



ARTICLE 19's submission

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

Introduction

ARTICLE 19 welcomes the opportunity to respond to the consultations organised by the UN Special Rapporteur on Freedom of Expression and Opinion (Special Rapporteur) on her forthcoming report to the Human Rights Council focused on challenges to freedom of opinion and expression in times of conflicts and disturbances.

Information operations in the context of armed conflict can have a variety of effects on the conduct of hostilities and influence the possibility of reaching a peaceful settlement. Oftentimes, all parties to an armed conflict perpetuate the conflict by promulgating a misleading, distorted image of the 'enemy'. Active combatants try to prevent information about their human rights abuses from getting to the general public. Disinformation and state propaganda activities during armed conflicts can have many harmful consequences for those most affected by the hostilities. For example, it may lead to the vilification of certain groups and encourage violence against them or distort information that is vital for the needs of civilians caught up in the hostilities.

Yet, it is precisely during times of conflict that freedom of expression and a free flow of information should be vigorously defended. Respect for freedom of expression is a crucial element in any long-term policy to promote peace and bring an end to the conflict as well as to protect lives in conflict situations. It is also necessary for adequate reporting on the conflict itself and for addressing human rights abuses – both as a cause of the conflict and a factor perpetuating it. At the same time, freedom of expression and information has been a serious casualty during armed conflicts. Often, there is a direct correlation between an increase in the intensity of a conflict and the severity of restrictions on freedom of expression. As could for instance be observed in the context of Russia's illegal invasion of Ukraine, responses by governments around the world, such as sanctions adopted against Russia, were often not grounded in international human rights law (IHRL).

ARTICLE 19, therefore, welcomes the decision of the Special Rapporteur to examine this issue and offer recommendations on how to address disinformation, misinformation and propaganda during armed conflicts in a way that complies with international standards on freedom of expression and information. Our submission outlines our key concerns in the following issues:

- The legal regimes applicable to different types of armed conflicts, including both international humanitarian law and international human rights law;

- How disinformation and state propaganda are regulated in international humanitarian law (IHL) and IHRL;
- The role of legacy media in both curbing and spreading disinformation and propaganda during armed conflicts;
- The role of social media companies in both curbing and spreading disinformation and propaganda during armed conflicts;
- The issues to freedom of expression and information raised by state and company responses to disinformation, with a particular focus on the responses following Russia's illegal invasion of Ukraine; and
- The concerning practice of internet shutdowns.

Finally, we will provide recommendations for the Special Rapporteur on how to protect and promote the right to freedom of opinion and expression while addressing disinformation and propaganda during armed conflicts.

Types of armed conflicts and the right to freedom of expression during armed conflicts

IHL – which seeks to limit the effects of armed conflicts – applies during international armed conflicts (IACs) and during non-international (also called internal) armed conflicts (NIACs).¹ Different IHL rules apply to IACs and NIACs – including rules which implicate freedom of expression and the right to information:

- IACs, which occur when at least one State resorts to armed force against another State – regardless of the intensity of the hostilities – are governed by the four Geneva Conventions of 1949 (Geneva Conventions) and Additional Protocol I of 1977 (AP I).²
- NIACs, which are armed conflicts, within the territory of a State, that involve at least one non-state armed group, are governed by much fewer IHL rules – Article 3 common to the Geneva Conventions and Additional Protocol II of 1977 (AP II). However, it is recognised that many of the rules applicable to IACs constitute customary international law, which also applies to NIACs.³ A NIAC may take place between the governmental armed forces and non-state armed groups, or between non-state armed groups only. It can only be said to exist if the non-State armed groups involved are sufficiently organised and the hostilities reach a certain level of intensity and are of a protracted nature (meaning that they extend over a certain period of time).⁴

¹ While the precise definitions of the different types have been subject of much debate, going into these detailed distinctions is considered beyond the scope of this submission. The definitions used are for the purposes of Articles 2 and 3 common to the Geneva Conventions.

² See Article 2 common to the Geneva Conventions for the definition of an IAC. See also *The Prosecutor v. Dusko Tadic*, Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction, ICTY, IT-94-1-A, 2 October 1995, para. 70 (“whenever there is a resort to armed force between states”).

³ See for a comprehensive study on customary international humanitarian law rules, ICRC, [Study on Customary International Humanitarian Law](#), 2005 (ICRC Study on Customary International Humanitarian Law).

⁴ The concept of NIAC is not defined in treaty law. In the *Tadic* Opinion and Judgment, the ICTY determined that a NIAC in the context of common Article 3 exists when the armed group involved in the conflict is sufficiently organised and the violence associated with the conflict is protracted in nature. See *The Prosecutor v. Dusko Tadic*, Opinion and Judgment, ICTY, IT-94-1-T, 7 May 1997, para 562.

IHL does not apply to situations of “internal disturbances and tensions”. They do not qualify as armed conflicts because either the violence has not yet reached a sufficient level of intensity or because the non-state armed group involved is not sufficiently organised. IHL describes such situations as “riots, isolated and sporadic acts of violence and other acts of a similar nature”.⁵

The question to what extent IHRL applies to situations of armed conflict – both IACs and NIACs – and the relationship between IHL and IHRL has been the subject of much discussions. It is today recognised that both IHL and IHRL apply during armed conflicts and that they provide complementary and mutually reinforcing protection. The rules of IHL prevail as *lex specialis* wherever they deal with a subject matter also tackled by human rights – in particular regarding the use of force and procedural safeguards for internment.⁶ This means that while the emergence of an armed conflict triggers the applicability of IHL, it does not suspend the applicability of IHRL.⁷ Only if the conflict causes a public emergency may a State derogate from its human rights obligations if strict substantive and procedural requirements are met.

IHL does not express explicit provisions that protect freedom of expression or information. There are a number of instances, in which questions impacting freedom of expression and information will be governed by IHL. This will be, for instance, around the protection of journalists and media professionals during armed conflicts – albeit these provisions are aimed at protecting individuals carrying out journalistic functions rather than at protecting the journalistic function as such – or the permissibility of propaganda. However, in most instances, there will be no IHL rules that apply to freedom of expression issues during armed conflicts. In those cases, freedom of expression and information will apply in the same manner in times of peace as they do in times of armed conflict. Yet, freedom of expression and freedom of information are often a serious casualty in both IACs and NIACs and restricted in a manner that does not comply with IHRL by meeting the requirements of the three-partite test (be prescribed by law, adopted in pursuit of a legitimate aim, be necessary and proportionate).

This is not to say that the legal assessment of a situation impacting freedom of expression and information during armed conflicts is always straightforward. Indeed, digital communications technologies have exacerbated the risks posed by manipulation of information and hate speech in times of armed conflicts and opened up new possibilities of adversarial conduct. The question arises whether and to what extent existing IHL – or IHRL – regulates such adversarial information operations or “hate speech” during armed conflicts and what the obligations of States and private actors are under IHL and IHRL. The illegal Russian invasion of Ukraine and responses of States around the world have added renewed urgency to those questions. At the same time, social media companies have had to grapple with how the spread of disinformation on their platforms may contribute to conflict situations.

We, therefore, invite the Special Rapporteur to focus her report not on all types of conflicts, violence and disturbances but on legal questions specifically dealing with freedom of expression and information during situations that meet the threshold to qualify as armed conflicts under IHL. This will

⁵ See Additional Protocol II of 1977 (AP II), Article 1(2).

⁶ See OHCHR, *International Legal Protection of Human Rights in Armed Conflict*, 2011, pp. 54-64. For example, the Human Rights Committee, in its General Comments Nos. 29 and 31, stated that the International Covenant on Civil and Political Rights (ICCPR) applied also in situations of armed conflict to which IHL was applicable. See General Comment No. 29 on Article 5 ICCPR regarding states of emergency, CCPR/C/21/Rev.1/Add.11, 31 August 2001, para. 3; and General Comment No. 31 on the nature of the general legal obligation imposed on states parties to the Covenant, CCPR/C/21/Rev.1/Add. 13, para. 11.

⁷ The idea that fundamental rights do not cease to exist during armed conflicts is further supported by Geneva Convention IV which sets out the core humanitarian law provisions relating to the protection of civilians in times of conflict. Article 27 of Geneva Convention IV has been interpreted broadly to include respect for the human rights of civilians in the widest sense. See ICRC, *Commentary on Article 27 of Geneva Convention IV*, 1958.

help clarify to what extent IHL addresses the issues of manipulation of information; what responses the international human rights framework provides; and thus, what obligations States, armed groups and private companies have under IHL and IHRL. At the same time, we hope that the Special Rapporteur will address the legal issues related to freedom of expression and other types of conflicts, violence and disturbances that do not meet the threshold in subsequent or separate reports.

Disinformation and State propaganda under IHL and IHRL

The concepts of disinformation, misinformation and state propaganda

Although the attention to the manipulation of information during both times of peace and armed conflicts has drastically increased in recent years, the discourse has at times suffered from a lack of clarity – concepts like ‘disinformation’, ‘misinformation’ or ‘propaganda’ have sometimes been used interchangeably. In the context of armed conflicts, ‘cognitive warfare’, ‘influence operations’, and ‘information operations’ are also often used without an agreed definition. This may be partly due to the fact that none of these concepts has an agreed definition in international law.

In our submission, we will mainly refer to the concepts of ‘disinformation’ and ‘state propaganda’. We observe that attempts to define the concept of ‘disinformation’ in national laws and regional standards often focus on prohibitions on “false” or “misleading” information that is knowingly shared to cause certain “harm” or detriment (e.g. to public health, social cohesion, economic gain or rights of individuals).⁸ To the extent that disinformation specifically targets groups at risk of discrimination it intersects with hate speech and can constitute incitement to violence.⁹ Disinformation may be disseminated by both private and public actors. ‘State propaganda’ on the other hand is often understood as information disseminated by official channels of States or channels controlled by them (e.g. state-controlled media), which is either false or misleading or true but framed in a way that has a manipulative effect on the target audience. ARTICLE 19 notes that both the concept of “disinformation” and the concept of “state propaganda” have been abused by power-holders, including governments, as a means to discredit media outlets, and journalists as well as suppress critical thinking and dissenting voices.

Regulation of disinformation and State propaganda under international law

Disinformation

International law (including IHL and IHRL) does not prohibit ‘disinformation’ *per se*. Where disinformation constitutes illegal ‘hate speech’ (which reaches the level of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence),¹⁰ States may be

⁸ See e.g. the Council of Europe, Information Disorder: Toward an interdisciplinary framework for research and policy making, Council of Europe report DGI(2017)09, 27 September 2017, which advocates for a distinction between “misinformation” (when false is shared, but no harm is meant), “disinformation” (when false information is knowingly shared to cause harm) and “mal-information” (when genuine information is shared to cause harm, often by moving information designed to stay private in the public sphere). The European Commission refers to disinformation as ‘false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit’. See European Commission, ‘A Multi-dimensional Approach to Disinformation’, 30 April 2018, p. 10.

⁹ For instance, Facebook was accused of allowing disinformation hate speech to fester on its site, helping incite the atrocities committed against the Rohingya in Myanmar. See e.g. New York Times, [A Genocide Incited on Facebook, With Posts From Myanmar’s Military](#), 15 October 2018.

¹⁰ Like disinformation and propaganda, the term “hate speech” is not defined in IHRL. For further guidance on the varying standards for defining and limiting “hate speech”, see ARTICLE 19, [‘Hate speech’ explained – A Toolkit](#), 2015 Edition.

under an obligation to prohibit it pursuant to Article 20 of the International Covenant on Civil and Political Rights (ICCPR). Beyond that, IHRL allows for the restriction of disinformation spread by individuals if the three-partite test of legality, legitimacy, necessity and proportionality set under Article 19(3) of the ICCPR and the applicable regional human rights instruments is met. In this context, it is important to note that the mere falsity or misleading nature of certain information will not satisfy the requirements under the tripartite test. IHRL only allows the restriction of disinformation when connected to one of the particular legitimate aims under Article 19(3) or Article 20 of the ICCPR. In addition, for restrictions on disinformation to be compatible with international standards on freedom of expression, they should be constantly applied in a proportionate manner.

It is important in this context to distinguish, on the one hand, between the human rights obligations of a State towards the individuals within their territory and towards foreign governments and foreign private actors on the other:

- Where information disseminated by an individual is restricted by their own government, they may invoke both their right to freedom of expression and freedom to impart information without interference.
- Where information is disseminated by a foreign government, the right to freedom of expression is not implicated because States cannot invoke human rights protections. However, in both these scenarios, the restricting measure may be in breach of the right of individuals *within* the jurisdiction of the restricting State to receive information without interference. ARTICLE 19 notes in this context that the right to receive and impart “information and ideas of all kinds”¹¹ by definition includes information that is correct or false. Therefore, individuals still have a right to receive foreign propaganda or even disinformation, unless the restrictions imposed by the restricting State meet the threshold for restrictions under international human rights law.

State propaganda

Like disinformation, state propaganda is not *per se* prohibited under international law¹². For example, from the perspective of a State against which an armed attack has occurred, propaganda is considered a legitimate act of self-defence as it may maintain unity, loyalty and confidence within the population at home and increase support from other States. However, not all propaganda is permissible under international law – in the context of armed conflicts it may be restricted under both IHL and IHRL, but only in narrow, specific instances.

It is well understood, and generally accepted, that state propaganda – directed both against the adversary as well as the population at home – can play a critical role after the commencement of an armed conflict. IHL is therefore generally permissive of propaganda activities, with only scarce and non-systematic limitations. It is notably permitted to engage in operations that qualify as so-called ruses of war – acts intended to mislead the adversarial party or to induce adverse forces to act recklessly.¹³ IHL also permits direct propaganda operations on the civilian population of the adverse belligerent party.¹⁴ It does, however, prohibit killing, injuring or capturing an adversary by making them

¹¹ See Article 19 ICCPR.

¹² See Tallinn Manual 2.0 on the international law applicable to cyber operations, Cambridge University Press 2017 (Tallinn Manual), p. 336.

¹³ For IACs, see Article 37(2) AP I. For NIACs, see ICRC Study on Customary International Humanitarian Law, *op.cit.*, Rule 57, pp. 203-204.

¹⁴ See Tallinn Manual, *op.cit.*, p. 421.

believe that they are entitled to, or are obliged to accord, protected status (perfidy).¹⁵ In addition, IHL seeks to limit the intentional exposure of prisoners of war for propaganda purposes by stipulating that “prisoners of war must at all times be protected [...] against insults and public curiosity”¹⁶ – a much-discussed provision following the dissemination on social media of videos of Russian Prisoners of War after the Russian invasion of Ukraine.¹⁷ Finally, under certain circumstances, State propaganda could be interpreted as constituting incitement to a breach of international humanitarian law. This may constitute a war crime and lead to the criminal liability of the person responsible under international criminal law.¹⁸

In terms of IHRL, the ICCPR expressly provides in Article 20(1) that “[a]ny propaganda for war shall be prohibited by law”. The prohibition extends to all forms of propaganda threatening or resulting in an act of aggression or breach of the peace contrary to the UN Charter, which means that propaganda for war in self-defence is permissible.¹⁹ Although Article 20 establishes an obligation for States to prohibit propaganda for war in domestic legislation, it has been suggested that the prohibition is also on State propaganda, and is not meant to only address propaganda for war by the media or other private actors.²⁰

Another issue is whether the prohibition covers only direct incitement to war or whether it also encompasses propaganda that supports an ongoing war and which may preclude a peaceful settlement.²¹ We submit that such an interpretation would be too broad since all States – including those with a certain level of democracy, accountability and rule of law – routinely convey a narrative that portrays their own war efforts in a favourable light.²² To the extent that propaganda constitutes illegal “hate speech” it will also be prohibited under Article 20(2) of the ICCPR.

The role of legacy media

Despite the rise of social media, reporting by legacy media continues to assume a central role in conflict and conflict resolution. On the one hand, sound public interest reporting is one of the key tools to counter the spread of disinformation, including in situations of armed conflicts. Efforts to support media diversity and independence and to promote media literacy are generally acknowledged to be effective in countering disinformation both in peace and times of armed conflicts.

At the same time, legacy media can also be a vehicle of propaganda and incitement to violence. This is certainly true for state-controlled media. In Myanmar, since the beginning of the February 2021 coup, the Myanmar military has taken control of all state-owned media platforms including newspapers, television and radio which it uses to disseminate its propaganda and disinformation about the post-coup-related situation in the country.

¹⁵ For IACs, see Article 37, AP I. For NIACs, see ICRC Study on Customary International Humanitarian Law, *op.cit.*, Rule 65, p 221.

¹⁶ See Article 13(2) of Geneva Convention III. There is no equivalent rule for NIACs as the “prisoner of war” status is only accorded in IACs.

¹⁷ See e.g. Washington Post, [Why you should think twice before sharing that viral video of an apparent Russian POW](#), 7 March 2022.

¹⁸ See Article 8 of the Rome Statute of the International Criminal Court.

¹⁹ See Andrei Richter, The Relationship between Freedom of Expression and the Ban on Propaganda for War, in: European Yearbook on Human Rights 2015, p. 494.

²⁰ See Sarah Joseph, Melissa Castan, The International Covenant on Civil and Political Rights (3rd Edition): Cases, Materials, and Commentary (Oxford 2013), Article 20, p. 628.

²¹ See Andrei Richter, *op.cit.*, p. 494.

²² See Sarah Joseph, Melissa Castan, *op.cit.*, Article 20, p. 628.

But also, media that is (at least seemingly) independent has been known to contribute to the incitement of violence and crimes during armed conflicts. A prominent example is the role of Radio Télévision Libre des Mille Collines (RTLM) in the Rwandan genocide. This was a nominally private radio station with only informal connections to high-level government officials and which initially appeared to be broadcasting fairly innocuous programmes. After the assassination of Melchior Ndadaye on 21 October 1993, the first democratically elected and first Hutu president of Burundi, RTLM programmes quickly became inflammatory and began to incite ethnic hatred. Its role has been regarded as crucial to creating the racial hostility that allowed the Rwandan genocide to occur.²³ More recently, mass media in Ethiopia have been reported to have contributed to fuelling ethnic violence in Ethiopia.²⁴

Another, distinct issue is that media institutions, although independent, are often overly dependent on official sources, including in countries with a strong and independent media, such as the United States. For instance, during the Vietnam War, media coverage seems to have suffered in the early years of the conflict by an overreliance on US government officials as sources of information and a reluctance to question official statements on national security issues.²⁵ In Iraq, the US administration also followed a – successful – policy of propaganda and control over the media during the lead-up to the campaign and the ensuing conflict and continuing violence.²⁶ Indeed, the US administration’s narrative was accepted and largely endorsed by the American media.

Since the Iraq war, the US administration has also been pursuing a policy of “embedding” reporters with troops on the ground. This – coupled with technological advancements – creates a sense of immediacy to the reports of journalists accompanying the troops.²⁷ At the same time, the practice has been criticised as reducing the objectivity of the embedded journalists covering the war, as their access to information is to a certain controlled by the armed forces, and embedded journalists may develop a viewpoint more sympathetic to the position pursued by the government.

If journalists or media facilities are closely involved in disseminating propaganda during an armed conflict that raises the question of at what point they may become legitimate military targets under IHL. This is by no means an abstract question as both journalists and media facilities are targeted in armed conflicts. A well-known example is the bombing of the Serbian State radio and television (RTS) building by NATO forces in 1999, which NATO justified as it neutralising a propaganda tool.²⁸

As mentioned earlier²⁹ while IHL does not provide any protections for the journalistic function as such, it does protect the individuals that report on armed conflicts. IHL distinguishes between “journalists engaged in professional missions” and “war correspondents”. The difference is that war correspondents are formally authorised to accompany armed forces. While both are considered civilians under IHL³⁰ and can therefore not be military targets, only war correspondents will receive prisoner of war status if captured, just like members of the armed forces.³¹ Media facilities will also generally be considered civilian objects, which cannot be the target of military attacks.

²³ See ARTICLE 19, [War of Words: Conflict and Freedom of Expression in South Asia](#), 2005, p. 13.

²⁴ See Accord, [The Interaction of Mass Media and Social Media in Fuelling Ethnic Violence in Ethiopia](#), 15 March 2022.

²⁵ See ARTICLE 19, *War of Words*, *op.cit.*, p. 9.

²⁶ *Ibid.*, p. 26.

²⁷ *Ibid.*

²⁸ See Alexandre Balguygallois, *Protection des journalistes et des médias en période de conflit armé*, IRRIC, March 2004, Vol. 86 No. 853, p. 52.

²⁹ *Ibid.*, p. 3.

³⁰ See Article 4A(4) of Geneva Convention III and Article 79 AP I, which reflects customary international law applicable in IACs and NIACs. See ICRC Study on Customary International Humanitarian Law, *op.cit.*, Rule 34.

³¹ See Art. 4(A)(4) of Geneva Convention III.

Both journalists and war correspondents will lose the protections afforded to them under IHL – and make them become legitimate military targets – if they take a direct part in the hostilities.³² The ICRC commentary to Article 51.3 of AP I states that "direct" participation means acts of war which by their nature or purpose are likely to cause actual harm to the personnel and equipment of the enemy armed forces."³³ In the same way, media facilities may become military targets if they make an effective contribution to military action. The debate whether journalists, war correspondents or media institutions spreading propaganda can become legitimate military targets or whether they only contribute to general "war sustaining activities" – hence not losing their protected status – can today be considered settled.³⁴ Disseminating propaganda, even though such activity supports a war effort, is not sufficient to legitimise media and media professionals as targets of military attacks. It is only if the propaganda incites war crimes or other international crimes that the media may become a legitimate military objective.³⁵

The role of social media platforms

Social media has undoubtedly been a powerful force for liberating free expression, enabling connections between people, providing tools for diverse voices to communicate ideas, perspectives and worldviews, and creating opportunities to mobilise politically to defend democracy and human rights. Over the years, however, social media platforms have increasingly become a driver of conflict. Both States and non-State actors have instrumentalised social media to crush dissent, incite armed violence, recruit members to join armed groups or contribute to crimes against humanity and war crimes. For example:

- In 2018, following the massacre of Rohingya Muslims by the military in Myanmar, the independent international fact-finding mission on Myanmar concluded that Facebook had been "a useful instrument for those seeking to spread hate in a context where, for most users, Facebook is the Internet".³⁶

In Ethiopia, violence between government forces and armed opposition groups from the country's Tigray region in November 2020 has killed thousands, including civilians, and displaced millions. The Bureau of Investigative Journalism (BIJ) found that posts on Facebook "inciting violence or making false claims designed to encourage hate between ethnic groups in Ethiopia have been allowed to circulate freely" and that Meta was aware that it was helping to directly fuel the growing tensions in Ethiopia.³⁷

- In Kenya, investigators found that in the lead-up to the August 2022 elections, "hate speech", incitement to violence against certain ethnic communities and manipulated content were spreading on TikTok with "[many of the videos containing] explicit threats of ethnic violence specifically targeting members of ethnic communities that are based within the Rift Valley region. Similar narratives stoked the post-election violence of 2007/2008, where over 1000 Kenyans died and thousands more were displaced."³⁸

³² See Article 51.3 of AP I.

³³ See ICRC Commentary on Article 51.3 of AP I.

³⁴ See Tallinn Manual, *op.cit.*, p. 528; ICRC, Interpretive Guidance on the notion of direct participation in hostilities, p. 51.

³⁵ See ICTY, Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, 13 June 2000, paras. 47, 55, 74 and 76.

³⁶ See Report of the Independent Investigative Mechanism for Myanmar, A/HRC/39/64, 12 September 2018, para. 74.

³⁷ See BIJ, [Facebook accused by survivors of letting activists incite ethnic massacres with hate and misinformation in Ethiopia](#), 20 February 2022.

³⁸ See Mozilla, [From Dance App to Political Mercenary, How disinformation on TikTok gaslights political tensions in Kenya](#), 7 June 2022.

ARTICLE 19 notes that in many ways, the problems associated with social media platforms are the same in times of peace and in times of armed conflicts, even though they are especially pronounced during armed conflicts. Their very business model, which is based on the vast collection of personal data and selling access to users' attention through targeted advertising, is harmful to human rights. Platforms are incentivised to develop recommendation algorithms that will amplify extremist, false and violent content as that tends to increase user engagement. The risks this poses for democracy, free discourse and human rights more generally are obvious. In times of armed conflicts, the implications of a business model that may thrive on spreading violence and disinformation can be devastating as there is a real risk that dehumanising and extremist content results in offline violence and harm.

The problematic business model of the biggest online platforms is routinely coupled with flawed content moderation processes. ARTICLE 19's recent report *Content Moderation and Freedom of Expression: Bridging the Gap between Social Media and Local Civil Society*³⁹ analyses current practices of content moderation in conflict-prone settings, namely Bosnia and Herzegovina, Indonesia, and Kenya. The findings in all three countries show that inadequate content moderation risks jeopardising peace and reconciliation processes and fuelling further societal tensions. Among the problems identified in the report were the inadequate levels of resources allocated by platforms to content moderation to moderate content in the local languages address certain country or regional contexts that they do not see as of strategic global importance; ineffective mechanisms set up by social media platforms to allow users to appeal content moderation decisions; insufficient transparency in content moderation practices, especially in relation to the lack of disaggregation of data on a per-country basis; and ineffective dialogues with local civil society actors.⁴⁰

The issues just described demonstrate that platforms' content moderation processes are often ill-equipped to appropriately tackle the spread of extremist and violent messages in armed conflict scenarios. Social media platforms, and in particular Meta, have also been criticised for their inconsistent responses to situations of armed conflict. Following Russia's invasion of Ukraine, platforms like Meta and Twitter were quick in adopting a range of *ad hoc* steps and employing significant resources to adjust their content moderation processes to the situation in Ukraine. At least in the case of Meta, the response included a number of policy changes, such as allowing for the call for the death of Russian President Vladimir Putin, using dehumanising language against Russian soldiers, and praising the Azov Battalion of the Ukrainian National Guard, which had previously been banned from the platform under its Dangerous Individuals and Organizations Policy.⁴¹

While quick measures to respond to the invasion were of course necessary, it did highlight the extent to which situations of violence in civilian populations of the Global South are often neglected. It further highlighted many of the issues that human rights organisations, including ARTICLE 19, have been raising for years with respect to the platforms' content moderation processes, such as a lack of transparency regarding content removal decisions or the fact that many content moderation policies fail to make allowance for public debate on individuals or groups banned due to their extremist nature.

³⁹ See ARTICLE 19, [Content Moderation and Freedom of Expression: Bridging the Gap between Social Media and Local Civil Society](#), June 2022.

⁴⁰ In terms of the consequences of under-resourced content moderation, the report found the findings of the TBIJ's investigation into the situation in Ethiopia particularly revealing. The BIJ has been able to document "a litany of failures" in Facebook's content moderation. Facebook 'is said to have frequently ignored requests for support from fact-checkers based in the country and some civil society organisations say they have not met with the company in 18 months. While content flagged as disinformation by Facebook's partner fact-checkers (who are mostly based outside Ethiopia) led to quick action by the platform, no action was taken when a local fact-checking organisation reported content. See ARTICLE 19, *Content Moderation and Freedom of Expression*, *op.cit.*, pp. 26 - 27.

⁴¹ See The Intercept, [Facebook's Ukraine-Russia Moderation Rules Prompt Cries of Double Standard](#), 13 April 2022.

Under the UN Guiding Principles on business and human rights, companies have obligations to respect human rights and to offer remedy. It is arguable that the business model of online platforms today, which has been shown to be inherently harmful for human rights, makes it impossible for them to respect human rights. Regulatory responses should recognise this and adopt solutions based on transparency, data protection and sound competition policies, including the unbundling of hosting from content curation and interoperability of large platforms⁴².

State responses to curb disinformation and state propaganda in armed conflicts

While disinformation and State propaganda can at times be problematic, so too are many of the responses of States and companies seeking to curb what they perceive as disinformation by both States and private actors. Keeping in mind that IHL is largely permissible of disinformation and state propaganda and that IHRL only allows restriction of disinformation and propaganda under the very limited conditions set out in Articles 19 and 20 of the ICCPR, ARTICLE 19 finds that many of the proposed measures adopted by States and private companies to tackle disinformation have not been compliant with IHRL. As noted earlier, in many instances, governments have used the concepts of disinformation and propaganda to silence dissenting voices. Recent and particularly problematic examples include proposed legislation in Turkey⁴³ and Nigeria⁴⁴.

In the context of armed conflicts, a much-discussed decision has been the EU sanction imposed on the Russian state-controlled media (first RT and Sputnik, and subsequently also Rossiya RTR/RTR Planeta, Rossiya 24/Russia 24, and TV Centre International) following Russia's illegal invasion of Ukraine. The ban was adopted with the justification that 'RT and Sputnik are essential and instrumental in bringing forward and supporting Russia's aggression against Ukraine'.⁴⁵ The EU Commission interpreted the enacted acts as imposing the obligation for search engines to remove any RT or Sputnik content.⁴⁶ In addition, RT's and Sputnik's content, as well as content re-posted *by users*, is to be removed from social media platforms.

The ban imposed by the EU has been criticised as problematic for various reasons. Among the key reasons are the fact that the ban prevents access to information for individuals in the EU, including journalists and researchers, which are precluded from developing a first-hand understanding of the narratives of Russian propaganda and from reporting on it. It has also been argued that a ban prevents EU citizens and the media to formulate a resilient counter-response to propaganda and disinformation disseminated by RT and Sputnik.⁴⁷ Prohibitions generally do not constitute an effective tool to address propaganda. The Government of Switzerland took another approach and concluded that opposing the false information spread by RT and Sputnik is much more likely to be effective in countering their propaganda.⁴⁸

The EU ban is reminiscent of the prohibition of listening to enemy broadcasts adopted by a few belligerent countries during the Second World War, including Germany. Since then, various human

⁴² See ARTICLE 19, [Taming Big Tech: Protecting freedom of expression through the unbundling of services, open markets, competition, and users' empowerment](#), December 2021.

⁴³ See ARTICLE 19, [Turkey: Scrap the 'disinformation bill' and stop digital censorship](#), 8 June 2022.

⁴⁴ See ARTICLE 19, [Nigeria: Prohibitions on false news and criminal defamation should be repealed](#), 8 November 2021.

⁴⁵ See European Commission, [Ukraine: Sanctions on Kremlin-backed outlets Russia Today and Sputnik](#), 2 March 2022.

⁴⁶ See Reason, [EU Orders Google to Vanish Russian Government Sites \(RT and SputnikNews\) from European Search Results](#), 11 March 2022.

⁴⁷ See Natali Helberger and Wolfgang Schulz, [Understandable, but still wrong: How freedom of communication suffers in the zeal for sanctions](#), 10 June 2022; Reuters, [Dutch journalists, rights group file lawsuit challenging EU ban on RT, Sputnik](#), 25 May 2022.

⁴⁸ InfoBae, [Switzerland expands sanctions against Russia but decides not to censor Russian media](#), 25 March 2022.

rights instruments have adopted a right to information. While it is to be determined whether the Russian state-controlled media as the agents of the State can invoke a right to freedom of expression, the ban does affect the right of individuals in the EU to freely receive information of “any kind” under Article 19 of the ICCPR. It has been also argued that while restrictions on the right to information are possible, the EU has not provided sufficient evidence that would such a ban is legitimate, proportionate, and necessary. In this respect, ARTICLE 19 notes that any justification on the basis of public order and security is unlikely to be pertinently convincing, given that the EU is not directly engaged in an armed conflict with Russia and in light of the limited distribution and impact of these media in the EU countries. The EU should demonstrate that RT and Sputnik’s programmes actually constitute a serious and immediate threat to public order and security to justify a ban in all EU Member States. In addition, many of the programs of the Russian state-controlled media, in particular as broadcast abroad, report on issues entirely unrelated to the war. We further note that in democratic countries and under the international freedom of expression standards, suspending or cancelling licenses for audiovisual media should be decided by independent regulators and not by political institutions.

At least equally problematic has been the ban imposed by Russia on several foreign news organisations such as the BBC, Voice of America or Deutsche Welle, for spreading what it alleged was false information about its war in Ukraine.⁴⁹ Russia also proceeded to amend its Criminal Code following its invasion of Ukraine and created new offences that punished discrediting the Russian armed forces, called for sanctions against Russia, and disseminated “false information” about the armed forces. As a consequence, both local media outlets (such as the Novaya Gazeta newspaper) and foreign journalists were forced to suspend their coverage of the conflict due to fears of being imprisoned under the new law for up to 15 years.⁵⁰

Internet shutdowns

There has been a sharp uptick in Internet shutdowns in recent years. In 2021, ARTICLE 19 documented 182 internet shutdowns in 34 countries, with seven new countries starting to use this tactic.⁵¹ Internet shutdowns are also often used in response to situations of armed conflicts or internal disturbances. Examples from 2021 include Ethiopia, Myanmar and Israel.⁵²

We note that during an armed conflict, online connectivity becomes ever more important. Internet shutdowns will put many civilians at risk because they lack access to vital information.

We note further that the use of internet shutdowns has been increasingly used to cover up the commission of human rights atrocities. ARTICLE 19 has, for instance, documented that in Myanmar, Internet shutdowns have been imposed across regions where resistance against the military has been most intense, and where arson attacks by the junta have been most widespread as both a punishment, and cover for atrocities.⁵³ Other examples include Iran⁵⁴, Uganda⁵⁵, Sudan⁵⁶, and the Central African

⁴⁹ See Reuters, [Russia blocks access to BBC and Voice of America websites](#), 4 March 2022.

⁵⁰ See Marko Milanovic, [The legal death of free speech in Russia](#), 8 March 2022.

⁵¹ See ARTICLE 19, [The Global Expression Report 2021](#), p. 13..

⁵² See Access Now, [The Return of Digital Authoritarianism – Internet shutdowns in 2021](#), 24 May 2022, pp. 21 - 24.

⁵³ See ARTICLE 19, [Myanmar: Internet shutdowns shrouding torchings and killings](#), 23 June 2022; or ARTICLE 19’s comprehensive report in internet shutdowns in Myanmar, [Unplugged in Myanmar](#), 9 July 2021.

⁵⁴ See ARTICLE 19, [Iran: Near total Internet shutdown in Khuzestan as food protests continue](#), 12 May 2022.

⁵⁵ See Association for Progressive Communications, [Joint letter on internet shutdown in Uganda](#), 23 February 2016.

⁵⁶ See Human Rights Watch, [Sudan – End network shutdown immediately](#), 12 June 2019.

Republic⁵⁷. Where employed during times of armed conflicts, Internet shutdowns that are used to conceal crimes and other serious human rights violations, contribute to a culture of impunity. In this context, freedom of expression is necessary for adequate reporting on the conflict itself and for addressing human rights abuses – both as a cause of the conflict and a factor perpetuating it.

Shutdowns can never be compliant with IHRL. In June 2016, the UN Human Rights Council explicitly condemned 'measures to intentionally prevent or disrupt access to or dissemination of information online in violation of international human rights law' and called on all States to refrain from and cease such measures.⁵⁸ In their report on Disinformation and freedom of opinion and expression, the Special Rapporteur has Internet shutdowns constitute an “inherently disproportionate response”.⁵⁹

Conclusion and recommendations

It is ARTICLE 19's view that enabling a strong culture for freedom of expression and information does not only serve to protect human rights and democracy in times of peace but also constitutes a powerful tool to prevent armed conflicts and encourage diplomacy. ARTICLE 19 further notes that resorting to censorship and prohibition of speech is not an effective tool to fight censorship and propaganda. A more effective antidote is indeed the promotion of a vibrant, pluralistic, professional, ethical and viable media ecosystem, which is entirely independent of those in power. Therefore, many of the recommendations we provide are those which we consider effective in tackling disinformation in times of peace. We believe that they will also provide the most effective means of preventing the most harmful aspects of armed conflicts.

The Special Rapporteur should recommend to **all parties to an armed conflict** (including non-State actors):

- Parties to an armed conflict should not consider journalists and media outlets to be legitimate military, even if they are perceived to engage in propaganda activities. They should recognise that only direct and demonstrable incitement of war crimes or other international crimes by media and media professionals can justify a loss of protected status.
- They should ensure that they do not spread propaganda and disinformation, which may incite violence or have other harmful effects and prolong an armed conflict. They should abandon intentional propaganda or disinformation campaigns and not engage in media manipulation of any kind.

The Special Rapporteur should recommend to **States**:

- States should adopt comprehensive right to information laws and comply with the principles of maximum transparency of public administration. States that currently have freedom of information laws should prioritise implementation and consider amendments to bring those laws in line with current international and regional standards and best practices. States should refrain from adopting unduly broad exceptions to the right to information on national security grounds.

⁵⁷ See Access Now, [Access submits evidence to International Criminal Court on net shutdown in Central African Republic](#), 17 February 2015.

⁵⁸ See Human Rights Council, [The promotion, protection and enjoyment of human rights on the Internet](#), UN Doc A/HRC/32/L.20, 27 June 2016, para 10.

⁵⁹ See Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, [Disinformation and freedom of opinion and expression](#), UN Doc A/HRC/47/29, 13 April 2021, para 51.

- Governments should ensure strong protections for whistle-blowers. Many States already have whistle-blower protections in freedom of information legislation or standalone laws. Those States should focus on consistent implementation to protect those raising concerns about government misconduct or various policy failures. Those without dedicated whistle-blower legislation should refrain from prosecutions or restrictions on those who release information in the public interest.
- Since some disinformation concerns target groups at risk of discrimination and intersects with 'hate speech', States should adopt positive policy measures to combat 'hate speech' and intolerance that are consistent with international human rights standards and best practices. The Human Rights Council Resolution 16/18 and the Rabat Plan of Action offer important guidance in this regard.
- States should take steps to ensure a free, independent and diverse media environment, in particular through clear regulatory frameworks that ensure self-governance and independence for the media and broadcasting sector. States may also consider supporting independent public service media with a clear mandate to serve the public interest. State authorities should end the harassment of journalists reporting on government activities and other issues of public interest. States should support and promote the existence and effective implementation of ethical standards by different media actors.
- Government officials and public authorities should ensure that they do not spread misinformation, and governments should abandon intentional propaganda or disinformation campaigns. Governments should at all times adopt measures to promote media and digital literacy, both generally and in relation to specific issues. This could include incorporating media and digital literacy lessons into school curriculum and engaging with civil society and social media platforms on similar efforts.
- Refrain from the practice of using embedded journalists. If journalists are given access to accompany armed forces, States should refrain from controlling the information journalists are given or take any measures that will affect their independent reporting.
- There are specific tools available to fight biased and misleading information, including rules on balance and accuracy in broadcasting; guarantees of the independence of media regulators; vibrant public service broadcasting with a special mission to include all viewpoints; a clear distinction between fact and opinion in journalism and transparency of media ownership.
- States should ensure that any responses to foreign wars are grounded in IHRL and that a human rights impact assessment is performed before such measures are adopted.
- States should under all circumstances refrain from imposing Internet shutdowns.

The Special Rapporteur should recommend that **media companies:**

- Media companies should abide by the highest ethical standards of objectivity and independence.
- They should refrain from disseminating any communication that could fuel tension and exacerbate conflict situations.
- They should remain independent from the government and remain impartial in their reporting and refrain from uncritically reporting information conveyed by official sources. They should avoid engaging in reporting that is too polarised along political, ethnic or religious lines.

- Retain an independent ownership structure.
- Refrain from participating in the practice of “embedded” journalists. If it is necessary to accompany armed forces, they should ensure that they retain independence and objectivity when reporting on activities of the armed forces.

The Special Rapporteur should recommend that **social media platforms** at minimum:

- Articulate clear, accessible and easily understood policies governing disinformation, propaganda and “hate speech” on their platforms in line with the “human rights by default” approach. All content rules, as well as reporting and appeals processes, should be available to the users in the language in which they engage with the platform.
- To the extent that social media companies develop policies specific to certain types of disinformation, propaganda or ‘hate speech’, these policies should likewise be precise and nuanced in line with the standard of legality set out in international human rights law.
- Decisions on content moderation should be made by people familiar with the relevant language and with sufficient awareness and understanding of the linguistic, cultural, social, economic and political dimensions of the relevant local or regional context.
- Content moderation processes should not disadvantage users on the basis of language, country or region.
- Social media companies should put in place internal complaint and dispute resolution mechanisms, including for the wrongful removal of content or other restrictions on their users’ freedom of expression. In particular, individuals should be given detailed notice of a complaint and the opportunity to respond prior to content removal. Internal appeal mechanisms should be clear and easy to find on company websites.
- Social media companies should publish comprehensive transparency reports, including detailed information on their decision-making processes; the tools they use to moderate content, such as algorithms and ‘trusted flagger’ schemes; and content removal requests received and actioned on the basis of their terms of service. They should also provide additional information on appeals processes, including the number of appeals, received and their outcomes. Such information should be disaggregated on a per-country basis.
- Social media companies should engage with local civil society actors to monitor and detect “hate speech” on social media tailored to the local context and languages. They should further collaborate with civil society groups focused on digital rights to ensure that content moderation and removal processes are aligned with community needs.