

EU: European Commission's Code of Conduct for Countering Illegal Hate Speech Online and the Framework Decision

June 2016

Table of contents

Introduction	3
Applicable international human rights standards.....	4
Compatibility of the EU Framework Decision with international law.....	7
Framework Decision and online communications.....	12
Analysis of the Code.....	14
The definition of “hate speech” is overbroad	15
Regulatory model	15
Failure to observe due process guarantees.....	16
Lack of civil society participation and consultation	16
Conclusion	18
About ARTICLE 19.....	19

Introduction

In this brief, ARTICLE 19 reviews the [Code of Conduct on Countering Illegal Hate Speech Online](#) (the Code of Conduct) that has been developed by the European Commission in collaboration with major information technology companies for its compliance with international standards on freedom of expression.

According to the European Commission, the Code of Conduct is the outcome of ongoing discussions between the European Commission and Facebook, Microsoft, Twitter and YouTube (the IT companies) as well as civil society.¹ It was initiated following the [EU Colloquium](#) on Fundamental Rights in October 2015 on ‘Tolerance and respect: preventing and combating Antisemitic and anti-Muslim hatred in Europe’ and the EU Internet Forum in December 2015. The Code of Conduct was published on 31 May 2016 in the wake of recent terrorist attacks in Europe and concerns among governments that social media is used by terrorist groups to ‘radicalise’ young people.

The Code of Conduct specifically refers to the [Framework Decision](#) on Combatting Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law (Framework Decision) as the legal basis for defining illegal hate speech under the Code.

In this analysis, ARTICLE 19 reviews both the legal basis for the Code of Conduct (i.e. the Framework Decision), the Code itself and the process that led to its adoption, under international standards on freedom of expression. Since the European Commission highlighted that the Code of Conduct is a part of series of approaches to address the problem of “online hate speech,” we hope that the Commission will use this analysis in its future activities in this area. We also urge the IT companies and other stakeholders to consider these recommendations in their cooperation with the European Commission and others.

¹ http://europa.eu/rapid/press-release_IP-16-1937_en.htm

Applicable international human rights standards

ARTICLE 19's comments on the Code of Conduct is informed by international human rights law, in particular the right to freedom of expression as protected by Articles 19 of the Universal Declaration of Human Rights, as well as Article 19 of the International Covenant on Civil and Political Rights (ICCPR).

While the right to freedom of expression is a fundamental right, it is not guaranteed in absolute terms. Under Article 19(3) of the ICCPR restrictions on the right to freedom of expression must be strictly and narrowly tailored and may not put in jeopardy the right itself. The determination whether a restriction is narrowly tailored is often articulated as a three-part test. Restrictions must:

- be provided by law, i.e. formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly;²
- pursue a legitimate aim as exhaustively enumerated in Article 19(3)(a) and (b) of the ICCPR; and
- conform to the strict tests of necessity and proportionality, i.e. if a less intrusive measure is capable of achieving the same purpose as a more restrictive one, the least restrictive measure must be applied.³

Further limitations on the right to freedom of expression are stipulated in Article 20 (2) of the ICCPR which requires states to prohibit “any *advocacy* of national, racial or religious hatred that constitutes *incitement* to discrimination, hostility or violence shall be prohibited by law.” Article 20 (2) of the ICCPR does not require States to prohibit all negative statements towards national groups, races and religions. However, States should be obliged to prohibit the advocacy of hatred that constitutes incitement to discrimination, hostility or violence. “Prohibition” allows three types of sanction: civil, administrative or, as a last resort, criminal.

The UN Rabat Plan of Action (Rabat Plan)⁴ has advanced a range of conclusions and recommendations for the implementation of Article 20(2) of the ICCPR; these correspond closely to the narrower political commitment of states in Resolution 16/18 to “criminalize incitement to imminent violence based on religion or belief. Namely, the prohibition in Article 20(2) of the ICCPR requires:

- *Conduct of the speaker*: the speaker must address a public audience, and their expression include:
 - advocacy
 - of hatred targeting a protected group based on protected characteristics,
 - constituting incitement to discrimination, hostility or violence.
- *Intent of the speaker*: the speaker must:
 - specifically intend to engage in advocacy of discriminatory hatred, and

² HR Committee, *L.M. de Groot v. The Netherlands*, No. 578/1994, UN Doc. CCPR/C/54/D/578/1994 (1995).

³ HR Committee, *Velichkin v. Belarus*, No. 1022/2001, UN Doc. CCPR/C/85/D/1022/2001 (2005).

⁴ [UN Rabat Plan of Action](#), 2012.

- intend for or have knowledge of the likelihood of the audience being incited to a discrimination, hostility or violence;
- *A likely and imminent danger of the audience actually being incited to a proscribed act*, as a consequence of the advocacy of hatred. A six-part “severity threshold” test, outlined below, assists in measuring whether the danger of incitement is sufficient to justify restrictions.
 - *Context*: analysis of the context should place the speech act within the social and political context prevalent at the time the speech was made and disseminated, for example the existence or history of conflict, existence or history of institutionalised discrimination, the legal framework and the media landscape;
 - *The identity of the speaker*: the position or status of the speaker in society should be considered, specifically the individual or organisation’s standing in the context of the audience to whom the speech is made and disseminated;
 - *The intent of the speaker*: it should be considered whether the speaker specifically intended to engage in the advocacy of hatred, to target a protected group, and for the proscribed outcome of discrimination, hostility or violence to actually occur;
 - *The content of the expression*: the words that were said and how they were said is critical, in particular with regard to what the audience understood by the content of the expression, and the form of the expression;⁵
 - *The extent or magnitude of the expression*: this includes elements such as the reach of the speech, its public nature, magnitude and the size of the audience;
 - *The likelihood and imminence* of discrimination, hostility or violence actually occurring: as an inchoate offence (where criminalised), there must be a reasonable probability that the speech would succeed in inciting actual action against the target group, recognising that such causation should be rather direct.

The six part test has been referenced positively by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression⁶ and the Committee on the Elimination of all forms of Racial Discrimination (the CERD Committee) has also based their guidance for respecting the obligation to prohibit certain forms of expression under Article 4 of the ICERD on this test.⁷

It should also be noted that there are some forms of “hate speech,” which may be understood as individually targeting an identifiable victim. This type of “hate speech” does not fit within

⁵ *Ibid.*, p. 35 – 36.

⁶ Report of the UN Special Rapporteur on the promotion and protection of the right of freedom of opinion and expression, A/67/357, 7 September 2012 (the 2012 Report of the SR on FOE), para. 45. The report cites an earlier draft of ARTICLE 19’s policy brief on incitement, prepared ahead of one of the regional expert meetings organised by OHCHR. Following these workshops the test was streamlined to six parts, which is also reflected in the Rabat Plan of Action.

⁷ UN Committee on the Elimination of Racial Discrimination (CERD), *General recommendation No. 35 : Combating racist hate speech*, 26 September 2013, paras. 15 - 16. The CERD Committee specifies that five contextual factors should be taken into account: the content and form of speech; the economic, social and political climate; the position or status of the speaker; the reach of the speech; and the objectives of the speech. The CERD Committee also specifies that States must also consider the intent of the speaker and the imminence and likelihood of harm.

the criteria of Article 20(2) of the ICCPR because the speaker does not seek to incite others to take an action against persons on the basis of a protected characteristic. These types of 'hate speech' include threats of violence, harassment and assault.

Importantly, expression may be inflammatory or offensive, but not meet any of the thresholds described above. This expression may be characterised by prejudice, and raise concerns over intolerance, but does not meet the threshold of severity, at which restrictions on expression are justified. This does not preclude States from taking legal and policy measures to tackle the underlying prejudices of which this category of 'hate speech' is symptomatic, or from maximising opportunities for all people, including public officials and institutions, to engage in counter-speech.

The same principles apply to electronic forms of communication or expression disseminated over the Internet.⁸ In particular, in General Comment No. 34, the HR Committee has said that:

43. Any restrictions on the operation of websites, blogs or any other Internet-based, electronic or other such information dissemination system, including systems to support such communication, such as Internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3. Permissible restrictions generally should be content-specific; generic bans on the operation of certain sites and systems are not compatible with paragraph 3. It is also inconsistent with paragraph 3 to prohibit a site or an information dissemination system from publishing material solely on the basis that it may be critical of the government or the political social system espoused by the government.⁹

While this does not necessarily imply that private companies cannot set the rules for the use of their services, the UN Special Rapporteur on freedom of expression has long held that censorship measures should never be delegated to private entities.¹⁰ In his June 2016 report to the UN Human Rights Council,¹¹ the UN Special Rapporteur on freedom of expression, David Kaye, enjoined States not to require or otherwise pressure the private sector to take steps that unnecessarily or disproportionately interfere with freedom of expression, whether through laws, policies, or extralegal means. In particular, he recommended that any demands, requests and other measures to take down digital content must be based on validly enacted law, subject to external and independent oversight, and demonstrate a necessary and proportionate means of achieving one or more aims under Article 19 (3) of the ICCPR.

Similarly, the four special mandates on freedom of expression recently reiterated in their 2016 [Joint Declaration on freedom of expression and countering violent extremism](#) that States should refrain from pressuring, punishing or rewarding intermediaries with the aim of restricting lawful content.

⁸ [General Comment 34](#), CCPR/C/GC/3, para 43.

⁹ HR Committee, Concluding Observations on the Syrian Arab Republic (CCPR/CO/84/SYR).

¹⁰ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 16 May 2011, [A/HRC/17/27](#), paras. 75-76.

¹¹ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 11 May 2016, [A/HRC/32/38](#).

Compatibility of the EU Framework Decision with international law

The Code of Conduct specifies that “illegal hate speech” should be understood as per the definition of this term under the Framework Decision and national laws transposing it (Framework Decision). Hence, ARTICLE 19 first wants to highlight the compatibility of the Framework Decision with international standards on freedom of expression.

Article 1(1)(a) of the Framework Decision requires the criminal prohibition of “publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin.” Article 1(1) limits the offence to intentional conduct, Article 1(1)(b) clarifies that the offence can be committed through the dissemination of any material, and Article 1(2) allows States to opt to punish “only conduct which is either carried out in a manner likely to disturb public order or which is threatening, abusive or insulting.” Article 3 prescribes “effective, proportionate and dissuasive criminal penalties,” with mandatory custodial sentences of between 1 and 3 years.

ARTICLE 19 finds that the offences in Article 1(1)(a) and (b) of the Framework Decision do not comply with international standards on the right to freedom of expression and information for the following reasons:

- **“Incitement to hatred” is too broad:** The offence of “incitement to hatred” in Article 1(1)(a) is formulated in terms that are too broad. It does not accurately transpose the obligation of States under Article 20(2) of the ICCPR or meet the three-part test for legitimate restrictions on the right to freedom of expression under Article 19(3) of the ICCPR.

Article 20(2) of the ICCPR only requires States to prohibit “incitement to discrimination, hostility or violence.” The “advocacy of hatred” is the vehicle for incitement, but “hatred” is not in itself a proscribed outcome. “Incitement to hatred” makes the proscribed outcome an emotional state or opinion, rather than the imminent and likely risk of a manifested action (discrimination, hostility, or violence). It should be noted that the right to freedom of opinion is an absolute right that cannot be qualified under any circumstances.

Thus, “incitement to hatred” does not meet the pressing social need that the obligation for restricting expression in Article 20(2) of the ICCPR addresses.¹² The subjective nature of “hatred” as a proscribed outcome also raises questions of legal certainty, and may open the door to arbitrary application. The circular definition given to the term “hatred” in the preamble to the Framework Decision is particularly concerning in this regard.¹³ Thus, the formulation of the offence under Article 1(1)(a) of the Framework Decision

¹² It should be noted that “incitement to hatred” is often used as shorthand for the obligation under Article 20(2) of the ICCPR. See, for example, the Rabat Plan of Action, *op. cit.*

¹³ The Framework Decision, preamble, paragraph 9, defines “hatred” as “referring to hatred based on race, colour, religion, descent or national or ethnic origin”. For recommended definitions of key terms in Article 20(2) of the ICCPR, see: the Camden Principles (at Principle 12) and the Rabat Plan of Action, *op. cit.*, para. 21. It should also be noted that “hatred” for the purposes of Article 20(2) of the ICCPR has a different meaning than “hatred” in the context of “hate crimes”, where the term is used to denote a bias motivation.

would probably not withstand scrutiny under international standards on freedom of expression.

- **Specific intent is not required:** A crucial and distinguishing element of incitement as prohibited by Article 20(2) of the ICCPR and Article 4(a) of the International Covenant on Elimination of Racial Discrimination (ICERD) is the intent of the speaker to incite others to discrimination, hostility or violence.

Article 1(1)(a) falls short of this standard as it fails to provide for specific intent in relation to each element of the offence. The use of the term “advocacy” in Article 20(2) of the ICCPR implies that negligence or recklessness are not sufficient for imposing sanctions.¹⁴ “Advocacy” also implies something more than intentional distribution or circulation, which is all that is seemingly required by Article 1(1)(b).¹⁵

- **Severity threshold is not specified:** The Framework Decision provides in its preamble that it is limited to combating “particularly serious” forms of racism and xenophobia.¹⁶ However, it provides little guidance to States on what is considered “particularly serious”, and how to reconcile these limitations with the right to freedom of expression (set out in Article 7).

Article 1(2) of the Framework Decision allows States to choose to limit the scope of the obligation to prohibit incitement to circumstances where a public order disturbance is likely, or where the language at issue is threatening, abusive or insulting. This reveals how broad the obligation is under Article 1(1)(a) of the Framework Decision for States that do not exercise this option. It allows States to restrict expression without regard to the likelihood of harm, the content of the expression or its impact on the audience.

- **There should be no specific prohibition on condoning, denying or grossly trivialising crimes:** Article 1(c) and (d) of the Framework Decision require States to criminalise “publicly condoning, denying or grossly trivialising” specific international crimes recognised under international humanitarian law “in a manner likely to incite to violence or hatred against” a protected group. While “incitement to genocide” must also be prohibited under international law, the threshold test for this offence is much higher than the one laid down in the Framework decision. We believe there are two reasons why the Framework Decision should dispense with the requirement for Member States to prohibit expression that condones, denies or grossly trivialises historical crimes:
 - First, specific offences prohibiting expression in respect of historical events do not meet a pressing social need, and are therefore not “necessary” restrictions on the right to freedom of expression. Any instance of incitement committed by way of condoning, denying or trivialising a crime committed against a protected group of people may, where necessary, be prosecuted through standalone provisions on incitement, or alternative provisions within the civil or administrative law. This should be done by reference to both Article 19(3) and Article 20(2) of the ICCPR, using the six-part test set out above.
 - Secondly, it is undesirable for States to interfere with the right to know and the search for historical truth by tasking itself with promoting or defending an

¹⁴ Rabat Plan of Action, *op. cit.*, para. 29(c).

¹⁵ Rabat Plan of Action, *op. cit.*, para. 21.

¹⁶ Framework Decision, preamble, para. 6.

established set of “historical facts”. It should be the role of free and open debate to establish historical truths, and not the role of States. Defending the right to freedom of expression is based on the importance of open discussion and the discovery of the truth, including the truth about historical personalities and events; prohibiting false arguments inevitably affects historical debate as well as the ability of historians to establish the truth. The existence of these laws in developed democracies can also embolden other countries to adopt analogous prohibitions that are abused to silence legitimate expression.¹⁷ Holding an individual criminally liable for denials of historical events amounts to an unacceptable restriction on the right to freedom of expression.

Removing Articles 1(1)(c) and (d) from the Framework Decision would find support in international human rights law and domestic jurisprudence. In General Comment No. 34, the HR Committee states “laws that penalise the expression of opinions about historical facts are incompatible with the obligations that the Covenant imposes on States parties in relation to the respect for freedom of opinion and expression.”¹⁸ This has been reiterated by the UN Special Rapporteur on the protection and promotion of the right to freedom of opinion and expression,¹⁹ who also calls on all States to repeal laws that prohibit discussion of historic events.²⁰

At the same time, the CERD Committee stress that public denials or attempts to justify crimes of genocide and crimes against humanity should only be prohibited by criminal law “provided that they clearly constitute incitement to racial violence or hatred.” They also endorse and reference the view of the HR Committee that “expression of opinions about historical facts should not be prohibited or punished.”²¹

The ECtHR has held that “it is an integral part of freedom of expression to seek historical truth.”²² In *Lehideux and Isorni v. France* the ECtHR noted that the “demands of pluralism, tolerance and broadmindedness” in a democratic society were such that a debate on matters of history must be permitted, despite the memories it might bring back of past sufferings and the controversial role of the Vichy regime in the Nazi Holocaust.²³ However, the jurisprudence of the ECtHR is not consistent in this regard.²⁴

The French Constitutional Court also found unconstitutional a law prohibiting the denial of the 1915 Armenian Genocide. The Court found that “by punishing anyone contesting the existence of ... crimes that legislators themselves recognised or qualified as such, legislators committed an unconstitutional attack on freedom of

¹⁷ See, ARTICLE 19, [Cambodia: Law Against Non-Recognition of the Crimes Committed During Democratic Kampuchea](#), ARTICLE 19, 27 June 2013; or [Rwanda: President should Pardon Agnès Uwimana Nkusi and Saidati Mukabibi](#), 11 April 2012.

¹⁸ HR Committee, General Comment No. 34, *op. cit.*, at para. 49

¹⁹ The 2012 Report of the SR on FOE, *op.cit.*, para. 55: “By demanding that writers, journalists and citizens give only a version of events that is approved by the Government, States are enabled to subjugate freedom of expression to official versions of events.”

²⁰ *Ibid.*, para. 78.

²¹ CERD Committee, General Recommendation No. 35, *op. cit.*, para. 14

²² ECtHR, *Chauvy and Others v. France*, Application No. 64915/01, 29 September 2004.

²³ ECtHR, *Lehideux and Isorni v. France*, Case No. 55/1997/839/1045, 23 September 1998, para. 55.

²⁴ See, for example, ECtHR, *Garaudy v. France*, App.No. 65831/01, 24 June 2003.

expression.”²⁵ The decision of the ECtHR in *Garaudy v. France*, should be read in light of this development.

- **Focus on criminal law:** as outlined above in the section on applicable standards, a variety of legal means should be used to prohibit incitement, primarily through the civil and administrative law; criminal sanctions should only be used exceptionally and as a last resort. Article 3 of the Framework Decision requires Member States to take necessary measures to ensure that offences concerning racism and xenophobia are punishable by “effective, proportionate and dissuasive” criminal penalties, prescribing a maximum of at least between 1 and 3 years of imprisonment. While Article 3 does not *require* mandatory custodial sentences, it does not recommend or suggest alternative or less severe sanctions.

ARTICLE 19 is concerned that the Framework Decision mandates the criminal prohibition of incitement, and seemingly prefers custodial penalties as sanctions. This potentially violates the principle of proportionality, as severe penalties are prescribed without requiring consideration of lesser sanctions in the criminal law or alternative modes of redress through civil or administrative law that would be less intrusive on the right to freedom of expression.²⁶

- *Criminalisation as an exceptional and last resort measure:* Article 20(2) of the ICCPR requires that States *prohibit by law* the advocacy of hatred that constitutes incitement to discrimination, hostility, or violence, it does not *require* criminalisation or that imprisonment be available as a sanction.

International standards indicate that criminal sanctions should be used to punish “only serious and extreme instances of incitement to hatred.”²⁷ The Venice Commission has affirmed that the use of criminal sanctions should be seen as “last resort measures to be applied in strictly justifiable situations, when no other means appears capable of achieving the desired protection of individual rights in the public interest,”²⁸ a principle also supported by the Rabat Plan of Action.²⁹ The Committee of Ministers for the Council of Europe emphasise that where criminal sanctions are imposed, there must be strict respect for the principle of proportionality, given that criminal sanctions generally constitute a serious interference with the right to freedom of expression.³⁰

In relation to Article 4(a) of the ICERD, which does *require* criminalisation of certain forms of racist expression, the CERD Committee has recently clarified that sanctions under this provision “should be reserved for serious cases, to be proven beyond reasonable doubt, while less serious cases should be addressed by means other than criminal law.”³¹

²⁵ French Constitutional Court, Decision no. 2012-647 DC of 28 February 2012

²⁶ The HR Committee have stated that restrictions on the right to freedom of expression “must be the least intrusive instrument amongst those which might achieve their protective function”, General Comment No. 34, *op. cit.*, para 34.

²⁷ The 2012 Report of the SR on FOE, *op. cit.*, para. 47, 2012. See also the Supreme Court of Canada, *R v. Keegstra*, [1990] 3 S.C.R. 697, 13 December 1990, ¶ 697 (Can.), para. 1.

²⁸ Venice Commission, Report on the Relationship Between Freedom of Expression and Freedom of Religion, 17 – 18 October 2008.

²⁹ The Rabat Plan of Action, *op. cit.*, para. 34.

³⁰ Recommendation No. R (97) 20 of the Council of Europe Committee of Ministers to Member States on “Hate Speech,” 30 October 1997, at Principle 5.

³¹ CERD Committee, General Recommendation No. 35, *op. cit.*, para. 12; see also: HR Committee, General

The Commission should also consider that criminal sanctions for “hate speech” may be counter-productive to the goal of effectively combating discrimination, particularly where that expression falls beneath the high threshold prescribed by Article 20(2) of the ICCPR.

The CERD Committee has noted “with concern” that broad or vague restrictions on freedom of speech have been used to the detriment of groups protected by the Convention, and stressed that measures to combat racist speech should not be used as a pretext to curtail expressions of protest at injustice, social discontent or opposition.³² This has been reiterated by the international mandates on freedom of expression in their 2006 Joint Declaration,³³ as well as by the UN Special Rapporteur on the protection and promotion of the right to freedom of opinion and expression.³⁴

The European Commission should also be mindful of the counter-productive impact “hate speech” prosecutions might have on the promotion of tolerance and respect for the rights of others. As the Council of Europe Committee of Ministers has warned, proponents of “hate speech” may exploit attempted criminal prosecutions to present themselves as “martyrs” or “victims”, or frame unsuccessful prosecutions as a vindication of their prejudicial views.³⁵ In both cases the prosecution serves as a platform to elevate the hate speaker’s views to a broader audience. The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression has echoed these concerns.³⁶

- *Incitement should be prohibited primarily through the civil and administrative law:* International standards foresee that, where effective, incitement should be sanctioned through a range of measures outside of the criminal law, principally through the civil and administrative law. We recall that the requirement of necessity means the least intrusive effective remedy should be employed when restricting speech to protect an overriding public or private interest.

Civil sanctions for incitement should include pecuniary and non-pecuniary damages, along with the right of correction and the right of reply.³⁷ The effectiveness of these sanctions is enhanced where civil society organisations are given standing to bring claims.³⁸ Some countries only prohibit incitement through the civil law,³⁹ and where

Comment No. 34, paras. 22- 25; 33-35.

³² *Ibid.*, para. 20.

³³ See, International Mechanisms for Freedom of Expression, Joint Declaration of 20 December 2006. The Special Mechanisms warned that “hate speech” legislation is often “abused by the powerful to limit non-traditional, dissenting, critical, or minority voices, or discussion about challenging social issues” and that the “resolution of tensions based on genuine cultural or religious differences cannot be achieved by suppressing the expression of differences but rather by debating them openly. Free speech is therefore a requirement for, and not an impediment to, tolerance.”

³⁴ The 2012 Report of the SR on FOE, *op.cit.*, para. 51.

³⁵ Council of Europe Committee of Ministers, Explanatory Memorandum to Recommendation No. R 97) (20), at Para. 16;

³⁶ The 2012 Report of the SR on FOE, *op.cit.*, para. 32.

³⁷ See Recommendation No. R (97) 20 of the Council of Europe Committee of Ministers to Member States on “Hate Speech,” 30 October 1997, at Principle 2; Rabat Plan of Action, *op. cit.*, para. 34; and Prohibiting incitement to discrimination, hostility or violence, ARTICLE 19, 2012, page 41 - 42.

³⁸ *Ibid.*

³⁹ Rabat Plan of Action, *op. cit.*, para. 16.

both criminal and civil sanctions are available, civil sanctions have been reported to be more effective.⁴⁰ Administrative sanctions and remedies should also be considered, including those identified and put in force by various professional and regulatory bodies.⁴¹

Framework Decision and online communications

The Framework Decision makes no provision on interpreting and implementing the obligations it contains in the context of online communications, giving no guidance to States on how to ensure fundamental rights including the right to freedom of expression should be protected in this context.

The UN Special Rapporteur on the protection and promotion of the right to freedom of opinion and expression has expressed concern that many attempts by States to tackle “hate speech” online have been misguided.⁴² These measures include requests for online intermediaries to screen or remove content, registration requirements to identify users’ real names, and the arbitrary blocking of websites.⁴³ The Rapporteur also points to the volume of content posted online everyday, and the cross-border nature of communications, as complicating implementation of any law on hate speech.⁴⁴

ARTICLE 19 believes that there should be specific guidance for safeguarding fundamental rights in the implementation of the Framework Decision in the online context. It is well established that the right to freedom of expression applies online.⁴⁵ Although there is no pressing social need for specific restrictions on content disseminated over the Internet,⁴⁶ enforcement mechanisms for existing restrictions must take into account the special characteristics of the Internet.⁴⁷

A series of principles should be emphasised to protect fundamental rights in relation to the enforcement of the Framework Decision in the online context:

- Self-regulation should be promoted as an effective tool for redressing harmful speech,⁴⁸
- Mandatory blocking of entire websites, IP addresses, ports, network protocols or types of uses (such as social networking) is an extreme measure – analogous to banning a newspaper or broadcaster – which can only be justified in accordance with international standards,⁴⁹

⁴⁰ For example, in Brazil it has been documented that sanctions have been levied more effectively through civil proceedings. See, Tanya Hernandez, “Hate Speech and the Language of Racism in Latin America”, 32 U. Pa. J. Int’l L. 805 2010 – 2011.

⁴¹ See Recommendation No. R (97) 20 of the Council of Europe Committee of Ministers to Member States on “Hate Speech,” 30 October 1997, Principle 2; Rabat Plan of Action, *op. cit.*, para. 34; and Prohibiting incitement to discrimination, hostility or violence, ARTICLE 19, 2012, p. 43.

⁴² The 2012 Report of the SR on FOE, *op.cit.*, para. 32.

⁴³ *Ibid.*

⁴⁴ *Ibid.*

⁴⁵ HR Committee, General Comment No. 34, *op. cit.*, at para. 11; UN Human Rights Council, Resolution 20 on “the promotion, protection and enjoyment of human rights on the Internet”, A/HRC/20/L.13, 29 June 2013.

⁴⁶ [Joint Declaration on Freedom of Expression and the Internet](#) 2011, para. 1(d).

⁴⁷ *Ibid.* See also HR Committee, General Comment No. 34, *op. cit.*, at paras 15 and 43.

⁴⁸ The 2011 Joint Declaration, *op.cit.*, para. 1(e).

⁴⁹ *Ibid.* See also Human Rights Committee, General Comment No. 34, *op. cit.*, para. 43.

- States should request the removal of content only through a court order and intermediaries should never be held liable for content of which they are not the authors;⁵⁰
- Decisions to block or compel the removal of content should be undertaken by a competent judicial authority or a body which is independent of any political, commercial, or other unwarranted influences to ensure that blocking is not used as a means of censorship;⁵¹
- The right of individuals to express themselves anonymously online must be fully guaranteed.⁵²

ARTICLE 19 has developed a number of policy documents that set out these principles in greater detail. These include *The Right to Blog* (2013),⁵³ and *Internet Intermediaries: Dilemma of Liability* (2013).⁵⁴

Moreover, we note that the Framework Decision not only fails to provide a clear benchmark in relation to the definition of ‘hate speech’, it also gives States a degree of flexibility in transposing its provisions in national law, including for determining what severity threshold speech should meet before being criminalised. Therefore, in practice, IT companies are encouraged to enforce via their Terms of Service widely different legal approaches to “hate speech” across Europe. Ultimately, we believe that this is likely to create more legal uncertainty for users and, most worryingly, lead to the application of the lowest common denominator when it comes to the definition of ‘hate speech’.

⁵⁰ The 2011 Joint Declaration, *op.cit.*, para. 2(a); the 2012 Report of the SR on FOE, *op.cit.*, para. 87.

⁵¹ The 2011 Report of the SR on FOE, *op.cit.*, para. 38.

⁵² The 2011 Joint Declaration, *op.cit.*, para. 6(d); the 2012 Report of the SR on FOE, *op.cit.*, para. 87.

⁵³ ARTICLE 19, [The Right to Blog](#).

⁵⁴ ARTICLE 19, [Internet Intermediaries: Dilemma of Liability](#), 2013.

Analysis of the Code

Under the Code of Conduct, the IT companies agree to take the lead on countering the spread of ‘illegal hate speech’ online by:

- Having in place effective mechanisms to review notifications regarding ‘illegal hate speech’ on their services so they can remove or disable access to such content;
- Having in place Community Guidelines clarifying that they prohibit the promotion of incitement to violence and “hateful” conduct.
- Reviewing the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary.

In particular, the Code of Conduct looks to strengthen notification processes between IT companies and law enforcement authorities by channelling communications between them through national contact points on both sides. The role of civil society organisations (CSOs) as “trusted reporters” of “illegal hate speech” is also highlighted, with the European Commission and Member States helping to ensure access to a representative network of CSO partners and “trusted reporters”.

The Code of Conduct contains further commitments from the IT companies to educate their users about the types of content not permitted under their rules and community guidelines, to share best practices between themselves and other social media platforms, and to continue working with the European Commission and CSOs on developing counter-narratives and counter-hate speech campaigns.

While the Code of Conduct does not put in place any mechanism to monitor the signatories’ compliance with it - and indeed is not binding or otherwise enforceable - the IT Companies and the European Commission agree to assess the public commitments in the Code on a regular basis. In addition, the European Commission, in coordination with Member States, commits to promote adherence to the commitments set out in the Code to other relevant platforms and social media companies.

The Code continues the [longstanding policy](#) of the European Commission to promote the adoption by the private sector of “voluntary” measures to tackle illegal content. It also reflects IT companies’ practices that have been in place for some time:

- Facebook, Twitter, Microsoft and Youtube have long had reporting or ‘flagging’ mechanisms in place.
- These companies have steadily “tweaked” their Community guidelines in the last year or so to reflect national legislation and concerns from Member States around hate speech and incitement to terrorism. Facebook updated its Terms of Service in [March 2015](#) and Microsoft in [May 2016](#). Twitter announced that it had been stepping up the enforcement of its rules in [February 2016](#).
- Companies such as Youtube or Facebook have been working with “trusted reporters” for some time, though these companies have so far not published any information about who these “trusted reporters” are and how they operate.

- That IT companies review removal notifications against their community guidelines and where necessary national laws, is also nothing new.

Therefore, in practice, it appears that the Code of Conduct is primarily making public and formalising certain aspects of the internal processes, which these IT companies already had in place to deal with complaints about certain types of content. One can only surmise that the commitment to review and remove “illegal hate speech” within 24 hours is an equally longstanding practice in these companies.

The definition of “hate speech” is overbroad

In ARTICLE 19's view, the Code of Conduct is particularly problematic as it encourages the removal of ‘illegal hate speech’ - and the ‘tweaking’ of Terms of Service - by reference to the EU's Framework Decision. As outlined above, we have serious concerns about the compatibility of the Framework Decision with international standards on freedom of expression. These problems will be further exacerbated if the IT companies rely on the Framework Decision as their assessment of prohibited expression will not meet international standards either.

Further, and in any event, the Code seems to encourage companies to go beyond the requirements of the Framework Decision since IT companies commit to make clear that they prohibit “hateful conduct”, i.e. a vague term that could encompass mere vulgar abuse.

In the current climate of fear over recent terrorist attacks, ARTICLE 19 is also concerned that “hate speech” is increasingly being conflated with “terrorist” and “extremist” content, concepts that are equally broad and often framed in vague and ambiguous terms.

Regulatory model

It appears that the Code of Conduct was initiated by the European Commission and was developed in consultations with the ICT companies. This resembles the co-regulatory model of codes of conduct (or broadcasting codes) typically applied for the broadcast media. ARTICLE 19 believes that the type regulatory model promoted by the European Commission is not only at odds with international standards on freedom of expression, it is also deeply inappropriate. Whilst the Code is non-binding and lacks any clear means of enforcing it, it is nonetheless the product of efforts by the European Commission and Member States to encourage private companies to censor content, which may well be lawful. In other words, it is a form of regulation-lite.

We also recall that the self-regulation of the content in the print media is generally accepted as the least restrictive means for the right to freedom of expression under international law. By contrast, international law permits greater regulation for broadcasters, recognising that the supply of broadcasting frequencies is limited and that a mechanism is necessary to allocate them across different operators, in order to create a diverse and pluralistic broadcasting landscape.

The Internet may be seen as occupying an intermediate position between the print media and broadcasting. In common with broadcasting, every operator on the internet requires the use of one or more unique addresses, similar to the frequencies of a radio or television station. But in common with the print media, there are no limits to the numbers of internet websites that can exist alongside each other. Hence, ARTICLE 19 believes that state regulation in this area

is undesirable and that self-regulation should be preferred instead.⁵⁵ This is also consistent with the 2011 Joint Declaration on freedom of expression and the Internet, in which the four special mandates on freedom of expression said:

- a. Greater attention should be given to developing alternative, tailored approaches, which are adapted to the unique characteristics of the Internet, for responding to illegal content, while recognising that no special content restrictions should be established for material disseminated over the Internet.
- b. Self-regulation can be an effective tool in redressing harmful speech, and should be promoted.

Failure to observe due process guarantees

Additionally, the Code of Conduct is problematic from a due process point of view. In particular:

- It puts companies - rather than the courts - in the position of having to decide the legality of content;
- It allows law enforcement to pressure companies to remove content in circumstances where the authorities do not have the power to order its removal because the content itself is legal;
- It deprives Internet users of a remedy to challenge wrongful removals since the vast majority of content removals are likely to be made on the basis of the company's terms of service.

The Code of Conduct also seems to indicate that the resources of law enforcement are increasingly devoted to the removal of content, such as “hate speech”, rather than the investigation and prosecution of those responsible for the allegedly unlawful conduct. In other words, States seem more concerned about the (in)accessibility of content rather than enforcing the law. While in some circumstances, the removal of content may be a more proportionate alternative than criminal liability, nonetheless, it signals governments' propensity to promote censorship rather than seeking to address the root causes of the social problems at issue. In practice, it is also likely to be counter-productive as it gives an incentive to individuals engaging in “hate speech” to migrate to other platforms with higher free speech standards. In the case of suspected terrorists, this is likely to lead to a whack-a-mole game as companies suspend “terrorist” accounts only to see new supporters create new profiles.

Lack of civil society participation and consultation

Finally, ARTICLE 19 is concerned that despite several references to civil society organisations (CSOs') in the Code of Conduct - and insofar as this form of government intervention may be inevitable - there was apparently no involvement of CSOs defending freedom of expression in the drafting of the Code of Conduct. We note, for instance, that [EDRI and AccessNow](#) walked out of discussions on the Code due to the lack of transparency of ongoing negotiations that led to its adoption. They have also denounced the lack of transparency of the negotiations

⁵⁵ Consistent with our policy on [Internet intermediaries: Dilemma of Liability](#) and our forthcoming policy on the Terms and Conditions of major Internet companies.

taking place at the EU Internet Forum on countering terrorism online. In our view, this severely undermines the credibility of the Code.

We are further concerned that, insofar as the Code promotes cooperation with “trusted reporters” or ‘CSOs’, no reference has been made to the need to ensure that free expression groups are consulted. Indeed, the Code fails to make it clear that any restriction on free expression should remain the exception rather than the rule. Instead, the Code only contains weak references to the protection of freedom of expression.

Conclusion

ARTICLE 19 is deeply concerned that, despite its non-binding character, the Code will lead to more censorship by private companies - and therefore a chilling effect on freedom of expression on the platforms they run. This is especially so in the absence of any independent or meaningful commitment to protect freedom of expression.

The Code of Conduct is likely to be trumpeted by governments and companies alike as a milestone in the fight against 'illegal hate speech'. ARTICLE 19 believes however that it is misguided policy on the part of governments, one that undermines the rule of law. For companies, it is likely to amount to no more than a public relations exercise. In the meantime, freedom of expression online is likely to be greatly diminished.

We therefore recommend the European Commission and the IT companies to revisit this initiative and make sure it complies with the international standards on freedom of expression. For the same reasons, we also urge the European Commission to revise the Framework Decision.

About ARTICLE 19

ARTICLE 19 advocates for the development of progressive standards on freedom of expression and freedom of information at the international and regional levels, and their implementation in domestic legal systems. The Law Programme has produced a number of standard-setting publications which outline international and comparative law and best practice in areas such as defamation law, access to information and broadcast regulation.

On the basis of these publications and ARTICLE 19's overall legal expertise, the organisation publishes a number of legal analyses each year, comments on legislative proposals as well as existing laws that affect the right to freedom of expression. This analytical work, carried out since 1998 as a means of supporting positive law reform efforts worldwide, frequently leads to substantial improvements in proposed or existing domestic legislation. All of our analyses are available at <http://www.article19.org/resources.php/legal>.

If you would like to discuss this analysis further, or if you have a matter you would like to bring to the attention of the ARTICLE 19 Law Programme, you can contact us by e-mail at legal@article19.org.