doi: 10.21783/rei.v11i1.892

SUBMETIDO: 07 JAN. 2025
ACEITO: 07 JAN. 2025

DECONSTRUCTED REALITY, TANGIBLE HARM: FROM ONLINE FAKE NEWS TO GENOCIDE

BÁRBARA LUIZA COUTINHO DO NASCIMENTO [1]

ABSTRACT: This paper studies the historical development of the international crime of incitement to genocide, from its origins in Nuremberg to its contemporary online format, identifying the role of the media and fake news in it. Showing this progression and describing its underlying patterns is a research effort for legally interpreting a phenomenon. After identifying the elements of the crime according to international law and presenting the new aspects brought by cybercommunications, the paper analyses the Myanmar case to demonstrate how the existing legal framework may be applied to social media posts spreading fake news. Lastly, concerning speeches falling outside the scope of the norm, the paper proposes a new crime prohibiting the conduct of systematically creating or distributing fake news online when such conduct constitutes computational propaganda with the intent to harm groups protected under the Genocide Convention and assesses the democratic justification of the proposal.

KEYWORDS: International Criminal Law; Incitement to Genocide; Online Fake News; Computational Propaganda.

INTRODUCTION

Can the intentional distribution of fake news using social media platforms be criminalised as incitement to genocide under international law? If so, how? When the intentional distribution of fake news online does not reach the threshold of incitement but constitutes computational propaganda with the intent to harm a group protected under the Genocide Convention, should it be criminalised? This paper seeks to answer these research questions.

Inciting genocide is a crime under international law, expressly prohibited by the 1948 Convention on the Prevention and Punishment of the Crime of Genocide (hereinafter 'the Convention' or 'Genocide Convention').[2] Its origins lie in the 1945/46 Nuremberg trial of Julius Streicher by the International Military Tribunal

[1] Master of Laws in Information Technology Law (The University of Edinburgh), Master of Laws in Theory and Philosophy of Law (UERJ), Public Prosecutor (Rio de Janeiro State Prosecutor's Office), 2022 Young Global Leader (World Economic Forum).

[2] Convention on the Prevention and Punishment of the Crime of Genocide (adopted 9 December 1948, entered into force 12 January 1951) 78 UNTS 277 (Genocide Convention) art III, c.

(IMT).[3] Streicher was the editor of an influential anti-Semitic magazine.[4] After the Convention came into force, the crime occurred during the Rwandan genocide, when radio broadcasts and newspapers were used.

With the emergence of the Internet, new ways of expressing opinions and thoughts have also emerged. Initially celebrated as an instrument that could lead to emancipation through access to information, the Internet showed its ugliest face in the Myanmar genocide case, which is currently under investigation at the International Criminal Court (ICC)[5] and on trial at the International Court of Justice (ICJ).[6]

The role of social media in spreading fake news and hate speech in the Myanmar genocide was recognised by a United Nations (UN) Fact-Finding Mission.[7] Although the Mission's Report focuses more on hate speech than on fake news, the role of fake news cannot be overlooked and demands specific research as fake news campaigns are more subtle but no less perverse. By undermining connection with reality, they are more penetrating. In Hannah Arendt's lessons, since the propaganda of totalitarian regimes cannot transform reality, they distort the way of experiencing and understanding it, adopting persuasive logic as a guide to action.[8]

This paper first establishes the international legal framework of the crime of incitement to genocide, historically built on written media and radio broadcasts. The Myanmar case will then be introduced to demonstrate how the identified framework can be applied to cybercommunications. Lastly, the paper proposes a new crime and assesses the democratic justification of the proposal.

2. THE INTERNATIONAL LEGAL FRAMEWORK OF INCITEMENT TO GENOCIDE

This section will review international case law on the crime of incitement to genocide. The objective is to show its historical development and unravel its elements, laying the basis for discussing whether these concepts are still valid and applicable to cybercommunications.

---

[3] Richard Wilson, 'Inciting Genocide with Words' (2015) 36 MichJIntlL 277 283.

[4] Maggi Eastwood, 'The Emergence of Incitement to Genocide Within the Nuremberg Trial Process: The Case of Julius Streicher' (PhD thesis, University of Central Lancashire 2006) <https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.438233> accessed 5 December 2022, vol I, 141-143.

[5] *Situation in the People's Republic of Bangladesh/Republic of the Union of Myanmar* (Decision) ICC-01/19-27 (14 November 2019)

[6] *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v Myanmar)* (Request for the Indication of Provisional Measures: Order) General List No 178 [2020] ICJ

[7] UNHRC 'Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar' (17 September 2018) UN Doc A/HRC/39/CRP.2

[8] Hannah Arendt, *Origins of Totalitarianism* (2nd edn, Meridian Book 1958) 471-472.

CC BY

A. THE ORIGINS OF THE CRIME

The 1945/46 Nuremberg trial of Julius Streicher inspired the criminalisation of incitement to genocide in international law. Streicher was a member of the Nazi Party, editor and owner of a magazine called *Der Stürmer*, which systematically used hate speech and fake news to persecute Jews.[9] For example, in a series of publications, *Der Stürmer* accused Jews of performing 'ritual murders' on Christian children, reporting events that occurred mainly during the Middle Ages as contemporary threats.[10] Streicher was also accused of lying under oath about his reasons for demolishing a synagogue in Nuremberg, which he testified in court was done for architectural reasons.[11]

When the IMT judged him, no international custom, precedent, or law existed criminalising the conduct of 'inciting mass murder through words'.[12] Besides, the IMT Charter established the crimes over which the IMT had jurisdiction and incitement to genocide was not expressly defined.[13]

Thus, in the absence of an explicit criminalisation of incitement, the possible legal construction under the Charter was for the prosecution to argue that in his speeches and writings, Streicher persecuted the Jewish people on 'political and racial grounds', committing a crime against humanity.[14] The strategy worked, and he was sentenced to death by hanging.[15]

Although the judgment discussed Streicher's long-term anti-Semitic propaganda, it quoted specific phrases to support the conviction, citing articles published in *Der Stürmer* written by his 'own hand which demanded annihilation and extermination' of Jews 'in unequivocal terms', such as '[a] punitive expedition must come against the Jews in Russia' and '[t]he Jews in Russia must be killed'.[16] In other words, the Tribunal sought phrases constituting direct[17] calls to action to convict Streicher.

On the other hand, the absence of evidence of direct calls to action led the IMT to acquit Hans Fritzsche, head of the Propaganda Ministry's Radio Division, of the charges of inciting and encouraging 'the commission of war crimes by deliberately

[9] Eastwood (n 3) vol I, 141-143.

[10] ibid 154-157.

[11] ibid 100-101.

[12] ibid 1. See also Wilson (n 2) 283.

[13] Agreement for the prosecution and punishment of the major war criminals of the European Axis (adopted 8 August 1945) 82 UNTS 279

[14] The International Military Tribunal for Germany, 'Judgement: Streicher' (The Avalon Project, Yale University) <https://avalon.law.yale.edu/imt/judstrei.asp> accessed 5 December 2022

[15] ibid

[16] ibid

[17] The concept of 'direct' incitement will be discussed in the next section.

CC BY

falsifying news to arouse in the German people those passions which led them to the commission of atrocities'.[18]

According to the judgment, Fritzsche 'did not urge persecution or extermination of Jews', and there was no evidence that he knew he was spreading 'false news'.[19]

Jurists agree that the IMT did not find Fritzsche's speech direct enough to be considered criminal.[20] However, one year after his Nuremberg trial, further incriminating evidence was found, and he was convicted by the German Denazification Court and sentenced to nine years of hard labour.[21]

Thus, it is possible to conclude that Streicher's and Fritzsche's trials demonstrate the need for international law to systematise the criminalisation of the conduct of inciting genocide.

B. THE ELEMENTS OF THE CRIME OF INCITEMENT TO GENOCIDE

In 1948, the UN General Assembly adopted the Genocide Convention. It criminalised 'direct and public incitement to commit genocide'.[22] Later, the Statute of the International Criminal Tribunal for Rwanda (ICTR)[23] repeated the Convention's text. The Rome Statute of the ICC reproduced its essence but slightly changed the wording, establishing criminal responsibility for a person who 'directly and publicly incites others to commit genocide'.[24]

Regarding the *actus reus* and the judicial interpretation of what should be considered 'public and direct incitement' in international law, the ICTR precedents became the primary source.[25]

An incitement is considered public when done in a 'public place' or through channels able to reach a large or indeterminate audience.[26] The idea is that the inciter wants to mobilise the audience to whom he speaks.[27] If incitement occurs in private, selectively, it loses the 'mob' factor and becomes complicity.[28]

---

[18] The International Military Tribunal for Germany, 'Judgement: Fritzsche' (The Avalon Project, Yale University) <https://avalon.law.yale.edu/imt/judfritz.asp> accessed 5 December 2022

[19] ibid

[20] Eastwood (n 3) vol II, 245; Susan Benesch, 'Vile crime or inalienable right: defining incitement to genocide' (2008) 48 VaJIntlL 485 510; Wibke Timmermann and William Schabas, 'Incitement to Genocide' in Paul Behrens and Ralph Henham (eds), *Elements of Genocide* (Routledge 2013) 156.

[21] Eastwood (n 3) vol II, 249.

[22] Genocide Convention (n 1)

[23] UNSC Res 955 (8 November 1994) UN Doc S/RES/955

[24] Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3 art 25(3)'e'.

[25] William Schabas, *Genocide in International Law: The Crime of Crimes* (2nd edn, CUP 2009) 325-326.

[26] *Prosecutor v Akayesu* (Judgment) ICTR-96-4-T, T Ch I (2 September 1998) [556]. See also Carsten Stahn, *A Critical Introduction to International Criminal Law* (CUP 2018) 97-98.

[27] Schabas (n 24) 329.

[28] ibid

CC BY

The concept of direct relies on the idea of an immediate call to criminal action, excluding indirect propaganda.[29] After the Genocide Convention entered into force, the ICTR developed the concept.

In *Akayesu*, the ICTR decided that to be considered direct, incitement must 'specifically provoke another to engage in a criminal act', excluding 'mere vague or indirect suggestion', that 'the direct element of incitement should be viewed in the light of its cultural and linguistic content', and that 'incitement may be direct, and nonetheless implicit'.[30]

In the Media Case,[31] the ICTR implicitly debated if lies and distortions of facts influenced the direct element of incitement.[32] Like the IMT, the ICTR found that spreading fake messages without calls to action could not be considered criminal. For example, in Rwanda, media broadcasts constantly exhorted an alleged Tutsi wealth over Hutus,[33] so the Tribunal analysed one specific broadcast affirming that the Tutsi had 'all the money' and considered that it was a 'generalisation' carrying 'hostility and resentment' but did not constitute direct incitement because it did 'not call on listeners to take action'.[34] Concerning one specific 'broadcast stating that 70% of the taxis in Rwanda were owned by people of Tutsi ethnicity', the ICTR did not rule on its veracity but affirmed that if it were false, it would indicate an intention 'to promote unfounded resentment and inflame ethnic tensions'.[35] Finally, the Appeals Judgment analysed linguistic structures of broadcasts on a case-by-case basis, only upholding the conviction of incitement to genocide when it found specific calls to action.[36]

Therefore, according to the current legal understanding, incitement to genocide is direct when it aims to trigger a course of action by the audience towards the commission of the crime (call to action). It does not matter if the wording is explicit or not. What matters is the meaning of the message to the audience and how people are expected to behave afterwards, regardless of whether someone acts or not.

As already mentioned, in Streicher's case, the IMT used quotes from him that fit the definition of direct incitement to support the conviction.[37] In contrast, in Fritzsche's case, the absence of evidence of direct calls to action led to acquittal.

Defined the *actus reus*, now the *mens rea* of the crime must be defined.

[29] Benesch (n 19) 508.

[30] *Prosecutor v Akayesu* (n 25) [557]

[31] *Media Case* (Judgment) ICTR-99-52-T, T Ch I (3 December 2003) [5]-[7]

[32] Timmermann and Schabas (n 19) 160.

[33] *Media Case* (Judgment) (n 30) [364]-[365]

[34] ibid [1021]

[35] ibid

[36] *Media Case* (Appeals Judgment) ICTR-99-52-A, A Ch (28 November 2007) [738]-[775]

[37] Timmermann and Schabas (n 19) 156.

CC BY

Incitement to genocide requires specific intent (*dolus specialis*),[38] meaning that the agent must want 'to destroy, in whole or in part, a national, ethnical, racial or religious group, as such',[39] which are the groups protected by the Convention. According to Schabas, 'this is of no practical difficulty, because the *mens rea* is generally obvious enough from the content of the message'.[40]

Lastly, incitement is an inchoate offence, meaning that the occurrence of the result is not necessary for the crime to be committed.[41] Thus, to achieve the preventive scope, the conduct is punishable even if no one is indeed incited, attempts or succeeds in committing genocide.[42]

3. INCITEMENT GOES ONLINE

The cases studied so far demonstrate that incitement to genocide has, throughout the 20th century, walked alongside different mass media. In a seemingly linear evolution, with the popularisation of cybercommunications, it would not be long before the 21st century had to face its own version of the conduct, as this section intends to demonstrate.

The Rohingya are an ethnic and religious Muslim minority historically suffering human rights violations in Myanmar.[43] In 1954, local authorities described them as an 'indigenous group'[44] and the Rohingya claim to have a 'longstanding connection' to their territory.[45] Nevertheless, especially after 1978, the country's government has systematically denied them recognition as nationals, insisting on the narrative that they are foreign invaders illegally immigrating from Bangladesh, calling them 'Bengali' to represent such status.[46] They have been denied birth certificates, citizenship, and the right to political participation.[47]

Violence against the Rohingya intensified during the 2010s, calling international attention.[48] From 2016 and escalating in 2017, the local military conducted what they called 'clearance operations', supposedly to contain terrorist actions.[49] The UN established a Fact-Finding Mission that found evidence that the operations caused disproportionate damage to Rohingya civilians, including village burning, forced

[38] Malcolm Shaw, *International Law* (8th edn, CUP 2017) 318-321.

[39] Genocide Convention (n 1) art II. See also Timmermann and Schabas (n 19) 166.

[40] Schabas (n 24) 326.

[41] Timmermann and Schabas (n 19) 147-148.

[42] ibid

[43] UNHRC (n 6) [458]

[44] ibid [473]

[45] *The Gambia v Myanmar* (n 5) [14]

[46] UNHRC (n 6) [460]-[490]

[47] ibid

[48] *The Gambia v Myanmar* (n 5) (Sep. Op. Cançado Trindade) [22]

[49] UNHRC (n 6) [751]

CC BY

displacement, sexual violence, and mass murder, among other acts considered genocide and crimes against humanity.[50]

In November 2019, the Gambia filed an application against Myanmar at the ICJ claiming a violation of the Convention,[51] which also obliges States to prevent genocide.[52] In January 2020, the Court indicated provisional measures. Among other determinations, the ICJ ordered Myanmar to guarantee that no acts of direct and public incitement to genocide were committed against the Rohingya in its territory.[53] Additionally, the ICC authorised the Prosecutor to open an investigation into the situation.[54] Procedurally, both cases are still incipient.

The role of the Internet, particularly social media, in spreading fake news and inciting hate against the Rohingya is noteworthy. The Mission's report describes the online public sphere in Myanmar as younger than other countries (due to censorship conducted by the dictatorship until 2011), permeated with hate speech and fake news, and falling easily into manipulation tactics.[55] Many people in Myanmar lack digital literacy, resulting in an online experience restricted to Facebook, which they consider a credible platform because authorities use it for official statements.[56]

Hence, although anti-Muslim and anti-Rohingya propaganda are longstanding problems in Myanmar, the Internet increased their complexity and reach. According to the Mission's report, ordinary discourse argues that there exists no ethnic group identified as Rohingya, stating that they are all 'Bengali' terrorists who illegally entered the country and created false claims to steal territory, depicting them as the ones who commit atrocities and lie.[57] This narrative was promoted, for example, on the Facebook account of the Myanmar President's spokesperson in 2012, days before violence resulted in the murder of ten Muslims.[58] Additionally, it has been echoed in Facebook posts of government departments, high-rank military officials, and other authorities.[59]

There were calls for ordinary people to engage in hostilities. For example, after implying that Myanmar's territorial integrity was at risk, in October 2017, Commander-in-Chief Min Aung Hlaing posted on Facebook that 'every citizen has the duty to safeguard race, religion, cultural identities and national interest',

[50] ibid [751]-[1482]

[51] *The Gambia v Myanmar* (n 5)

[52] Genocide Convention (n 1) art I.

[53] *The Gambia v Myanmar* (n 5) [86]

[54] *Situation in the People's Republic of Bangladesh/Republic of the Union of Myanmar* (n 4)

[55] UNHRC (n 6) [1342]-[1354]

[56] ibid

[57] ibid [702]-[1379]

[58] ibid [705]

[59] ibid [1324]-[1338]

CC BY

adding that 'the national defence duty falls on every citizen'.[60] This post seems to fit the definition of direct and public incitement, as will be demonstrated in the next section.

In another event, a viral social media post accused Muslim men of rape. Later, it was proved that the accusation was fabricated, but not before causing riots that killed one Muslim and one Buddhist.[61]

Among the online tactics used, observers reported fake accounts and trolls.[62] Militaries were accused of 'creating fan pages for local Burmese pop stars and celebs (…) accumulating over a million combined followers that abruptly swapped into propaganda accounts to spread anti-Rohingya messaging'.[63] 'Clickbaits' drew users' attention to other kinds of content and redirected them to anti-Rohingya hate messages.[64] False pages were created imitating those belonging to Rohingyas and pretending that they were spreading violence.[65] Lastly, pages pretending to be independent news sources spread fake and hate messages against the Rohingya.[66]

The examples shown appear to constitute a systematic campaign of fake news and hate speech against the Rohingya. Local civil and military authorities, each with hundreds of thousands to millions of Facebook followers, were accused of neglecting their duties and increasing the spread of fake news.[67] They deceived public opinion by denying the atrocities committed against the Rohingya and by distributing information proven false on Facebook, affirming, for example, that forced displacement was due to people fleeing for fear of terrorists instead of being caused by the military.[68] They downgraded severe allegations of sexual violence as

[60] ibid [1341]

[61] ibid [744],[1325]

[62] Progressive Voice and others, 'Hate Speech Ignited: Understanding Hate Speech in Myanmar' (8 October 2020) Joint Report <https://hrp.law.harvard.edu/wp-content/uploads/2020/10/20201007-PV-Hate-Speech-Book-V-1.4-Web-ready1.pdf> accessed 5 December 2022 62. See also Steve Stecklow, 'Why Facebook is losing the war on hate speech in Myanmar' *Reuters* (15 August 2018) <www.reuters.com/investigates/special-report/myanmar-facebook-hate> accessed 5 December 2022.

[63] Joshua Citarella, 'There's a new tactic for exposing you to radical content online: the 'slow red-pill'' *The Guardian* (15 July 2021) <www.theguardian.com/commentisfree/2021/jul/15/theres-a-new-tactic-for-exposing-you-to-radical-content-online-the-slow-red-pill> accessed 5 December 2022

[64] Progressive Voice and others (n 61) 62.

[65] ibid 63.

[66] ibid. See also Samantha Bradshaw and others, 'Country Case Studies Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation' (2021) Oxford Computational Propaganda Research Project <https://medleysites.oii.ox.ac.uk/demtech/wp-content/uploads/sites/127/2021/03/Case-Studies_FINAL.pdf> accessed 5 December 2022 272-273.

[67] UNHRC (n 6) [1327]-[1329]

[68] ibid [1339]-[1423]

CC BY

'rumours' and shared fabricated photos allegedly showing Rohingyas burning their own houses.[69]

Comments on social media posts tend to follow the narrative, engage in violent speech patterns, advocate hatred, consider the Rohingya invaders, instigate their 'eradication', and silence those who oppose the narrative.[70]

It remains uncertain to what extent political bots (robots 'programmed with human attributes or abilities in order to pass as genuine social media users' 'used for political manipulation')[71] were used. Facebook did not share significant data with international bodies[72] but excluded accounts from Myanmar for 'coordinated inauthentic behaviour'.[73] An independent analysis conducted on Twitter accounts, which has also been used but to a lesser extent, suggested that the Myanmar digital crowd worked in an orchestrated but not automated way.[74] However, the Oxford Computational Propaganda Research Project found that bots are used in Myanmar and suspected human-curated bots could be running.[75] These are called cyborgs and can avoid automated detection.[76]

Consequently, supposing that at least some of the comments on hate posts come from real users, while not an element of the crime of direct and public incitement to genocide, the vociferous way people react to aggressive posts demonstrates that such posts are capable of triggering intense emotions in the audience. Precisely the kind of emotion that can trigger a causal course of brutal events. On the other hand, if reactions to posts are staged, it reveals another element in the intention to create the general hate climate.

The Mission asks whether such profusion of online hate speech and fake news is linked to tangible harm, concluding that a connection between Facebook posts and the violent climate exists, calling for further research on its extension.[77]

The international community noticed Facebook's disregard for the Myanmar crisis. According to its Community Standards, the company should have removed much of the false and hateful content aforementioned but was publicly criticised

[69] ibid [1270]-[1340]

[70] ibid [1312]-[1319]

[71] Samuel Woolley and Philip Howard, 'Introduction: Computational Propaganda Worldwide' in Samuel Woolley and Philip Howard (eds), *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (OUP 2019) 6.

[72] UNHRC (n 6) [1351]

[73] Facebook, 'Removing Myanmar Military Officials From Facebook' (28 August 2018) <https://about.fb.com/news/2018/08/removing-myanmar-officials/> accessed 5 December 2022

[74] Stecklow (n 61)

[75] Bradshaw and others (n 65) 274.

[76] Samuel Woolley and Philip Howard, 'Conclusion: Political Parties, Politicians, and Computational Propaganda' in Samuel Woolley and Philip Howard (eds), *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (OUP 2019) 242.

[77] UNHRC (n 6) [1325]-[1354]

CC BY

for responding slowly, ineffectively, and being uncooperative with international bodies.[78] Contributing factors to Facebook's failure were the lack of human reviewers who understood the cultural context of Myanmar and the technological incompatibility of Burmese language fonts with the platform's system.[79]

After the problem was exposed, Facebook adopted measures to try to overcome the challenges, such as hiring country specialists, using artificial intelligence to flag suspicious content, and acting on misinformation.[80] As a result, it removed hatred and misleading content, and banned people and organisations that promoted them, preserving data for further investigations.[81]

Nevertheless, it was not enough. Hate speech and fake news continued to spread in the country's social media environment.[82] Only after a coup d'état happened in 2021, the platform decided to comprehensively ban the military, including media and commercial entities linked to them.[83] Still, critics questioned the measure's effectiveness since the military would have other Facebook profiles to use covertly.[84]

Additionally, international bodies considered the measures adopted by the Myanmar government inadequate.[85] Instead of promoting respect for human rights and holding accountable public officials who engaged in hateful rhetoric, the government was accused of taking advantage of arguments advocating that it should curb hate speech and fake news, using them to silence human rights defenders and journalists who used the Internet to denounce abuses and criticise authorities.[86] Hatred was fostered; human dignity was stifled.[87] The values protected by free speech were inverted.

Concluding, the Myanmar case shows that cybercommunications added complexity to an already acute problem. Propaganda against a vulnerable group was computationally enhanced, gaining the characteristics of computational

[78] ibid

[79] ibid

[80] UNGA 'Report of the Special Rapporteur on the situation of human rights in Myanmar, Thomas H. Andrews' (4 March 2021) UN Doc A/HRC/46/56, Annex I [16]

[81] ibid

[82] ibid

[83] Facebook, 'An Update on the Situation in Myanmar' (11 February 2021) <https://about.fb.com/news/2021/02/an-update-on-myanmar> accessed 5 December 2022

[84] Billy Perrigo, 'Facebook's Ban of Myanmar's Military Will Be a Test of the True Power of Social Media Platforms' *Time* (1 March 2021) <https://time.com/5943151/facebook-myanmar-military-ban/> accessed 5 December 2022

[85] UNGA 'Report of the independent international fact-finding mission on Myanmar' (12 September 2018) UN Doc A/HRC/39/64 [73]. UNHRC (n 6) [1327]

[86] UNHRC (n 6) [1355]-[1359]

[87] ibid

CC BY

propaganda, such as scale and anonymity.[88] While the author of some Burmese posts can be promptly identified, who is behind the online orchestrated attacks? How to deal with something so quick and scalable on the one hand, and underhanded and fragmented on the other hand?

4. APPLYING THE INTERNATIONAL LEGAL FRAMEWORK OF INCITEMENT TO GENOCIDE TO THE INTENTIONAL DISTRIBUTION OF FAKE NEWS USING SOCIAL MEDIA

The Myanmar case allows us to question whether the existing legal framework stands the test of time.

To answer it, we will work with the Commander-in-Chief's post stating that 'every citizen has the duty to safeguard race, religion, cultural identities and national interest' and that 'the national defence duty falls on every citizen'.[89] His official unrestricted profile had 2.9 million followers,[90] therefore able to 'reach a large or indeterminate audience', constituting public incitement. The audience were Myanmar citizens, a conclusion drawn from the use of Burmese fonts, language, and the message itself.[91] It was posted in 2017,[92] contemporarily to the moment when the Rohingya genocide was a possibility since the 'clearance operations' had already started.[93]

What about the content? The idea of a legal duty is not to permit an action (which would constitute a right) but to require action.[94] It gives the idea of compelling people to act. Those who violate their duties are themselves against the law. He had already posted other messages calling the Rohingya 'Bengali'.[95] In a previously built context that denies the existence of a group as such and falsely treats this group as illegal immigrants, the text intends to persuade readers to the logical conclusion that citizens are obliged to act against Rohingya invaders.[96] However, the conclusion is invalid because it is built on false premises, as demonstrated. In Arendt's lessons, it is the deconstruction of reality coupled with the persuasive logic of propaganda being used to guide action.[97] Hence, besides public, this social media post displays all characteristics needed to constitute direct incitement, since it aims at triggering a course of action by its audience, and it spreads a false conclusion.

[88] Woolley and Howard, 'Introduction: Computational Propaganda Worldwide' (n 70) 7.

[89] UNHRC (n 6) [1341]

[90] ibid [1329]

[91] ibid [1341]-[1352]

[92] ibid

[93] ibid [537]-[573]

[94] Leif Wenar, 'Rights', *The Stanford Encyclopedia of Philosophy* (Spring edn, 2021) <https://plato.stanford.edu/entries/rights> accessed 5 December 2022

[95] UNHRC (n 6) [1336]-[1341]

[96] ibid

[97] Arendt (n 7) 471-472.

CC BY

The fake narrative also implies the *mens rea*. As described, the ICTR suggested that falsehood indicated intent.[98] Therefore, by disguising premises to make a false conclusion appear to be the result of logical reasoning, the agent's manipulative intent shows itself. Lastly, the specific genocidal intent derives from the fact that the message stimulates the destruction of an ethnic group as such.

Now we can return to the question posed at the beginning of this paper, as to whether the intentional distribution of fake news using social media platforms can be criminalised as incitement to genocide under international law, and answer it: yes, as demonstrated, it can. The process was also demonstrated, answering the second research question.

Nevertheless, this was only one example and not every social media post spreading fake news will reach the threshold of incitement, just as not all examples from the past reached.

5. SHOULD THE DISTRIBUTION OF ONLINE FAKE NEWS THAT DOES NOT CONSTITUTE INCITEMENT BE CRIMINALISED UNDER CERTAIN CIRCUMSTANCES?

What about online speeches falling outside the scope of the norm? To what extent, if at all, should they be criminalised when carrying falsehood?

The precedents reviewed throughout this paper show how fake news has been an important element present in cases of incitement to genocide since judgments expressly discussed evidence of lies, disinformation and misrepresentation of facts in defendants' publications and broadcasts. However, it is an element that triggers a contradictory response from international case law. If, on the one hand, international Courts recognise the role of systematic campaigns to deconstruct reality as a determining factor in creating the climate conducive to genocide, on the other hand, they are reluctant to recognise it as enough for convictions, seeking specific phrases in the discourse promoted by defendants.

In other words, international Courts admit that fake news campaigns add to the social context needed for genocide to become a real possibility but do not hold criminally responsible those who intentionally contribute only to this aspect, seeking messages calling to action, even if implicitly. When international Tribunals could not find those calls, it led to acquittal, as Fritzsche's Nuremberg judgment showed. However, when the German Denazification Trial sentenced him, '[t]he court made it clear that Fritzsche had been convicted for anti-Semitic propaganda per se, without additional calls for acts of violence'.[99] Therefore, Courts diverged.

Thus, what do Streicher's misrepresentation of Jews accusing them of drinking children's blood,[100] Rwandese broadcasts accusing Tutsis of being 'unjustifiably wealthy',[101] and Myanmar faked photos showing the Rohingya burning their own

[98] *Media Case* (n 30) [1021]

[99] Benesch (n 19) 511.

[100] Eastwood (n 3) vol I, 156.

[101] *Media Case* (n 30) [365]

CC BY

villages,[102] all have in common? They use false narratives to progressively build up hate and prejudice against a minority; to 'inflame ethnic tensions' in the words of the ICTR.[103] Although such discourses cannot constitute direct incitement under international case law, what is their purpose if not to achieve the same result but through multiple actions? The ICTR recognised that hate propaganda loaded the gun in Rwanda.[104] Hence, to trigger a course of action is the outcome desired by those who engage in the aforementioned fake news campaigns, but it is done step-by-step.

Criminal law requires limitation, but seeking specific calls to action (explicitly or implicitly) seems to be too strict of a limitation because the current international legal framework is unable to catch some types of extremely harmful speech, as shown. Thus, the law falls short of effectively protecting human rights and preventing genocide.

Besides, the Myanmar case shows that the problem escalates when computational resources are used to create an orchestrated campaign.

Therefore, international law must go beyond the current framework to effectively prevent genocide. How? One possibility is to criminalise the conduct of systematically creating or distributing online fake news when such conduct constitutes computational propaganda with the intent to harm groups protected under the Genocide Convention. However, how is this different from hate propaganda? Is this solution democratically justifiable? These questions will be discussed in the next section.

A. PROPOSING A NEW CRIME

The debate on criminalising hate propaganda occurred during the drafting of the Convention.[105] The Secretary-General's Draft suggested that '[a]ll forms of public propaganda tending by their systematic and hateful character to provoke genocide, or tending to make it appear as a necessary, legitimate or excusable act shall be punished'.[106] It explained that this crime focused on speech falling outside the scope of direct and public incitement, acknowledging that genocidal 'propaganda is even more dangerous than direct incitement to commit genocide' because it induces ordinary citizens to believe that the existence of the victim group

[102] UNHRC (n 6) [1270]-[1340]

[103] *Media Case* (n 30) [1021]

[104] ibid [953]

[105] Matthew Lippman, 'The Drafting of the 1948 Convention on the Prevention and Punishment of the Crime of Genocide' (1985) 3 BostonUIntlLJ 1 31-48. Wibke Timmermann, 'The Relationship between Hate Propaganda and Incitement to Genocide: A New Trend in International Law Towards Criminalization of Hate Propaganda?' (2005) 18 LJIL 257 279. Benesch (n 19) 508.

[106] Draft Convention on the Crime of Genocide (26 June 1947) UN Doc E/477 art III.

CC BY

is 'a mortal danger for the nation or for society'.[107] However, criminalisation was rejected for fear that it would limit free speech.[108]

Regarding jurists, Timmermann proposes 'that hate propagandists should be prosecuted for direct and public incitement to genocide if their hate speech is engaged in with the specific intent to commit genocide and creates a substantial danger of genocide'.[109] However, endorsing Shaw's view (1989, cited in Schabas, 2009), Schabas excludes the possibility of interpreting genocidal propaganda as incitement.[110]

Thus, how can the focus on cybercommunications coupled with fake content, rather than hate content, provide a new perspective on the phenomenon?

Such a perspective allows us to address those novel ingredients that aggravate the problem by proposing a new crime encompassing the concept of computational propaganda in its definition. By doing so, the phenomenon is better described because both the content of messages and the process used to disseminate them are equally relevant. Thus, the accuracy of the *actus reus* is improved.

The concept of computational propaganda encompasses the idea of enhancing traditional propaganda with computational tools.[111] It 'describes the use of algorithms, automation and human curation to purposefully manage and distribute misleading information over social media networks' aiming at 'the manipulation of public opinion'.[112] It 'typically involves one or more of the following ingredients: bots that automate content delivery; fake social media accounts that require some (limited) human curation; and junk news - that is, misinformation about politics and public life'.[113] All these elements have been identified in the Myanmar situation.

Therefore, the proposal here is to criminalise the conduct of creating or distributing computational propaganda with the specific intent to harm a group protected under the Genocide Convention ('national, ethnical, racial or religious'[114] minorities[115]). This definition encompasses the idea of fake narratives and demands assessing the harmful intent also according to the computational tools chosen to broadcast messages instead of just examining linguistic structures.

---

[107] ibid 32-34

[108] Lippman (n 104) 31-47. Schabas (n 24) 324. Timmermann (n 104) 279.

[109] Timmermann (n 104) 257.

[110] Schabas (n 24) 334.

[111] Woolley and Howard, 'Introduction: Computational Propaganda Worldwide' (n 70) 7.

[112] ibid 4-5.

[113] ibid

[114] Genocide Convention (n 1) art II.

[115] On the discussion about defining protected groups, see: William Schabas, 'Groups protected by the Genocide Convention: conflicting interpretations from the International Criminal Tribunal for Rwanda' (2000) 6 ILSAJIntl&CompL 375

CC BY

B. DEMOCRATIC JUSTIFICATION

Would such criminalisation be democratically justifiable?

According to international law, restrictions on freedom of expression must be necessary to achieve specific purposes, such as the rights of others and public order.[116]

What other alternatives could there be to stop such harmful online speech?

One could think of platform regulation, but the Myanmar situation shows that it does not solve the problem. Burmese inciters were successful in using a variety of tactics to circumvent technical obstacles. For example, when posts were excluded from Facebook, they kept the texts online on other websites.[117] Since '[i]n Myanmar, most users share posts by copying and pasting the content, rather than by using the share function', their circulation could still be renewed.[118] Regarding human moderation, there have been reports of biased approaches and moderators unable to understand the metaphorical language commonly used.[119] Lastly, when banned from Facebook, inciters migrated to other social media platforms.[120]

Hence, platform regulation was not enough to stop the use of computational propaganda against the Rohingya and protect them from harm. However, this does not mean that regulation is irrelevant. Regulation can make it harder for hatred and falsehood to spread. Still, it is necessary to go further to adequately protect vulnerable groups. This is the reason why criminalisation is necessary.

Regarding the risks of criminalising computational propaganda for free speech, one could argue that broad criminalisation of speech causes chilling effects, legally understood in this context as the inhibition of expressing legitimate speech for fear that it could result in sanctions.[121]

However, the European Commission acknowledged that media campaigns (although against judges and prosecutors) and online attacks (although against journalists) can also cause chilling effects.[122] Therefore, conflicting rights would at least pose the same risk.

But it goes beyond that. The problem becomes more serious when a vulnerable group is the one attacked. Fiss shows how hate speech silences 'disadvantaged groups', nullifying their speech and compromising the liberal idea that the remedy for free speech abuses would always be more speech.[123] Consequently, if powerful

[116] International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 art 19(3).

[117] UNHRC (n 6) [1328]

[118] Progressive Voice and others (n 61) 63-64.

[119] ibid 64-65.

[120] ibid 65-66.

[121] *Eon v France* App no 26118/10 (ECtHR, 14 March 2013) [61]

[122] Commission, '2020 Rule of Law Report: The rule of law situation in the European Union' (Communication) COM(2020) 580 final 11, 20.

[123] Owen Fiss, *The Irony of Free Speech* (Harvard University Press 1996) 15-19.

CC BY

actors exclude marginalised groups from the public debate, States could reduce influential voices to achieve a balance.[124]

The process used by computational propaganda tries to 'manufacture consensus' and silence dissent.[125] The Myanmar case shows that it can generate the same effect that Fiss observed regarding hate speech.

From a political philosophy standpoint, this is a problem because suppressing the voice of a group affects more than individual rights to freedom of expression; it affects their collective capacity to fight for their rights. When a group is silenced, it is depoliticised and loses the ability to claim its rights.[126] The loss of political participation, which leads to the loss of rights and dehumanisation, is part of the genocidal process,[127] as shown.

When the crime here proposed restricts victim groups to the most vulnerable ones, it represents an effort to mitigate the risk of overcriminalising behaviour on the one hand and, on the other hand, focuses on protecting those for whom more speech would not be a solution. Therefore, it is democratically justifiable because it aims to restore equality in public debate and protect the capacity of vulnerable groups to defend their rights.

We can now answer the third research question: the intentional distribution of fake news online that does not reach the threshold of incitement but constitutes computational propaganda with the intent to harm a group protected under the Genocide Convention should be criminalised because it is a necessary and proportionate legal measure to effectively prevent genocide, as demonstrated.

6. CONCLUSION

This paper demonstrated how the existing legal framework of the international crime of incitement to genocide may be applied to cybercommunications under certain circumstances.

Notwithstanding, it found a problem: because the current framework is limited by the concept of direct incitement, it historically could not (and still cannot) reach some types of extremely harmful speech, such as hate propaganda and fake news campaigns without additional calls to criminal action.

When presenting the Myanmar case, the paper concluded that adding cybercommunications to the phenomenon increased the complexity of the problem.

To adequately protect human rights, it proposed to criminalise the conduct of creating or distributing computational propaganda with the specific intent to harm a group protected under the Genocide Convention.

The concept of computational propaganda was applied because it better represents the phenomenon by encompassing the content of fake messages and the

[124] ibid

[125] Woolley and Howard, 'Introduction: Computational Propaganda Worldwide' (n 70) 4.

[126] Jacques Rancière, *Dissensus: On Politics and Aesthetics* (Continuum 2010) 37-39.

[127] *The Gambia v Myanmar* (n 5) [55]

CC BY

dissemination process. Hence, it is a new perspective that adds precision and limits criminalisation.

Lastly, it found that such criminalisation is democratically justifiable because it is a necessary and proportionate measure to prevent genocide.

REFERENCES

Agreement for the prosecution and punishment of the major war criminals of the European Axis (adopted 8 August 1945) 82 UNTS 279.

Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v Myanmar) (Request for the Indication of Provisional Measures: Order) General List No 178 [2020] ICJ.

Arendt H, *Origins of Totalitarianism* (2nd edn, Meridian Book 1958).

Benesch S, 'Vile Crime or Inalienable Right: Defining Incitement to Genocide' (2008) 48 VaJIntlL 485.

Bradshaw S and others, 'Country Case Studies Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation' (2021) Oxford Computational Propaganda Research Project <https://medleysites.oii.ox.ac.uk/demtech/wp-content/uploads/sites/127/2021/03/Case-Studies_FINAL.pdf> accessed 5 December 2022.

Citarella J, 'There's a New Tactic for Exposing You to Radical Content Online: The "Slow Red-Pill"' *The Guardian* (15 July 2021) <www.theguardian.com/commentisfree/2021/jul/15/theres-a-new-tactic-for-exposing-you-to-radical-content-online-the-slow-red-pill> accessed 5 December 2022.

Convention on the Prevention and Punishment of the Crime of Genocide (adopted 9 December 1948, entered into force 12 January 1951) 78 UNTS 277.

Draft Convention on the Crime of Genocide (26 June 1947) UN Doc E/477.

Eastwood M, 'The Emergence of Incitement to Genocide Within the Nuremberg Trial Process: The Case of Julius Streicher' (PhD thesis, University of Central Lancashire 2006) https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.438233 accessed 5 December 2022.

CC BY

Facebook, 'An Update on the Situation in Myanmar' (11 February 2021) https://about.fb.com/news/2021/02/an-update-on-myanmar accessed 5 December 2022.

Facebook, 'Removing Myanmar Military Officials From Facebook' (28 August 2018) https://about.fb.com/news/2018/08/removing-myanmar-officials/ accessed 5 December 2022.

Fiss O, *The Irony of Free Speech* (Harvard University Press 1996).

International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171.

Lippman M, 'The Drafting of the 1948 Convention on the Prevention and Punishment of the Crime of Genocide' (1985) 3 BostonUIntlLJ 1.

Media Case (Appeals Judgment) ICTR-99-52-A, A Ch (28 November 2007).

Media Case (Judgment) ICTR-99-52-T, T Ch I (3 December 2003).

Perrigo B, 'Facebook's Ban of Myanmar's Military Will Be a Test of the True Power of Social Media Platforms' *Time* (1 March 2021) https://time.com/5943151/facebook-myanmar-military-ban/ accessed 5 December 2022.

Progressive Voice and others, 'Hate Speech Ignited: Understanding Hate Speech in Myanmar' (8 October 2020) Joint Report <https://hrp.law.harvard.edu/wp-content/uploads/2020/10/20201007-PV-Hate-Speech-Book-V-1.4-Web-ready1.pdf> accessed 5 December 2022.

Prosecutor v Akayesu (Judgment) ICTR-96-4-T, T Ch I (2 September 1998).

Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3.

Schabas W, *Genocide in International Law: The Crime of Crimes* (2nd edn, CUP 2009).

Stecklow S, 'Why Facebook Is Losing the War on Hate Speech in Myanmar' *Reuters* (15 August 2018) <www.reuters.com/investigates/special-report/myanmar-facebook-hate> accessed 5 December 2022.

CC BY

Timmermann W, 'The Relationship Between Hate Propaganda and Incitement to Genocide: A New Trend in International Law Towards Criminalization of Hate Propaganda?' (2005) 18 LJIL 257.

UNGA, 'Report of the Independent International Fact-Finding Mission on Myanmar' (12 September 2018) UN Doc A/HRC/39/64.

UNHRC, 'Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar' (17 September 2018) UN Doc A/HRC/39/CRP.2.

UNSC Res 955 (8 November 1994) UN Doc S/RES/955.

Wilson R, 'Inciting Genocide with Words' (2015) 36 MichJIntlL 277.

Woolley S and Howard P, 'Introduction: Computational Propaganda Worldwide' in Woolley S and Howard P (eds), *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (OUP 2019).

Woolley S and Howard P, 'Conclusion: Political Parties, Politicians, and Computational Propaganda' in Woolley S and Howard P (eds), *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (OUP 2019).

CC BY