Россия должна сделать с Украиной' (РИА Новости, 5 April 2022) <https://ria.ru/20220403/ukraina-1781469605.html> accessed 27 December 2022; Илья Бер, 'Говорил ли Александр Дугин, что „украинцев надо убивать, убивать и убивать"?' (Delfi, 23 August 2022) <https://rus.delfi.ee/statja/120055354/govoril-li-aleksandr-dugin-chto-ukraincev-nado-ubivat-ubivat-i-ubivat> accessed 27 December 2022; Александр Максюк, 'Армен Гаспарян – адвокат геноцида. Почему в России открыто призывают убивать украинцев' (Сегодня, 13 April 2022) <https://www.segodnya.ua/strana/podrobnosti/advokat-genocida-pochemu-armyanin-gasparyan-prizyvaet-russkih-ubivat-ukraincev-1614729.html> accessed 27 December 2022

[433] Jim Waterson, 'Who is Graham Phillips, the YouTuber accused of 'war crimes'?' (The Guardian, 20 April 2022) <https://www.theguardian.com/world/2022/apr/20/who-is-graham-phillips-the-youtuber-accused-of-war-crimes> accessed 27 December 2022

[434] Kevin Rawlinson and Heather Stewart, 'Family of captive Briton tell of distress at seeing him on Russian TV' (The Guardian, 19 April 2022) <https://www.theguardian.com/world/2022/apr/19/family-captive-briton-aiden-aslin-distress-russian-tv> accessed 27 December 2022

prosecution of foreigners. Akin to Phillips, though, millions of Russians themselves publish similar propagandistic content designed to promulgate hatred and hostility, dehumanise Ukrainians and incite to illegal actions against them.

Conditions for bringing individuals distributing disinformation in the context of the international armed conflict launched by Russia against Ukraine:

I. Determining the appropriate forum for reviewing the charges against individuals accused in the commission of crimes which include dissemination of disinformation:

- The ICC has jurisdiction over the crimes committed during the armed conflict in Ukraine based on the agreement regarding the ad hoc jurisdiction of the Court. At the same time, the issue of immunities of the highest State officials remains relevant and shall be addressed by the Court (in view of the interpretation of Article 98 of the Rome Statute);
- The special criminal tribunal can be created based on the Resolution of the UN General Assembly, having the issues of disinformation and illegal propaganda explicitly addressed in its Statute. Otherwise, the current modes of liability would exclude the potential for convicting individuals whose activities have not directly led to prohibited consequences or where the link between an expression and consequences is not established.

II. The practice of the international criminal tribunals provides relatively extensive guidance on how information crimes shall be treated from the perspective of the modes of liability, causation links and interrelation with the actual consequences, as well as qualification of the particular expressions in the light of the general social and political environment where such expressions are pronounced.

III. Liability of the media outlets shall be reviewed in a manner similar to the individual liability for international crimes, establishing the framework where individuals cannot be shielded by the names of legal entities or present their role in the operation of such entities as insignificant.

IV. Recommendations

The crimes of Russia in Ukraine extend far beyond the actual shooting and shelling of civilian objects in breach of all possible norms of international law. Quite frequently, they are rooted in mass disinformation campaigns, which significantly fuelled atrocities and contributed to the escalation of violence and suffering. Digital Security Lab Ukraine finds it necessary to bring responsible actors to justice on all possible levels, *i.e.* in the dimension of individual criminal liability on the domestic level, the State responsibility, individual criminal liability under the ICL, and on the political level regarding the individuals, who are shielded by the immunities from liability. In this regard, we recommend the following:

In the sphere of **the IHRL:**

- Ukraine shall initiate the procedures against Russia in the HRC and CERD Committee concerning its non-fulfilment of obligations to prohibit and refrain from propaganda for war and hate speech under Article 20 of the ICCPR and Articles 4(a) and 4(b) of the CERD to obtain authoritative reports of Conciliation Commissions establishing authoritative findings for further use in litigation against Russia;
- Ukrainian civil society organisations shall coordinate with a view to submit a communication to the UN Human Rights Council on Russia's non-fulfilment of its positive obligations to abstain from propaganda for war and hate speech to further appoint a UN Special Rapporteur on the matter;
- Ukraine shall work on a diplomatic track with the States Parties to the ICCPR to ensure proper implementation of Article 20 of the ICCPR, including the prohibition on propaganda for war, into the national legislation and the use of the prohibitions contained therein to prevent the spread of Russian disinformation.

In the dimension of the **PIL**:

- Ukraine shall enter diplomatic communication with the Parties to the International Convention concerning the Use of Broadcasting in the Cause of Peace concerning the potential application to the ICJ with regard to the violation by Russia of its provisions, as well as the invalidity of the Soviet Union's reservation;
- Ukraine shall supplement its claim under the Genocide Convention with additional evidence regarding the commission of incitement to genocide by Russia and actors whose actions are attributable to Russia under the law on State responsibility;
- In case of the creation of a special tribunal for adjudicating the issues related to the armed conflict launched by Russia against Ukraine, Ukraine shall

bring up the claims regarding the violation of the Geneva Conventions and Additional Protocols thereto and customary international law applicable in times of armed conflict concerning the dissemination of disinformation and related damage experienced due to such activities;
- Ukraine shall collect the evidence of speech crimes and internationally wrongful acts, as well as the proofs of attributability of such actions to Russia to further claim reparations for violation of international obligations by Russia.

In the area of the **ICL:**

- Ukraine shall conduct strategic diplomatic communication with the UN Member States regarding the adoption of the Resolution on establishing a special criminal tribunal over Russian propagandists. The Statute of the tribunal shall necessarily include incitement to international crimes as a separate crime, not as a mode of liability;
- Ukraine shall cooperate with the ICC with regard to war crimes and crimes against humanity within the scope of the ICC's jurisdiction, providing the Office of the Prosecutor with the necessary evidence;
- Following the issuance of the arrest warrants by the ICC and the special tribunal, the States shall adequately cooperate with these bodies to apprehend and transfer individuals, mentioned in the arrest warrants to the relevant judicial bodies.

For whom the bell tolls:
responsibility for disinformation in wartime

Digital Security Lab Ukraine
2022