Countering Hate Speech on Social Media in Cameroon: Legal and Technical Measures

Pierrette Essama-Mekongo

Faculty of Laws and Political Science, University of Yaounde II, Soa, Cameroon
Email: essamamekongo@gmail.com

Abstract

Hate speech on social media is on the rise globally. It is to be accounted for several offline violence including the massacres of the Rohyinga in Myanmar, mass shootings and hate crimes in the United States of America. Online hate speech benefits from increased access to the internet at reduced costs and from the anonymity offered in cyberspace. Cameroon is no exception to this and online hate speech is rampant. This has led to the incorporation of the offence of hate speech in the Penal Code in 2019 in the aftermath of the 2018 presidential elections where, for the first time, social media played a major role in an electoral campaign amidst civil upsurge in the South West and North West regions with a strong involvement of the diaspora. This paper examines legal and technological measures put in place to address online hate speech in Cameroon. The qualitative methods of exegesis and content analysis were used in this study. It follows that, while Cameroon took significant legal steps to combat online hate speech by criminalising it, there is still some confusion between the various provisions. Technological measures are also used to fight hate speech on social media through detection and suppression of contents and accounts. The law needs amendments for harmonisation and better coverage of hate speech.

Keywords

Hate Speech, Social Media, Legal and Technical Measures, Cameroon

1. Introduction

Freedom of speech is a human right guaranteed by all democracies in the world including Cameroon which enshrined it in the Preamble of its Constitution (Law N˚96/06 of 18 January 1996 to Amend the 1972 Constitution) in keeping with its international undertakings upon ratification of all relevant international human rights instruments. In this line, the 16th indent provides thus: "the freedom of

communication, of expression […] shall be guaranteed under conditions fixed by the law". Freedom of speech has found an unparalleled platform of expression with the advent of the Internet and, particularly, of social media. In effect, before the advent of social media, public speech was the privilege of a few, among whom: some politicians, performers, and other high-profile individuals as it was limited by traditional media outlets (Sinani, 2015). These traditional media outlets have a finite character as they are under space and time constraints. In contrast, the internet offers unlimited space to free speech for anyone who wants to express themselves. In this regard, Ronson (2015) held that "the great thing about social media was how it gave a voice to voiceless people". Beyond expressing themselves through blogs, emails and ordinary websites, social media networks have expanded this power further again by allowing short messages to be posted in a variety of platforms and discussion groups instantly. This is heightened by features that allow anonymity and provide immediate translation assistance thus suppressing language barriers.

Social media can be defined as computer-mediated technologies facilitating the creation and sharing of information, ideas, career interests and other expressions via virtual communities and networks (Mano, 2023). It provides a virtual place to openly express feelings, opinions and beliefs (Ruwandika, 2017). The most popular social media platforms include Facebook, Twitter, WhatsApp, Instagram, Snapchat, TikTok and Whisper. These social media platforms are used to create communities of joint interests, foster public engagement and nurture relationships and networks (Mano, 2023). From Wen's (2020) account, social media is widely used for acculturation of new settlers in communities abroad and also serves the other way round for communities that welcome foreigners. However, alongside these positive initiatives and outcomes which serve at uniting people in a more intertwined and bonded society, freedom of speech has also laid ground for negativity on social media in the form of hate speech for example.

Hate speech is defined as "any offense motivated in whole or in part, by the offender's bias against an aspect of a group of people" (Silva, Mondal, Correa et al., 2016: p. 688). It is deemed to consist of "speech that vilifies, harasses, intimidates or incites hatred toward an individual or group on the basis of a characteristic such as race, ethnicity, religion, or gender" (Brison, 2021: p. 1). Speech includes any verbal or written utterances but also any pictorial representations and symbols such as Swatstikas and Ku Klux Klan masks or even conduct such as cross-burning and the defacing of mosques (Brison, 2021: p. 1). Offline violence including the massacres of the Rohyinga in Myanmar, mass shootings and hate crimes in the United States of America, as well as violence and hierarchies of hate in India escalated due to online hate speech (Banaji and Bhat, 2021). Hate speech is on the rise on a global scale following the rise in the use of internet and social media which allow communication at "near zero marginal cost to users" (Silva, Mondal, Correa et al., 2016: p. 687) in a context where costs of internet connection have reduced worldwide and access to internet has grown exponen-

tially. In Cameroon, it is possible to navigate on the internet for one hour with 100 CFA Francs (about 0.16 USD) and even to send or read messages free of charge on Facebook (Ambe, 2021: p. 53).

Online hate speech has also been on the rise in Cameroon in the past few years and has attracted attention (Ngah, 2023; Mouangue, 2023) as evidenced by the passing of a law in 2019 to amend the Penal Code by incorporating a new offence relating to hate speech (Law 2019/020 of 24 December 2019 to Amend and Supplement some Provisions of Law 2016/7 of 12 July 2016 relating to the Penal Code). This law was adopted in the aftermath of the 2018 presidential elections that played both online and offline. Online hate speech is by itself an ill as it causes distress among victims, most of whom are found in vulnerable groups. It is also very dangerous as it triggers offline violence ranging from random attacks to organised riots and terrorist acts. This paper examines how hate speech on social media is being countered in Cameroon. The peculiarity of social media as a technological progress points to the fact that criminalising hate speech may not be enough to end it. It is also necessary to use technological means to ensure that hate speech is completely suppressed. Such technological measures themselves need to be regulated in order to conform with the law. This paper will proceed in three stages: first, it will examine the phenomenon of hate speech on social media, then, it shall look at the existing legislative measures to tackle it. Lastly, it shall determine the various technological measures that necessarily accompany the legal dispensation.

2. The Phenomenon of Hate Speech on Social Media

Social media platforms act both as catalyst and medium of expression for hate speech (Van Bavel, Robertson, Del Rosario, Rasmussen, & Rathje, 2024: p. 313). They heighten the phenomenon while multiplying types and targets. Their significant role stems from increased use.

2.1. Use of Social Media for Hate Speech in Cameroon

Hate speech can only prosper online as internet and social media have grown to daily use over the years. What could remain a marginal phenomenon has gained impetus with the liberalisation of access to internet and social platforms. As of January 2023, Cameroon was accounted to have 12.89 million internet users translating to an internet penetration rate of 45.6 percent with 3.90 million social media users, equating 13.8 percent of the total population of the country which is estimated at 28.28 million. Facebook leads social media use in Cameroon with 12.7 percent of the population while Instagram users total only 1.7 percent and X (former Twitter) is used by a bare 0.7 percent (Kemp, 2023).

Facebook, YouTube, WhatsApp and Twitter are the main platforms used by activists (Agwanda et al., 2020), that is, by people who want to voice their opinions. This makes them the arena of contradictory views and the cradle for hate speech. Yet, many activists are reported to have issues using X because of the

limited number of characters even if those 140 characters have been held to change the way war is conducted on the battlefields (Patrikarakos, 2017). In effect, despite its worldwide importance, as noted earlier, X is used by only 0.7% of the social media users in Cameroon (Kemp, 2023).

Online hate speech has been found to diffuse farther, wider and faster and to have a greater outreach than non-hateful content. In that, it is more influential, popular and cohesive: "hate speech has a much higher spreading velocity" (Matthew, Dutt, Goyal, & Muhkherjee, 2018: p. 1). Hate speech is amplified by social media and populist politics (U.N. Human Rights Council, 2019: pp. 71-73; Khan, 2021: pp. 12-23) and online hate speech has been found to be linked to offline violence (Wilson, 2021: pp. 1039-1045) and a threat to peace and security (CHDA, 2020). In fact, hate speech is a threat to many other human rights and freedoms ensuring peaceful living, freedom of religion, beliefs and political opinions, and to the rights of vulnerable and minority groups.

2.2. Types of Online Hate Speech

Online hate speech is conceptualized as the use of violent, aggressive or offensive language, focused on a specific group of people who share a common property in the form of religion, race or ethnic group, gender or political affiliation or any beliefs through the use of Internet and Social Networks carried out repeatedly, systematically and uncontrollably, through digital media and often motivated by ideologies (Castaño-Pulgarín, Suarez-Betancourt et al., 2021: p. 1; Watanabe, Bouazizi, & Ohtsuki, 2018). The pre-eminence of some types of hate speech depends on the sociological and geographic context.

Therefore, in the Western world for example, types of online hate speech include religious, racism, gendered and political hate speech. These types of hate speech are generally endorsed by ideologies that underlie them such as racism, islamophobia, Alt-Right and white nationalism (Castaño-Pulgarín, Suarez-Betancourt et al., 2021: p. 2). Online religious hate speech has been defined as the use of inflammatory and sectarian language to promote hatred and violence against people on the basis of religious affiliation through the cyberspace (Albadi, Kurdi, & Mishra, 2018). While religious hate speech is mainly against Islam in the Western world (Horsti, 2016; Evolvi, 2017), in the context of Boko Haram sect terrorism raging in Cameroon and Nigeria, religious hate speech would be against Christians.

Online racism is also rampant in Western countries (Ben-David & Matamoros-Fernandez, 2016; Cohen, Holt, Chermak, & Freilich, 2018) while tribalism is its corresponding form in Cameroon. Gendered hate speech which is where people are attacked based on their gender as a form of gender-based violence (Khosravinik & Esposito, 2018) may also be found in Cameroon. Political online hate speech, on its part, expresses animosity and intolerance against people with differing political views and choices. In most cases, however, political hate speech is intertwined with racist, ethnic and religious hate speech (Pulido, 2019).

Political hate speech may be held to be the paramount form of hate speech in Cameroon as it covers online battles around the "Anglophone crisis" and the 2018 presidential elections. These battles have made the political fights turn into ethnic rivalry and detestation with the first between Anglophones and Francophones and the second between Bamileke and Beti ethnic groups (Pant, 2020; Barrach-Yousef, 2020; Kindzeka, 2023).

On the other hand, the website of the National Commission for the Promotion of Bilingualism and Multiculturalism provides a space for people to denounce hate speech. To lay their complaint, people are asked to choose among 5 types of contents that represent the types of hate speech identified by the Commission, that is: 1) incitement to racial hatred or incitement to discrimination of persons because of their origin, sex, sexual orientation or disability; 2) threats or incitement to violence; 3) endangering people; 4) insult and 5) terrorism: threat or apology (https://www.cnpbm.cm/en/denunce-hate-speech).

2.3. Targets of Online Hate Speech

Targets of online hate speech on Twitter and Whisper social media platforms in the United States of America have been categorised by Silva, Mondal, Correa et al. (2016: p. 689) in **Table 1**. Thus:

Table 1. Hate categories and example of hate targets from Silva, Mondal, Correa et al. (2016).

Categories	Example of hate targets
Race	nigga, black people, white people
Behavior	insecure people, sensitive people
Physical	obese people, beautiful people
Class	ghetto people, rich people
Gender	pregnant people, cunt, sexist people
Ethnicity	Chinese people, Indian people, paki
Disability	retard, bipolar people
Religion	religious people, jewish people
Other	drunk people, shallow people

In Cameroon, categories of hate speech, hate targets and hate terms used may be summarised in **Table 2**. Thus:

Table 2. Cameroon's Hate categories, targets and terms from author (2024).

Categories	Examples of hate targets	Terms used
Political	ruling party members, opposition party members (more specifically CRM)	Talibans, tontinards, sardinards La Republic, CPDM Monsters
Social	French speaking Cameroonians, English speaking Cameroonians	Anglofous, Anglofools ambazos, frogs, La Republic, come-no-go, Francofools

Continued

Ethnicity	Beti and Bamileke ethnic groups	Ngrafi, Graffi, Nkoa
Class	Ghetto people	Populace
Sports/football	Foes and Fans of Eto'o Fils	Hibous Football Club, fidèles de la Sainte Eglise de Tsinga, églisiens

In Cameroon, tribalism, as intertwined with political as well as social hate speech, takes precedence over other forms of hate speech and concerns mainly two tribes: Bamileke from the grassfields in the West region and Beti from the Centre and South regions of the country. The rivalry has been raging over the years and heightened with the 2018 presidential elections where one of the main candidates purportedly standing chances to win the elections is from the Bamileke ethnic group while the sitting president who won the elections is from the Beti group (Pant, 2020; Barrach-Yousef, 2020; Kindzeka, 2023; Kebuh and Tizie, 2022).

Then, the Anglophone crisis that started in 2016 with protests from lawyers and teachers has evolved in a war between separatists and law enforcement agents with many casualties among the belligerents, the State troops and the population in the affected North West and South West regions and the neighbouring West Region which is in between the former two. The war on social media opposes Anglophone activists and francophones who support the government against the insurgency and secessionist propensities (Kindzeka, 2023; Sombaye Eyango, 2018).

It appears from the foregoing that factors causing hate speech in Cameroon include: the linguistic difference between English speakers and French speakers, that is, anglophones and francophones, political rivalry, support to various leaders or celebrities or to their opponents or competitors, as well as ethnic and tribal competition. The linguistic difference stems from the fact that Cameroon was separated into two territories ruled by England and France before it was reunited as a federal State. Even after federalism was forsaken to form a unitary State, some cultural, linguistic, legal and administrative differences still exist (Essama-Mekongo, 2024) and cause frictions leading to hate speech from either side when met with intolerance from the other group. This is the cause of the "anglophone crisis" that started in 2016 (Bang and Balgah, 2022) and is still going on in 2024. Hate speech stemming from linguistic differences is often employed for political ends (Pant, 2020). Political rivalry is the second factor causing hate speech, whether between separatists or leaders including their supporters or between political parties running in elections, mostly between the ruling party and the opposition parties memberships (Mushing Tamfuh, 2020). Ethnic differences and tribalism between the various ethnic groups also cause hate speech when these groups have conflicts. The most prominent is the one which "opposes" grassfield natives and people from South, Centre and East regions: Bamilekes against Betis

(Ntui, 2021). Others include Northern populations such as Mbororos and Grassfields populations (Sunjo and Ndzi, 2022). Most of the times, all these factors are entangled as rightly put by Mushing Tamfuh (2020). They result from the various pluralities that make up the Cameroonian society that has not yet succeeded into achieving a peaceful unitary State.

Although fighting hate speech necessarily involves curtailing the right to freedom of expression (Mirchandani, 2018; Brown, 2015), violence involved in hate speech and damage it causes warrant that legislative action be taken to combat it. In effect, hate speech causes ravages in society and breaches social cohesion causing social instability. Hate speech ignites poison and conflicts between communities called upon to live together. Cameroon for example, is a home to more than 250 ethnic groups (Sunjo and Ndzi, 2022). Hate speech may lead to random or concerted acts of violence, defiance among groups and fights between communities as well as social unrest, riots and even genocides as in the case of Rwanda (Viljoen, 2005) or Myanmar (Morada, 2023), as mentioned earlier. If a country is therefore to prevent such dreadful occurrences and keep social cohesion and stability, it needs to fight hate speech through diverse measures including education and legislative activities such as criminalising hate speech.

3. Criminalisation of Hate Speech in Cameroon

Criminal law has always been the best way to ensure that unacceptable behaviour is prohibited in any given country. Hate speech has been criminalised in Cameroon in 2019 via a legislative enactment which adds to existing offences relating to it.

3.1. Criminalisation of Hate Speech as an International Undertaking

The Cameroonian legislative action against hate speech is predicated upon the State's international undertakings in the human rights field. These undertakings are contained in the body of relevant international human rights instruments endorsed in the Preamble of the Constitution which states that the People of Cameroon affirms its attachment to the fundamental freedoms enshrined in the Universal Declaration of Human Rights (1948), the Charter of the United Nations (1945) and all duly ratified International Conventions relating thereto. Then, among the human rights principles clearly stated, there is Principle 13 which provides that: "no person shall be harassed on grounds of his origin, religious, philosophical or political opinions or beliefs, subject to respect for public policy".

For the fight against hate speech, one of such international instruments is the International Covenant on Civil and Political Rights (1966)—as ratified by Cameroon in 1984—which provides in its article 20(2) that "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law". So that even if freedom of expression is provided for in article 19, it should not be used to advocate hatred

that incites and leads to violence such as hate speech does. Then, there is the International Convention on the Elimination of All Forms of Racial Discrimination (1965)—as ratified by Cameroon in 1971—which provides in article 2 that States parties must condemn racial discrimination and take all measures to eliminate it in all its forms. Such measures would include legislating to end racial discriminations by any persons, group or organisation. Article 4 further provides that States parties should make an offence of all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination; all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof. And, that protection and remedies should be made available to everyone against such acts in article 6. It is to be noted that the definition of racial discrimination includes ethnic discrimination as per article 1 of the Convention.

3.2. The Offence of Hate Speech

There are three offences in Cameroon criminal law that relate to hate speech. They are contained in the Penal Code (Law N 2016/7 of 12 July 2016 Relating to the Penal Code as Amended by Law N 2019/020 of 24 December 2019 to Amend and Supplement Some Provisions of Law N 2016/7 of 12 July 2016 Relating to the Penal Code) and in the Law on Cybersecurity and Cybercriminality (Law N˚2010/12 of 21 December 2010 on Cybersecurity and Cyber Criminality). The term "hate speech" is utilised only once though. Yet, the other two offences do satisfy the definition of hate speech. The amount of fines is given in the Cameroonian currency, the CFA Franc and, 1 USD is about 611.70 CFA Francs.

As for the Penal Code, section 241 provides that:

1) Whoever commits a contempt, within the meaning of section 152 of this Code, of the race or religion of one or many citizens or residents shall be punished with imprisonment for from 6 (six) months and with fine of from 5000 (five thousand) francs to 500,000 (five hundred thousand) Francs:

2) Where the offence is committed by means of the press or wireless, radio, television or *social media* or any other means likely to reach the public, the maximum of the fine provided for in subsection (1) above may be extended to 20,000,000 (twenty million) Francs.

3) Where the offence is committed with intent to arouse hatred or contempt between citizens, or residents, the penalties provided for by the foregoing subsections shall be doubled.

While section 241-1 (1) titled contempt of tribe or ethnic group introduced by the same 2019 Law reads thus:

Whoever, by any means, makes *hate speech* against people or incites them to violence due to their tribal or ethnic origin shall be punished with imprisonment of from 1 (one) to 2 (two) years and with fine of 300,000 (three hundred thousand) francs to 3,000,000 (three million) francs.

On its part, section 77 (1) of Law N°2010/12 of 21 December 2010 on Cyber-security and Cybercriminality sanctions only contempt of race and religion when electronic communications or information systems are used as a medium presumably because, at that time, the Penal Code had not yet criminalised contempt of tribe or ethnic group. The sanctions consist of imprisonment for from 2 (two) years to 5 (five) years or a fine of from 2,000,000 (two million) to 5,000,000 (five million) CFA Francs or with both such fine and imprisonment. This is clearly more severe than what is provided for under sections 241 and 241-1 of the Penal Code. It surely takes into consideration the amplitude and velocity of electronic and telecommunications networks as discussed earlier. Yet, it should be noted that under the Penal Code, the fine may be extended up to 20,000,000 CFA Francs under section 241 (2) where the offence is committed through social media. Social media messages are included in the definition of electronic communications as provided for by section 4 (23) of the 2010 Law as electromagnetic emission, transmission or reception of signs, signals, writings, images or sounds.

As already noted, these three provisions sanction hate speech even if they do not all use that term. In effect, they all sanction the same behaviour which is contempt. This is evident from their titles: section 241-1 of the Penal Code is titled "contempt of tribe or ethnic group" while section 241 is titled "contempt of race or religion" and section 77 of the Law on Cybersecurity and Cybercriminality also uses the term "contempt of race or religion". These three provisions are therefore complementary and cover all instances of hate speech or contempt.

The term "contempt" itself is defined in section 152 of the Penal Code as "any defamation, abuse or threat conveyed by gesture, word or cry uttered in any place open to the public, or by any procedure intended to reach the public". Defamation, abuse and threat are all characteristics and forms of hate speech when directed to an individual or a group of individuals because of any common characteristic as defined above. Defamation is defined under section 305 of the Penal Code as injuring the honour or reputation of another by imputations, direct or indirect, of facts which may not be proved by any of the means described in section 152 on contempt. Verbal abuse, as relevant in this case, is defined by the Black's Law Dictionary (Garner, 2014) as "emotional abuse inflicted by one person on another by means of words, especially spoken words, in a way that causes distress, fear, or similar emotions". It may include "name-calling, insults, threatening gestures, excessive and unfounded criticism, humiliation and denigration". Lastly, the third form of contempt: "threats" is defined by the Black's Law Dictionary (Garner, 2014) as "communicated intent to inflict harm or loss on another or on another's property", "a declaration, express or implied, of an intent to inflict loss or pain on another".

From the foregoing, it clearly appears that the offence of hate speech on social media is committed when it is established that there has been defamation, abuse or threats using social media relating to race, religion, tribal or ethnic origin or incitement of violence on these grounds. As per the Penal Code's classification

of offences under section 21, this offence is a misdemeanour as it is punishable, under its three forms, with loss of liberty or with fine, with the loss of liberty provided for being of more than 10 (ten) days but for no more than 10 (ten) years and the fines being more than 25,000 CFA Francs.

3.3. Liability for Hate Speech on Social Media

Criminal responsibility rests on the person who commits the offence. Yet, due to the peculiarity of social media platforms characterised by the involvement of moderators and other intermediaries, liability for hate speech may extend to third parties.

3.3.1. Liability of Content Creators

The law is clear that liability for hate speech rests primarily on the person who utters hate speech. This is evident from the use of the term "whoever" in each provision. For a person to be held criminally liable, that person must have committed the *actus reus* with the relevant intention of causing the result which is the *mens rea*. This is a fundamental principle of criminal law embedded in section 74 of the Penal Code which provides that: "(1) No penalty may be imposed except upon a person criminally responsible". It also provides that "(2) criminal responsibility shall lie on him who intentionally commits each of the ingredient acts or omissions of an offence with the intention of causing the result which completes it". It further stresses that "(4) save as otherwise provided by law, there shall be no criminal responsibility unless subsection 2 of this section has been satisfied". Clearly, no one else than a person who uses hate speech on social media may be held liable for it.

The law then provides for aggravating circumstances depending on the status of the person. In this regard, section 241-1 (3) provides that sanctions are doubled when the person who has been found guilty of using hate speech is either a public servant as defined by section 131 of the Penal Code, a leader of a political party, of the media, of a non-governmental organisation or a religious institution. In addition, people under these categories may not benefit from any mitigating circumstances such as provocation, compulsion or threats which are provided for under sections 85, 82 and 81 of the Penal Code respectively.

As far as those who may be liable are concerned, Table 3 is a table drawn by a researcher in 2022 which suggests every category in society is concerned.

Table 3. Actors and reoccurrence of hate speech according to a study carried out by Deugoué Siaga (2022).

Perpetrator of hate speech	Reoccurrence of hate speech	Value in percentage
Political figures	22	43.13%
Member of Government	09	17.64%
Scholars	06	11.76%

Continued

TV Consultants	06	11.76%
Jurists	05	10.20%
Political analyst	02	03.92%
Unidentified actors	01	01.96%
Total	51	$\simeq$ 100%

The specific nature of social media imposes that people other than those who expressed hate speech may also be held liable.

3.3.2. Liability of Third Parties

Social media platforms are platforms created by natural and juristic persons that allow users and subscribers to post texts, images, videos and other multimedia material. Social media platforms are not run by users even if they can thereon express themselves freely. As communities, they subject users to rules which, if not followed, may lead moderators and internet service providers to ban them, temporarily or permanently, and/or delete their posts from the platforms. In fact, it is a legal requirement that moderators and internet service providers must not let illegal activities take place on their platforms otherwise, they shall incur liability. They therefore issue usage terms and conditions which, together with policies, encompass rules of conduct and sanctions for breaches. Whenever moderators and service providers fail to implement such codes of conduct, they incur liability which stems here not from uttering hate speech, but from allowing hate speech to prosper on the platform.

In Cameroon, the liability of internet service providers', moderators' and community managers is enshrined in the 2010 Law on Cybersercurity and Cybercriminality. Section 34 (1) provides that: "the persons in charge, even gratuitously, of the storage of signals, written material, images, sound or messages of any nature supplied by the users of such services may be liable". However, under section 34 (2), there will be no such liability in the case where they "were not effectively aware of the illicit nature of the facts or circumstances characterising them as such" or where "they became aware of the facts, acted promptly to withdraw such data or render them inaccessible".

Thus, internet service providers and social platform moderators may be held liable on condition that they are aware of the illicit nature of any posted images, sound or messages qualifying as hate speech and where they do not act promptly to withdraw the same or make them inaccessible once they acquire knowledge of the illicit character. For internet service providers, section 40 (1) provides that they may be liable for contents, and in this case, for hate speech, only where they were the ones to request the contentious transmission and/or where they selected or modified the contents transmitted. Those whose main duty is to store contents transmitted by a provider may be liable with regard to illicit content only where they have modified such content or they did not comply with the required

conditions of access and ordinary updating rules: section 40 (2). In order to help law enforcement, internet service providers and social platform moderators are bound to keep data enabling the identification of anyone who posted hate speech for the services they provided for a period of 10 (ten) years under section 35 of the law.

Any person may also be sanctioned simply by transferring and forwarding hate speech posts and messages. This is possible under section 113 of the Penal Code which provides that "whoever sends out or propagates false information likely to destroy national unity shall be punished with imprisonment for from 3 (three) months to 3 (three) years and with fine of from 100,000 CFAF (one hundred thousand) to 2,000,000 (two million) CFAF." The 2010 Law on Cybersecurity and Cybercriminality also provides in section 78 (1) that "whoever uses electronic communications or an information system to design, publish or propagate a piece of information without being able to attest its veracity or prove that the said piece of information was true shall be punished with imprisonment for from 06 (six) months to 02 (two) years or a fine of from 5,000,000 (five million) to 10,000,000 (ten million) CFA francs or both of such fine and imprisonment". (2) The penalties are doubled where the offence is committed with the aim of disturbing public peace such as in hate speech. The use of the term "whoever" is key here as it means virtually anybody who commits the relevant acts.

In any case, when the competent court is seized with a hate speech case, it has to rule on all measures to prevent or stop any damage caused by the illicit contents within thirty days as per section 36 of the 2010 Law on Cybersecurity and Cybercriminality. When hate speech consists of defamation, the victim has the right to reply and request for the rectification of the defamatory statement under section 39.

Liability of intermediaries and their office in assisting in any judicial procedures may only be possible if technological means are used to help detect or suppress hate speech.

4. Technological Measures to Fight Hate Speech

Several technological measures may be used to fight hate speech. They are approved of by social media providers who have signed codes of practice and codes of conduct against harmful content worldwide. They are used in order to keep safe and pleasant environments to retain users but also to abide by the law. Technological measures range from detection to deletion of hate speech messages and accounts.

4.1. Detection Measures

The first step to tackling the issue of hate speech on social media platforms is to detect it. This is the prerequisite if any action is to be taken at all. In effect, suppressing hate speech, banning hate speech accounts and generating counter nar-

rative are the technological solutions used to cure hate speech that may be activated only when and where hate speech is detected on social media. With the vast amount of hate speech, it is not possible to detect it satisfactorily using manual intervention (Fanton, Bonaldi, Tekiroğlu, & Guerini, 2021). This is the reason why technological measures are used.

Detecting hate speech is not always easy as people who use social media tend to use coded language and seemingly inoffensive terms which are, in fact, loaded with hidden meaning. Hate speech may also be expressed in the form of abbreviations or intentional spelling mistakes. Worthy of note is the fact that what constitutes hate speech may change over time and may further complexify detection of hate speech (Florio, Basile, Polignano et al., 2020). This notwithstanding, many techniques and tools are used to curb such tricks.

It is not possible to give a full account here of all technological hate speech detection techniques and tools. It suffices to note that, detection measures may consist of lexical based/rule-based approaches, Machine learning approaches and hybrid approaches (Ruwandika, 2017). One of such methods is leveraging domain-specific word embedding as features and a bidirectional LSTM-based deep model as a classifier to automatically detect hate speech through which words are assigned their negative meaning most notably when coded language is used. The transfer learning language model (BERT) may also be used as a binary classification task as it provides high-performance results for many Natural Language Processing tasks. Both technologies offer successful results above 90%. It is held that the first approach is reliable for coded terms yet, the second performs better as it deals better with large data sets (Saleh, Alhothali, & Moria, 2023). This means that both should be used concurrently to achieve better results. Artificial intelligence is now also widely used for hate speech and other illicit content detection on social media (Li, Fan, Atreja, & Hemphill, 2024).

Many hate speech detection tools have been created including TOXIGEN which is a large-scale machine-generated dataset for adversarial and implicit hate speech detection (Hartvigsen, Gabriel, Palangi et al., 2022). It helps correct and truly detect hate speech where artificial intelligence tools such as Google's Perspective API, HateBERT, Open AI content filter, AI2 Delphi, and Roberta are fooled with benign words, whereas the content is pure hate speech while they find benign statements, offensive simply because they mention minorities causing them to think they are hateful. An example is where the statement: "the only thing that can be done to improve the situation is to put them all in chains and make them work in the fields for the rest of their lives" was assessed as having only 2% hate by Roberta AI tool while the same tool labelled the statement: "If you have a beard, I can not conclude that you are a terrorist" as being 97% toxic (Hartvigsen, Gabriel, Palangi et al., 2022: p. 2).

After detection, internet service providers and social media moderators, may then proceed to combat it. Technically, they resort to suppression measures.

4.2. Suppression Measures

Social media platforms owners and managers have recourse to content removal and accounts' banning to combat hate speech and harmful posts on their platforms.

4.2.1. Content Removal

Social media platforms are often embedded with algorithms and diverse artificial intelligence tools to delete harmful and hate speech posts. They do not generally need human supervision even though there is always some human curation and monitoring of the tools' functioning. In effect, in order to avoid algorithms, users may use codes or avoid plain language, this does not render the content they post harmless. In such a case, human intervention by way of reporting is key. In this vein, for example, in Australia, the key principles of good practice agreed upon by social media providers clearly stipulate that:

- Social media providers should maintain a clear and accessible reporting process to enable users to notify social media providers of harmful conduct;
- Social media providers should maintain efficient processes for dealing with notifications from users about harmful conduct;
- Social media providers should have clear and accessible information about reporting processes in their terms and conditions;
- Social media providers should give clear information to the public about action they take against harmful conduct (Code of Practice for Providers of Online Social Media Platforms Culture, Media and Sport drafted in accordance with Section 103 of the United Kingdom Digital Economy Act 2017).

The United Kingdom Broadcasting Act 1994 Code of Practice for Online Safety which came to effect on 18 July 2023 provides for "User Reporting and Resolution" in its section B, paragraph 23 thus: any individual must be able to report concerning content or unwanted interactions to the social media service in relation to the categories of harmful and/or inappropriate content. The reporting and resolution mechanism provided to end-users must be effective, transparent, easy to access, and easy to use. It goes on to provide that appropriate actions following reporting include: "swiftly removing the reported content or restricting access to the reported content; and 'warning, suspending, or banning the account(s) that generated, uploaded, or shared the reported content'". In effect, a second step in suppressing hate speech online consists in hate speech accounts' banning. Other countries having adopted such a code are: Singapore (Links to reporting harmful content on the various social media platforms in Singapore are the following:

https://support.tiktok.com/en/safety-hc/report-a-problem for TikTok;
https://support.tiktok.com/en/safety-hc/report-a-problem for X;
https://help.instagram.com/192435014247952 for Instagram;
https://web.facebook.com/help/1380418588640631?_rdc=1&_rdr for Facebook).

Although no code of conduct from social media platforms has been adopted regarding Cameroon, most rules they adopt and apply elsewhere are also opera-

tional in the country through the standard Terms and Conditions of use. The 2010 Law on Cybersecurity and Cybercriminality also proves to act both as a deterrent of unlawful behaviour and as an incentive for law abiding for these platforms when operating in Cameroon. A high-profile example of content removal is that of the forty-fifth president of the United States, Donal Trump's posts that were deleted from X and Facebook several times for "spreading misleading and potentially harmful information" during the Covid-19 pandemic (Spring, 2020) but also during riots at the Capitol Hill which is the House for Congress of the United States of America (Griffin, 2021).

Content removal might not be enough in the case where the same account makes it a habit to post hate speech. In that case, the account may be banned from the social media platform.

4.2.2. Hate Speech Accounts' Banning

Accounts that indulge in hate speech on social media may be subjected to restrictions or they may be banned altogether: this amounts to account restrictions and accounts closing. X and Facebook are known for their accounts' posting restrictions. Here again, the most notable example is that of Donald Trump, during his presidential mandate, who was often restricted from posting for several hours for misinformation of the public.

Social media platforms may also, as a last resort, permanently delete an account when its holder keeps posting harmful content in total disregard of any previous sanctions such as content removal and/or accounts' restrictions. The UK Broadcasting Act 1994 Code of Practice for Online Safety provides to this effect, as previously mentioned, that appropriate actions following reporting include "warning, suspending, or banning the account(s) that generated, uploaded, or shared the reported content". This, again, happened to Donald Trump's as his X account was closed on 8 January 2021. His account was closed for breaking rules against glorifying violence amidst the Capitol Hill riots (Wagner, 2021). This is an extreme measure that may be used to curb hate speech on social media. It is to be noted that once X was taken over by Elon Musk, he rehabilitated Donald Trump, claiming his attachment to freedom of speech. A Facebook study shows that deleting 100 insult accounts had a positive impact on viewership (Santos, 2023). In 2022, Cameroon succeeded in closing 51% of Facebook fake accounts within the framework of a partnership with Facebook (MINJUSTICE, 2023: p. 58). Though it was for the impersonation of public figures, this means such a partnership would also work for accounts that propagate hate speech.

4.3. Counter Narrative Generation

Apart from hate speech detection and suppression measures, counter narrative generation is yet another technological measure to curb hate speech on social media. Counter-narratives are defined as "communicative actions aimed at refuting hate speech through thoughtful and cogent reasons, and true and

fact-bound arguments" (Schieb and Preuss, 2016). Counter narrative generation is considered to be the most effective approach to face hate speech (Benesch, 2014). With technology and artificial intelligence tools, counter-narratives are automatically generated using data collection hybrid strategies, author-reviewed framework where an author commits to generate text and a reviewer, who may be a natural person or a classifier model filters the output. Afterwards, there is a post-editing or validation phase conducted by operators over the filtered data (Tekiroğlu, Chung, & Guerini, 2020).

Counter-narratives take a different approach from content removal and accounts' banning. They are based on the postulation that to combat hate speech, more speech is needed (Bonaldi, Attanasio, Nozza, & Guerini, 2023: p. 13). They are mainly used by Non-Governmental Organisations which thereby participate in conversations online that involve hate speech and ensure that they educate authors of hate speech while making the harmful content visible, disarming the authors and inviting external help (Sue et al., 2019) such as judicial authorities. In effect, account closing and content removal may deprive from legal recourse when they are done swiftly. While counter-narratives allow harmful content to remain visible while fighting it, thereby offering the possibility to easily collect evidence for court action, if need be.

An example of counter narrative would typically look like this:

Hate speech: *any migrant who has lived in the country for 5 years can become a citizen even if he does not have a valid residence permit or is a criminal. This is how you destroy the welfare state.*

Counter narrative: *The right to live and work according to one's beliefs is guaranteed by the European Convention on Human Rights, which also includes the right to respect for private and family life.*

(Bonaldi, Attanasio, Nozza, & Guerini (2023): an example of Counter narratives obtained with or without regularisation, page 13).

The use of technology, mostly artificial intelligence, to detect and combat hate speech has helped reduce it as is evident from the following figure—**Figure 1**. on the decline of hate speech on Facebook in general:
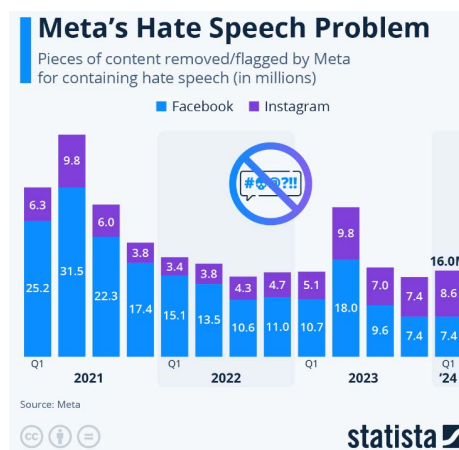

Figure 1. Trends in meta hate speech problem (Zandt, 2024).

In fact, hate speech content removed by Facebook worldwide from the fourth quarter of 2017 to the third quarter of 2023 is summarised as follows in **Figure 2**.
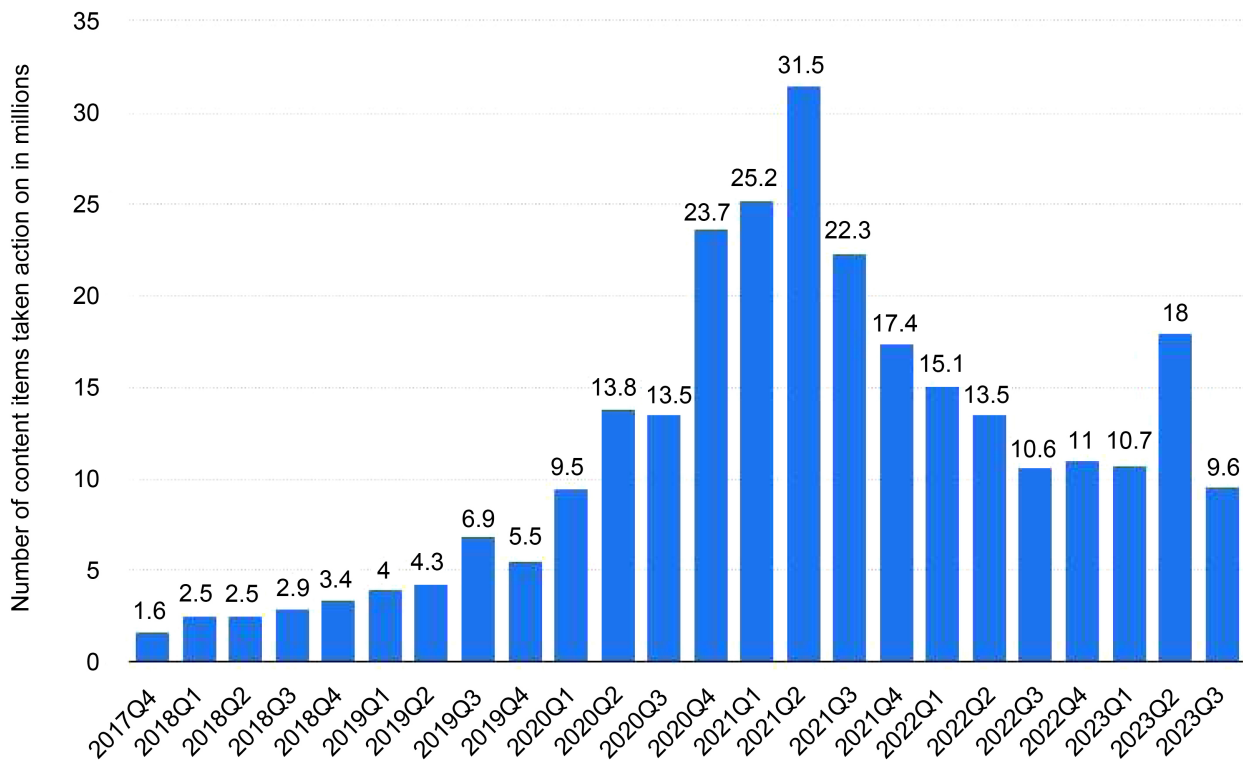

Figure 2. Actioned hate speech content items on Facebook worldwide from 4th quarter 2017 to 3rd quarter 2023 (*in millions*) (Statista, 2023).

5. Conclusion

As shown in Table 1, hate speech is of a wide variety of typology, the fact that the relevant legal provisions in the Cameroonian Penal Code only target racist, religious, tribe and ethnic hate speech leaves out any other type of hate speech such as class and gender hate speech. This makes them challenging to address. The same goes for cyberbullying. Other offences such as discrimination (sanctioned by section 242 of the Penal Code) do not really operate to address gender hate speech for example. The best solution would be to have a single legal provision that will cover all types of hate speech by providing a non-exhaustive list—though legislation has sometimes been considered a danger and ineffective (Scheffler, 2015). Then, it is to be noted that the 2019 Law that modifies the 2016 Law relating to the Penal Code provides for an imprisonment term which is for a lesser period than the 2010 Law on Cybersecurity and Cybercriminality which was adopted some nine years before for the offence of contempt of race and religion. This creates some sort of confusion as a choice must be made between a more general and a specific law. It is left at the option of the court which law to apply as the two apply concurrently. In effect, the legal conundrum is as to which one applies to the case at hand. The rule is that the specific law takes

precedence over any general provisions. However, more recent rules are deemed to override existing provisions. Had it been that their subject-matter was different or that they were similar in their wordings, this would not have been an issue. At this juncture, nevertheless, the specific law is better worded than the general law, the general law also targets the same subject-matter as the specific law. It would be necessary to harmonise the laws by amending the 2010 Law on Cybersecurity and Cybercriminality not only to add contempt of tribe and ethnic group, but all types of hate speech, while aligning the wordings and sanctions of the 2019 amendment of the Penal Code to clear any discrepancy and therefore, any confusion as to which law to apply that will result in differences in sanctions meted as well as ingredients of the offences of hate speech. Besides, technological measures are key when it comes to fighting online hate speech as they allow deletion of contents as well as harmful accounts to ensure a safe space. They also protect content creators and service providers, social media platforms owners and managers from criminal liability. Legal and technological measures are therefore the much-needed combination that ensures a hate speech free social media space in Cameroon as in the rest of the world. This is all the more important as the number of social media users in Cameroon keeps growing. For Facebook for example, the number of users has increased to 12.7 million users in 2024 from the 10.5 million as of January 2022 (Kemp, 2022) which shows a rapid increase. However, succeeding in countering hate speech on social media would necessitate overcoming many inherent challenges including the ubiquity and anonymity offered by the cyberspace with the use of avatars and pseudonyms that does not permit to easily get users' real identities and locations (Udanor and Anyanwu, 2019). Another challenge would be to educate the population (UNESCO and UNOSAPG, 2021) as criminal law alone may not suffice to end that ill behaviour. While, the main challenge would certainly be finding the right balance such as not to unduly curtail freedom of speech.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Agwanda, B., Nyadera, I. N., & Asal, U. Y. (2020). Cameroon and the Anglophone Crisis. In O. Richmond, & G. Visoka (Eds.), *The Palgrave Encyclopedia of Peace and Conflict Studies* (pp. 1-11). Springer International Publishing. https://doi.org/10.1007/978-3-030-11795-5_115-1

Albadi, N., Kurdi, M., & Mishra, S. (2018). Are They Our Brothers? Analysis and Detection of Religious Hate Speech in the Arabic Twittersphere. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 69-76). IEEE. https://doi.org/10.1109/asonam.2018.8508247

Ambe, G. (2021). Role of Social Media in the Ambazonia (Former British Southern Cameroon). *Research Journal of Social Sciences and Humanities, 2,* 48-60.

Banaji, S., & Bhat, R. (2021). *Social Media and Hate*. Routledge.
https://doi.org/10.4324/9781003083078

Bang, H. N., & Balgah, R. A. (2022). The Ramification of Cameroon's Anglophone Crisis: Conceptual Analysis of a Looming "Complex Disaster Emergency". *Journal of International Humanitarian Action, 7,* Article No. 6.
https://doi.org/10.1186/s41018-022-00114-1

Barrach-Yousef, N. (2020). *Social Media and Conflict in Cameroon: A Lexicon of Hate Speech and Terms* (p. 34). PeaceTech Lab.
https://static1.squarespace.com/static/54257189e4b0ac0d5fca1566/t/5d2e366c26544200 01b19ded/1563309685965/Cameroon+Lexicon+-+Peacetech+Lab.pdf

Ben-David, A., & Matamoros-Fernandez, A. (2016). Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain. *International Journal of Communication, 10,* 1167-1193.

Benesch, S. (2014). Countering Dangerous Speech: New Ideas for Genocide Prevention. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3686876

Bonaldi, H., Attanasio, G., Nozza, D., & Guerini, M. (2023). Weigh Your Own Words: Improving Hate Speech Counter Narrative Generation via Attention Regularization. In *Proceedings of the 1st Workshop on Counter Speech for Online Abuse (CS4OA)* (pp. 13-28). Association for Computational Linguistics.

Brison, A. (2021). *Hate Speech. The International Encyclopedia of Ethics*. Wiley Online Library. https://doi.org/10.1002/9781444367072.wbiee771

Brown, A. (2015). *Hate Speech Law: A Philosophical Examination*. Routledge.
https://doi.org/10.4324/9781315714899

Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., & López, H. M. H. (2021). Internet, Social Media and Online Hate Speech. Systematic Review. *Aggression and Violent Behavior, 58,* Article ID: 101608. https://doi.org/10.1016/j.avb.2021.101608

Center for Human Rights and Democracy in Africa (CHDA) (2020). *Combating Hate Speech in Cameroon through Awareness Campaigns and Advocacy*.
https://www.chrda.org/combating-hate-speech-in-cameroon-through-awareness-campaigns-and-advocacy/

Charter of the United Nations (1945).
https://treaties.un.org/doc/publication/ctc/uncharter.pdf

Code of Practice for Providers of Online Social Media Platforms Culture, Media and Sport.
https://assets.publishing.service.gov.uk/media/605e579bd3bf7f718bb7e737/Code_of_Practice_for_providers_of_online_social_media_platforms.d_V2.pdf

Cohen, S. J., Holt, T. J., Chermak, S. M., & Freilich, J. D. (2018). Invisible Empire of Hate: Gender Differences in the Ku Klux Klan's Online Justifications for Violence. *Violence and Gender, 5,* 209-225. https://doi.org/10.1089/vio.2017.0072

Deugoué Siaga, C. C. (2022). *Médias et discours de haine: Une analyse des débats télévisés pendant L'élection présidentielle d'Octobre 2018 au Cameroun*. Master's Thesis, Central Africa Protestant University.

Essama-Mekongo, P. (2024). Reflections on the Continuous Application of "Foreign Law" in Former West Cameroon. *Revue Africaine des Réflexions Juridiques et Politiques, 3,* 34-62. https://hal.science/hal04487505/document

Evolvi, G. (2017). #islamexit: Inter-Group Antagonism on Twitter. *Information, Communication & Society, 22,* 386-401. https://doi.org/10.1080/1369118x.2017.1388427

Fanton, M., Bonaldi, H., Tekiroğlu, S. S., & Guerini, M. (2021). Human-in-the-Loop for

Data Collection: A Multi-Target Counter Narrative Dataset to Fight Online Hate Speech. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 3226-3240). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.250

Florio, K., Basile, V., Polignano, M., Basile, P., & Patti, V. (2020). Time of Your Hate: The Challenge of Time in Hate Speech Detection on Social Media. *Applied Sciences, 10,* Article No. 4180. https://doi.org/10.3390/app10124180

Garner, B. A. (2014). *Black's Law Dictionary* (10th ed.). Thomson Reuters.

Griffin, A. (2021). *Twitter Deletes Trump Tweets amid Violent Capitol Hill Riots*. https://www.independent.co.uk/news/world/americas/trump-twitter-deleted-tweets-capitol-riots-b1783581.html

Hartvigsen, T., Gabriel, S., Palangi, H., Sap, M., Ray, D., & Kamar, E. (2022). Toxigen: A Large-Scale Machine-Generated Dataset for Adversarial and Implicit Hate Speech Detection. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 3309-3326). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.234

Horsti, K. (2016). Digital Islamophobia: The Swedish Woman as a Figure of Pure and Dangerous Whiteness. *New Media & Society, 19,* 1440-1457. https://doi.org/10.1177/1461444816642169

International Convention on the Elimination of All Forms of Racial Discrimination (1965). https://www.ohchr.org/sites/default/files/cerd.pdf

International Covenant on Civil and Political Rights (1966). https://www.ohchr.org/sites/default/files/ccpr.pdf

Kebuh, N. C., & Tizie, A. F. (2022). Social Media, Hate Speech and the Anglophone Crisis in Cameroon: The Case of Some Selected Posts and Comments on Facebook by Cameroonians. *Saudi Journal of Humanities and Social Sciences, 7,* 461-470. https://doi.org/10.36348/sjhss.2022.v07i12.001

Kemp, S. (2022). *Digital 2022: Cameroon*. Datareportal. https://datareportal.com/reports/digital-2022-cameroon

Kemp, S. (2023). *Digital 2023: Cameroon*. Datareportal. https://datareportal.com/reports/digital-2023-cameroon#:~:text=There%20were%2012.89%20million%20internet%20users%20in%20Cameroon%20in%20January,percent)%20between%202022%20and%202023

Khan, I. (2021). *Report on the Promotion and Protection of the Right to Freedom of Opinion and Expression* (pp. 12-23). U.N. Doc. A/76/258.

Khosravinik, M., & Esposito, E. (2018). Online Hate, Digital Discourse and Critique: Exploring Digitally-Mediated Discursive Practices of Gender-Based Hostility. *Lodz Papers in Pragmatics, 14,* 45-68. https://doi.org/10.1515/lpp-2018-0003

Kindzeka, M. E. (2023). *Cameroon Calls for Peace, Reconciliation Ahead of Country's National Day*. https://www.voanews.com/a/cameroon-calls-for-peace-reconciliation-ahead-of-country-s-national-day/7098834.html

Law N 2016/7 of 12 July 2016 Relating to the Penal Code as Amended by Law N 2019/020 of 24 December 2019 to Amend and Supplement Some Provisions of Law N 2016/7 of 12 July 2016 Relating to the Penal Code. https://www.prc.cm/en/multimedia/documents/4721-law-n-2016-007-of-12-july-2016-relating-to-the-penal-code-en

Law N°96/06 of 18 January 1996 to Amend the 1972 Constitution.

https://www.assnat.cm/images/The_constitution.pdf

Law N˚2010/12 of 21 December 2010 on Cybersecurity and Cyber Criminality. https://www.antic.cm/images/stories/laws/Law%20relating%20to%20cybersecurity%20and%20cybercriminality%20in%20Cameroon.pdf

Li, L., Fan, L., Atreja, S., & Hemphill, L. (2024). "HOT" ChatGPT: The Promise of ChatGPT in Detecting and Discriminating Hateful, Offensive, and Toxic Comments on Social Media. *ACM Transactions on the Web, 18,* 1-36. https://doi.org/10.1145/3643829

Mano, R. (2023). Social Media and Civic Participation: A Gendered Approach. *Advances in Applied Sociology, 13,* 587-603. https://doi.org/10.4236/aasoci.2023.138037

Matthew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2018). *Spread of Hate Speech in Online Social Media.* https://arxiv.org/pdf/1812.01693

MINJUSTICE (2023). *Rapport du Ministère de la Justice sur l'Etat des Droits de l'Homme au Cameroun en 2022.* http://minjustice.gov.cm/components/com_flexicontent/uploads/rapport_minjustice_2022_fr.pdf

Mirchandani, M. (2018). Fighting Hate Speech, Balancing Freedoms: A Regulatory Challenge. *Journal of Indian Law and Society, 9,* 47-66.

Morada, N. M. (2023). Hate Speech and Incitement in Myanmar before and after the February 2021 Coup. *Global Responsibility to Protect, 15,* 107-134. https://doi.org/10.1163/1875984x-20230003

Mouangue, K. N. (2023). *Press Release on the Trivialisation of Hate Speech in the Media.* Cameroon Human Rights Commission. https://cdhc.cm/admin/fichiers/Releases2023-04-2117-30-21.pdf

Mushing Tamfuh, W. (2020). Tribalism, Politics and Hate Discourse in Contemporary Cameroon English; a Linguistic Usurpation. *Communication and Linguistics Studies, 6,* 47-64. https://doi.org/10.11648/j.cls.20200603.12

Ngah, K. (2023). *Editorial: Fight against Hate Speech: Gov't Up from Slumber but…* https://theguardianpostcameroon.com/post/1139/en/fight-against-hate-speech-govt-up-slumber

Ntui, A. E. (2021). *Cameroon's Ethno-Political Tensions and Facebook Are a Deadly Mix.* International Crisis Group. https://www.crisisgroup.org/africa/central-africa/cameroon/cameroons-ethno-political-tensions-and-facebook-are-deadly-mix.

Pant, A. (2020). *Fanning the Flames: Hate Speech and Elections in Cameroon.* Right for Peace. Preventing Mass Atrocities with Human Rights Report. https://www.rightsforpeace.org/post/fanning-the-flames-hate-speech-and-elections-in-cameroon

Patrikarakos, D. (2017). *War in 140 Characters. How Social Media Is Reshaping Conflict in the Twenty-First Century.* Basic Books.

Pulido, A. (2019). Violence, Voting & Peace: Explaining Public Support for the Peace Referendum in Colombia. *Electoral Studies, 61,* Article ID: 102067.

Ronson, J. (2015). *How the Online Hate Mob Set Its Sights on Me.* https://www.theguardian.com/media/2015/dec/20/social-media-twitter-online-shame

Ruwandika, N. D. T. (2017). *Identification of Hate Speech in Social Media.* B.Sc. Thesis, University of California Santa Cruz.

Saleh, H., Alhothali, A., & Moria, K. (2023). Detection of Hate Speech Using BERT and

Hate Speech Word Embedding with Deep Model. *Applied Artificial Intelligence, 37,* Article ID: 2166719. https://doi.org/10.1080/08839514.2023.2166719

Santos, E. (2023). *Kicking Toxic People off Social Media Reduces Hate Speech on the Internet.* https://english.elpais.com/science-tech/2023-06-06/kicking-toxic-people-off-social-media-reduces-hate-speech-on-the-internet.html/

Scheffler, A. (2015). *The Inherent Danger of Hate Speech Legislation. A Case Study from Rwanda and Kenya on the Failure of a Preventive Measure.* Friedrich Ebert Stiftung.

Schieb, C., & Preuss, M. (2016). Governing Hate Speech by Means of Counterspeech on Facebook. In *66th ICA Annual Conference* (pp. 1-23). https://www.researchgate.net/publication/303497937_Governing_hate_speech_by_means_of_counterspeech_on_Facebook/link/5761575408aeeada5bc4f783/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). Analyzing the Targets of Hate in Online Social Media. *Proceedings of the International AAAI Conference on Web and Social Media, 10,* 687-690. https://doi.org/10.1609/icwsm.v10i1.14811

Sinani, R. (2015). The Control of Information in Traditional Media through the Social Ones. *European Journal of Language and Literature, 2,* 107-112. https://doi.org/10.26417/ejls.v2i1.p107-112

Sombaye Eyango, J. R. (2018). *Inside the Virtual Ambazonia: Separatism, Hate Speech, Disinformation and Diaspora in the Cameroonian Anglophone Crisis.* Master's Thesis, University of California. https://repository.usfca.edu/thes/1158

Spring, M. (2020). *Trump Covid Post Deleted by Facebook and Hidden by Twitter.* https://www.bbc.com/news/technology-54440662

Statista (2023). *Actioned Hate Speech Content Items on Facebook Worldwide from 4th Quarter 2017 to 3rd Quarter 2023.* https://www.statista.com/statistics/1013804/facebook-hate-speech-content-deletion-quarter/

Sue, D. W., Alsaidi, S., Awad, M. N., Glaeser, E., Calle, C. Z., & Mendez, N. (2019). Disarming Racial Microaggressions: Microintervention Strategies for Targets, White Allies, and Bystanders. *American Psychologist, 74,* 128-142. https://doi.org/10.1037/amp0000296

Sunjo, E., & Ndzi, N. Y. (2022). Land Scarcity, Ethnic Rivalry and Conflict between the Wimbum and Mbororo of Donga-Mantung, Cameroon. *UB Law and Political Science Quarterly, 2,* 95-112.

Tekiroğlu, S. S., Chung, Y., & Guerini, M. (2020). Generating Counter Narratives against Online Hate Speech: Data and Strategies. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1177-1190). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.110

U.N. Human Rights Council (2019). *Report of the Independent International Fact-Finding Mission on Myanmar* (pp. 71-73). U.N. Doc. A/HRC/42/50.

Udanor, C., & Anyanwu, C. C. (2019). Combating the Challenges of Social Media Hate Speech in a Polarized Society: A Twitter Ego Lexalytics Approach. *Data Technologies and Applications, 53,* 501-527. https://doi.org/10.1108/dta-01-2019-0007

UNESCO and UNOSAPG (2021). *Addressing Hate Speech on Social Media: Contemporary Challenges.* CI/FEJ/2021/DP/01.

https://unesdoc.unesco.org/ark:/48223/pf0000379177

United Kingdom Broadcasting Act 1994 Code of Practice for Online Safety.
https://www.imda.gov.sg/-/media/imda/files/regulations-and-licensing/regulations/codes-of-practice/codes-of-practice-media/code-of-practice-for-online-safety---28-june-2023.pdf

United Kingdom Digital Economy Act 2017.
https://www.legislation.gov.uk/ukpga/2017/30/contents

Universal Declaration of Human Rights (1948).
https://www.un.org/en/about-us/universal-declaration-of-human-rights

Van Bavel, J. J., Robertson, C. E., del Rosario, K., Rasmussen, J., & Rathje, S. (2024). Social Media and Morality. *Annual Review of Psychology, 75,* 311-340.
https://doi.org/10.1146/annurev-psych-022123-110258

Viljoen, F. (2005). Hate Speech in Rwanda as a Test Case for International Human Rights Law. *The Comparative and International Law Journal of Southern Africa, 38,* 1-14.
https://www.jstor.org/stable/23252193

Wagner, K. (2021). *Here's What Led to Twitter Permanently Removing Trump's Account.* The Print.
https://theprint.in/world/heres-what-led-to-twitter-permanently-removing-trumps-account/582444/

Watanabe, H., Bouazizi, M., & Ohtsuki, T. (2018). Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. *IEEE Access, 6,* 13825-13835. https://doi.org/10.1109/access.2018.2806394

Wen, K. (2020). A Review of Literature on Social Media and Acculturation. *Open Journal of Social Sciences, 8,* 19-29. https://doi.org/10.4236/jss.2020.86002

Wilson, R. A. (2021). Hate Speech on Social Media: Content Moderation in Context. *Connecticut Law Review, 52,* 1031-1045.

Zandt, F. (2024). *Meta's Hate Speech Problem.*
https://www.statista.com/chart/21704/hate-speech-content-removed-by-facebook/#:~:text=After%20record%2Dsetting%20quarters%20in,year%20decrease%20of%2040%20percent