Int J Semiot Law (2022) 35:2187–2191
https://doi.org/10.1007/s11196-022-09953-y

Editorial Introduction

Jacob Mchangama[1] · **Natalie Alkiviadou**[1]

Accepted: 21 October 2022 / Published online: 7 November 2022

In the digital age, social media has supercharged the ability for billions of people to impart instantly and access a range of semiotic systems across borders and real time. But, with this ground-breaking development in terms of human communication comes an increase in the visibility and manifestation of phenomena such as hate speech. There are 5 billion internet users worldwide and 4.62 billion social media users around the world. Authoritarian regimes as well as liberal democracies are placing increasing pressure on social media platforms to deal with allegedly harmful content. National and regional legislative measures that dramatically enhance platform liability for content developed by users include the German Network Enforcement Act (NetzDG) and the EU's forthcoming Digital Services Act (DSA). Platform liability (at the risk of penalties) places free speech at risk and potentially shrinks civic space. The German NetzDG blueprint for intermediary liability has been followed by over 20 countries around the world, including Belarus, Turkey, Venezuela, and Russia [1]. Such measures render private companies, not bound by International Human Rights Law (IHRL) arbiters of fact and law. To meet obligations and avoid hefty fines, social media platforms are adopting a "better safe than sorry" approach, increasingly relying on Artificial Intelligence (AI) proactively to remove even contentious areas of speech such as hate speech. The use of AI, even without human supervision, is a necessity when it comes to content that could never be ethically or legally justifiable, such as CASAM. As highlighted in a Council of Europe report, the use of AI for hate speech regulation directly impacts the freedom of expression, which raises concerns *vis-à-vis* the rule of law and, in particular, notions of legality, legitimacy and proportionality [2]. Moreover, regardless of the technical specifications of a particular mechanism, proactive identification (and removal) of hate speech is a prior restraint of speech with all the legal issues that this entails.

✉ Jacob Mchangama
jacob@justitia-int.org

✉ Natalie Alkiviadou
natalie@justitia-int.org
https://futurefreespeech.com/
https://justitia-int.org/en/

[1] Justitia, Copenhagen, Denmark

1 The Importance of Freedom of Expression[1]

The origins of free speech can be traced back to the Athenian democracy some 2500 years ago. However, it was not until the 18th century that free speech was codified as an individual and inalienable legal right protected in laws and constitutions as what the influential Cato´s Letters called the 'the great bulwark of liberty.' After World War II, on a United Nations level, Article 19 of the Universal Declaration of Human Rights (UDHR) provided that 'everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.' Drafted in the sixties and in force in the seventies, Article 19 of the International Covenant on Civil and Political Rights (ICCPR), one of the basic tenets of International Human Rights Law (IHRL), provided for the freedom of opinion and expression:

1. Everyone shall have the right to hold opinions without interference;
2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice;
3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:

    (a) For respect of the rights or reputations of others;
    (b) For the protection of national security or of public order (ordre public), or of public health or morals.

As well as the limitations found in Article 19(3), Article 20 of the ICCPR contains a specific prohibition on two types of expression providing that:

1. Any propaganda for war shall be prohibited by law.
2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.

Notwithstanding the fundamental position held by the freedom of expression in the international legal framework, 'this freedom does not enjoy such a position of primacy among rights that it trumps equality rights' [3]. However, as noted by General Comment 34 of the Human Rights Committee (HRC) which monitors the implementation of the ICCPR, Article 19 of the ICCPR embraces 'even expression that may be regarded as deeply offensive' [4].

[1] For a detailed historical overview of free speech see: Jacob Mchangama, 'Free Speech: A History from Socrates to Social Media' (Basic Books 2022).

Either way, as private entities, social media platforms are not signatories to or bound by such documents. Nevertheless, as the former Special Rapporteur for Freedom of Opinion and Expression David Kaye has argued, IHRL is a means to facilitate a more rights-compliant and transparent model of content moderation. At the same time, its global nature may also prove useful in dealing with the differences in national perception and legislation that characterize the global ecosystem of online expression. Yet, applying IHRL to private companies is a difficult task involving a plethora of challenges and dilemmas [5].

Interestingly and beyond the legal ambit are the civic perspectives towards free speech. Justitia's global survey entitled "Who Cares about Free Speech" [6] asked citizens in thirty-three countries questions about their attitude towards free speech in principle and tested their attitudes when confronted with controversial speech and trade-offs. Support for the principle of free speech was found to be very high, averaging around 90% in all countries, dropping substantially when put to the test against supposedly competing values such as statements offensive to religion and minority groups or statements disclosing information that could destabilize the national economy. To assess the actual support for free speech in a country, the survey included a composite measure, the Justitia Free Speech Index [7] based on answers to eight "tough" questions. The top scorer is Norway with eighty points average approval on all eight questions while Pakistan is at the bottom with only thirty-eight points.

2 The Notion of Hate Speech

Hate speech does not enjoy a universally accepted formulation, with most States and institutions adopting their own understanding of what hate speech entails [8] without defining it. One of the few documents, albeit non-binding, which has sought to elucidate the meaning of hate speech, is the Recommendation of the Council of Europe Committee of Ministers on hate speech [9]. It provides that this term is to be:

> understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerant expression by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.

In the framework of academic commentary, a plethora of definitions has been put forth to describe hate speech. In exploring its different formulations, Belavusau underlines that hate speech is 'deeply rooted in the ideologies of racism, sexism, religious intolerance, xenophobia, and homophobia'[10]. In addition, he argues that pinpointing the grounds from which hate speech may arise is also a tricky task and poses the question of where limits are to be drawn. According to Matsuda, hate speech which is discussed in the sphere of racism, contains three central elements: namely, that the message is 'of racial inferiority, the message is directed against historically oppressed groups and the message is persecutory, hateful and degrading' [11]. McGonagle offers a broad interpretation of hate speech in terms of threshold

but not in terms of content and target groups, arguing that 'virtually all racist and related declensions of noxious, identity-assailing expression could be brought within the wide embrace of the term' [12]. Smolla defines it as a 'generic term that has come to embrace the use of speech attacks based on race, ethnicity, religion and sexual orientation or preference' [13]. Although some common elements can be discerned from these approaches to hate speech and the variations therein, Kiska (2012) noted that 'hate speech seems to be whatever people choose it to mean' [14].

Turning to social media platforms themselves, this table demonstrates the (growing) areas of 'protected characteristics' in hate speech policies and shows a comparative overview with national criminal law (Denmark) and the ICCPR.

Protected characteristics	Facebook	YouTube	Danish Penal Code (Section 266 b)	Art 20(2) ICCPR
Race/skin color	✓	✓	✓	✓
National origin	✓	✓	✓	✓
Ethnic origin	✓	✓	✓	
Religion	✓	✓	✓	✓
Disability	✓	✓	✓	
Sexual orientation	✓	✓	✓	
Gender, gender identity, gender characteristics.	✓	✓	✓	
Age (if referenced with other protected characteristics)	✓	✓	–	
Caste	✓	✓	–	
Immigrant / immigrant status	✓	✓	–	
Refugees	✓	–	–	
Migrants	✓	–	–	
Asylum seekers	✓	–		
Serious disease	✓	–	–	
Employment (if referred to with other protected characteristics)	✓	–	–	
Victims of a major violent event + relatives	–	✓	–	
Veteran status	–	✓	–	

3 The Contributions

In light of above developments, drawing the line between hate speech and the legitimate exercise of freedom of speech is more challenging than ever. To tackle the complex question of 'free speech versus hate speech,' this special issue looks at aspects of free speech and hate speech over time and space. Papers analyze issues such as sexist hate speech, the impact of the COVID-19 pandemic on speech, as well as good and bad national and regional practices *vis-à-vis* the treatment of hate speech. The series starts off with an assessment of the contemporary European challenges of free speech, looking at increasing platform liability, the approach of the European Court of Human Rights to this issue and the impact of AI on online speech. It continues with two thematic issues, namely truth and humour. The first paper in this stream looks at what kinds of speech ought to be protected under a free speech principle. The author seeks to advance an account of "assertion", found in the speech act theory, that can identify speech which contributes to truth-discovery in a nuanced way. After the issue of truth comes that of humour which is depicted in a paper which assesses case law on humour,

free speech and hate speech at the European Court of Human Rights. This is the first research output of the Dutch Research Council's project 'Forensic Humour Analysis: Rethinking Offensive Humour and its Legal Regulation' (2022–2027). The next stream looks at national approaches to hate speech, specifically hate speech in South Africa and Japan. Papers then assess particular types of hate speech, namely sexist hate speech and its treatment by International Human Rights Law and (ii) anti-semitism, considered through a comparative of linguistic analysis versus legal judgment. The special volume also incorporates two works on speech during the pandemic, looking at the rise of conspiracy theories but also covid masks as semiotic expressions of hate. Related to this is a study of fake news detection through a forensic linguistic analysis. The closing article assesses an analysis of the semio/pragmatic conditions for the production of performativity inherent in hate speech across different cultural universes of discourse. It takes a cue from the decision of the Court of Justice of the European Union in the case of *Ewa Glawischnig-Piesczek v. Facebook Ireland* Limited which raised the issue of the transcultural/trans-territorial signification of hate speech and hate crimes.

References

1. Mchangama, J., and N. Alkiviadou. 2020. *The Digital Berlin Wall Act 2*:. How the German Prototype for Online Censorship went Global: Justitia.
2. Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications'. 2017. Council of Europe, DGI(2017) 12: 18.
3. Farrior, Stephanie. 1996. Molding The Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech. *Berkley Journal of International Law* 14: 1.
4. General Comment 34 CCPR/C/GC/34. 2011. *Article 19: Freedom of Opinion and Expression*. United Nations Human Rights Committee.
5. For a discussion of these dilemmas and recommendations for ways forward please consult: Mchangama, J., N. Alkiviadou, and R. Mendiratta, A Framework of First Reference – Decoding a Human Rights Approach to Content Moderation in the era of "platformization" (2021) Justitia.
6. Svend-Erik. 2021. *Skaaning and Suthan, Krishnarajan 'Who Cares about Free Speech? Findings form a Global Survey'*. Justitia.
7. Justitia's Free Speech Index. 2021. https://justitia-int.org/report-who-cares-about-free-speech-findings-from-a-global-survey-of-free-speech
8. Council of Europe Committee of Experts for the Development of Human Rights Report. 2007. Chapter IV, 123, para. 4.
9. Council of Europe's Committee of Ministers Recommendation 97 (20) on Hate Speech.
10. Uladzislau, Belavusau. 2013. Freedom of Speech: Importing European and US Constitutional Models in Transitional Democracies. Routledge, 41.
11. Mari, J., and Matsuda. 1987. 'Public Response to Racist Speech: Considering the Victim's Story'. 87 Michigan Law Review 8.
12. McGonagle, Tarlach. 2001. Wrestling Racial Equality from Tolerance of Hate Speech. *23 Dublin University Law Journal* 21: 4.
13. Claudia, E., Haupt. 2005. Regulating Hate Speech - Damned If You Do and Damned If You Don't: Lessons Learned from Comparing the German and U.S. Approaches. 23 *Boston University International Law Journal* 2: 304.
14. Roger, Kiska. 2012. 'Hate Speech: A Comparison Between The European Court of Human Rights and the United States Supreme Court Jurisprudence'. 25 *Regent University Law Review* 1: 110.