Legal perspectives on online hate speech in Europe

Recent developments and open challenges

Pietro Dunn

PhD candidate @ University of Bologna & University of Luxembourg – pietro.dunn2@unibo.it

Academic Fellow @ Bocconi University, Milan – pietro.dunn@unibocconi.it

FASTLISA

Defining hate speech

Hate speech for the purpose of the Recommendation entails the use of one or more particular forms of expression – namely, the advocacy, promotion or incitement of the denigration, hatred or vilification of a person or group of persons, as well any harassment, insult, negative stereotyping, stigmatization or threat of such person or persons and any justification of all these forms of expression – that is based on a non-exhaustive list of personal characteristics or status that includes “race”, colour, language, religion or belief, nationality or national or ethnic origin, as well as descent, age, disability, sex, gender, gender identity and sexual orientation.

ECRI, General Policy Recommendation No. 15 (2016)

FASTLISA

Rationale(s) for hate speech regulation

Prevention of crime, violence, discrimination

Protection of individuals from psycho-physical harm

Protection of targeted groups' dignity

FASTLISA

The International Law Framework

INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS (ICCPR)

Article 20(2): “Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law”.

INTERNATIONAL CONVENTION ON THE ELIMINATION OF ALL FORMS OF RACIAL DISCRIMINATION (ICERD)

Article 4(a): “States Parties … shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as 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; …”

Some characters of online hate speech

PERMANENCE

- Ability of hateful content to thrive online and spread, also thanks to, e.g., hyperlinking tools
- It may depend on the architecture of the platforms themselves
- Enhances the harm inflicted on the targeted persons by making it more difficult to remove hate speech contents
- Helps the development of “cyber-cesspools” (Leiter, 2010)

ITINERANCY

- Ability of online content to be easily moved across the cyber-space
- This way, even when hate speech content is removed, it may be reposted easily elsewhere
- Intertwined with permanence, contributes to making it easier for poorly-formulated content, which would probably not have been capable of reaching public audiences in the past, to spread

Some characters of online hate speech

ANONYMITY

- Component of online freedom of expression (cf. ECtHR, *Standard Verlagsgesellschaft v Austria*, 2021)
- However, issue for law enforcement
- Even more, it contributes to people's disinhibition on the Internet (especially for low-profile hate speech)
- Anonymity + physical distance contributes to the dehumanization of targets and thus to "moral disengagement"

CROSS-JURISDICTIONAL NATURE

- Enhances the harm of hate speech
- Issue for international co-operation
  - Co-operation within the EU?
  - Co-operation between the EU and third countries (US)?

FASTLISA

Algorithmic content moderation & curation and hate speech

- Content moderation (*stricto sensu*): removal of hate speech content, suspension or deletion of accounts
  - Probabilistic tools: inevitable margin of error
  - Higher false positives v higher false negatives (freedom of expression v fight against hate speech)?
  - The danger of biased algorithms and the impact on victimised categories
  - Content moderation (*stricto sensu*): removal of hate speech content, suspension or deletion of accounts
- Content curation: design & presentation of contents, recommender systems etc. to improve individuals' experiences (and engagement)
  - Controversial content like hate speech generates engagement
  - The risk of shadowbanning

FASTLISA

New approaches to speech regulation

- Balkin: distinction between old-school and new-school forms of speech regulation
  - «Old-school speech regulation»: focuses on sanctioning individuals for their illegal utterances
  - «New-school speech regulation»: focuses on enforcing moderation and/or curation duties vis-*à*-*vis* online platforms
- European strategies are increasingly moving towards new-school approaches

FASTLISA

The ECHR Framework on Hate Speech

Condemning hate speech?

“Tolerance and respect for the equal dignity of all human beings constitute the foundations of a democratic, pluralistic society. That being so, as a matter of principle it may be considered necessary in certain democratic societies to sanction or even prevent all forms of expression which spread, incite, promote or justify hatred based on intolerance (including religious intolerance)”

Gunduz v Turkey (2003)

FASTLISA

Condemning hate speech?

“The Court reiterates its finding that comments that amount to hate speech and incitement to violence, and are thus clearly unlawful on their face, may in principle require the States to take certain positive measures … It has likewise held that inciting hatred does not necessarily entail a call for an act of violence or other criminal acts. Attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for the authorities to favour combating racist speech in the face of freedom of expression exercised in an irresponsible manner”

Beizaras and Levickas v Lithuania (2020)
[Cf. *Valaitis c Lithuania*, 2023]

FASTLISA

The “two-tiered” approach of the ECtHR

ARTICLE 10 ECHR

- Need to evaluate if the measure is consistent with the conditions set out in Article 10(2):
  - Lawfulness of the interference;
  - Legitimate aim;
  - “Necessary in a democratic society” (proportionality)
- General application

ARTICLE 17 ECHR

- The utterance is considered to be an abuse of freedom of expression: no protection of Article 10 ECHR
- Often applied in cases of:
  - Holocaust denial
  - Islamophobia
  - Antisemitism

Hate speech and intermediary liability

- *Delfi AS v. Estonia* (2015)
  - Imposing an online news portal to pay damages for having failed to promptly remove defamatory comments posted by anonymous users does not amount to a violation of right to freedom of expression entrusted to Article 10 of the ECHR
- Subsequent case law (*MTE v. Hungary*, *Pihl v. Sweeden*, *Høiness v. Norway*) upheld *Delfi* but came to different conclusions
  - Factors taken into account include severity of the comments, the categorization of the ISP as a business activity etc.
- BUT: hate speech as an exceptional case?

FASTLISA

The EU Framework on Hate Speech

EU Framework Decision 2008/913/JHA

- Only racist and religious hate speech is subjected to EU harmonisation under criminal law
- The EU generally has no juridisction in matters of criminal law
  - However, some forms of anti-hate speech law exist with respect to other fields of the law (e.g., AVMSD)
- Legal base for Framework Decision 2008/913/JHA is Article 67(3) TFEU
  - The Union shall endeavour to ensure a high level of security through measures to prevent and combat crime, racism and xenophobia, and through measures for coordination and cooperation between police and judicial authorities and other competent authorities, as well as through the mutual recognition of judgments in criminal matters and, if necessary, through the approximation of criminal laws

FASTLISA

EU Code of Conduct on Illegal Hate Speech

- Code of conduct on countering illegal hate speech online adopted on 31 May 2016
- However, some European countries subsequently stated their intentions to introduce regulation concerning hate speech
- Notion of hate speech refers to Framework Decision 2008/913/JHA
- In fact, the impact assessments have revealed a partial failure of the Code

FASTLISA

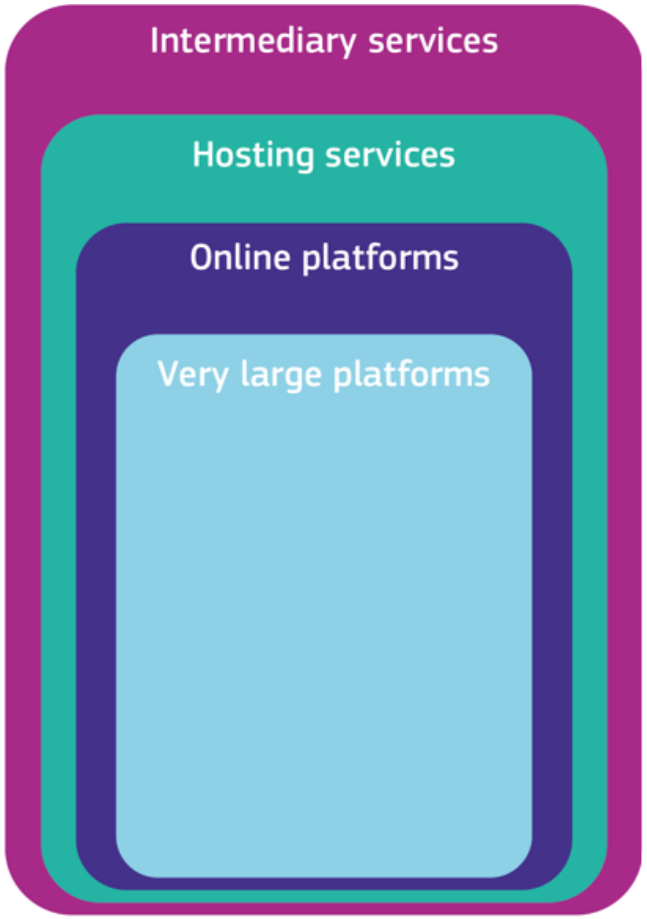
The Audiovisual Media Services Directive

- Directive 2010/13/EU, amended by Directive (EU) 2018/1808: new obligations for providers of video-sharing platforms
- Article 28b: «[…]Member States shall ensure that video-sharing platform providers under their jurisdiction take appropriate measures to protect: […]
  (b) the general public from programmes, user-generated videos and audiovisual commercial communications containing incitement to violence or hatred directed against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the Charter;
  (c) the general public from programmes, user-generated videos and audiovisual commercial communications containing content the dissemination of which constitutes an activity which is a criminal offence under Union law, namely […] and offences concerning racism and xenophobia as set out in Article 1 of Framework Decision 2008/913/JHA.

FASTLISA

The Digital Services Act

- Horizontal approach to content moderation
- Same liability exemptions for host providers as the e-Commerce Directive
- New “due diligence obligations for a transparent and safe online environment”: four levels of duties depending on the size and nature of the provider of intermediary services (asymmetric, risk-based approach)



FASTLISA

Terms and conditions (Art. 14)

- Obligation to provide clear, plain, intelligible, user-friendly and unambiguous terms and conditions concerning illegal content and moderation practices
- Application: necessary to give due regards to all legitimate interests and fundamental rights, including freedom of expression, as protected by the CFREU

Notice and action mechanisms (Art. 16)

- Provision applicable to all providers of hosting services
- Obligation to put in place mechanisms to allow individuals and entities to notify the presence of potentially illegal content
- Such notification will be considered sufficient to trigger the provider's liability for the content under Article 6
- In the case of online platforms and VLOPS, notifications from so-called "trusted flaggers" will need to be processed immediately (Art. 22)

FASTLISA

Risk assessment and mitigation (Arts. 34-35)

- Provisions applicable to VLOPs
- Obligation to operate yearly assessments of systemic risks connected to their services (e.g. dissemination of illegal content or negative effects on fundamental rights and freedoms) (Art. 34)
- Duty to put in place mitigation measures to reduce such risks, for example by adapting their AI tools for content moderation (Art. 35)

FASTLISA

Safeguards for individuals against over-removal

TRANSPARENCY REQUIREMENTS

- Art. 14: terms and conditions must be clearly presented to the users
- Art. 15: obligation for all providers to publish regular transparency reports concerning moderation practices
- Art. 17: need for a ‘statement of reasons’ concerning restrictions to users operated by host providers
- Additional transparency requirements for online platforms and VLOPs

PROCEDURAL SAFEGUARDS

- Art. 14: terms and conditions must be enforced in a diligent, objective and proportionate manner, with due regard to the rights and legitimate interests of all paeries, including freedom of expression, freedom and pluralism of the press, other rights set by the Nice Charter
- Art. 20: obligation for online platforms and VLOPs to introduce a complaint-handling system, which cannot be decided solely based on automated procedures, and with the possibility for the user to contact a human interlocutor

FASTLISA

Codes of conduct (Art. 45)

- The DSA follows a co-regulatory strategy by recognizing the possibility to adopt voluntary codes of conduct at Union level
- Implementation of these codes serves as evidence for providers to prove their compliance with the DSA; failure to comply with the codes serves as evidence against them
- See, e.g., 2022 Strengthened Code of Practice on Disinformation

The notion of illegal content under the DSA

- The DSA does not define what **illegal content** is
- In this respect, it must therefore be integrated by EU law as well as by national laws (when compliant with EU law!) concerning what is to be identified as illegal content
- The EU is deploying a wide array of sectoral (vertical) laws harmonising content moderation governance across the Union (e.g., DSM Copyright Directive, AVMS Directive Refit, TCO Regulation, CSAM Regulation proposal)
- **What about hate speech?**

FASTLISA

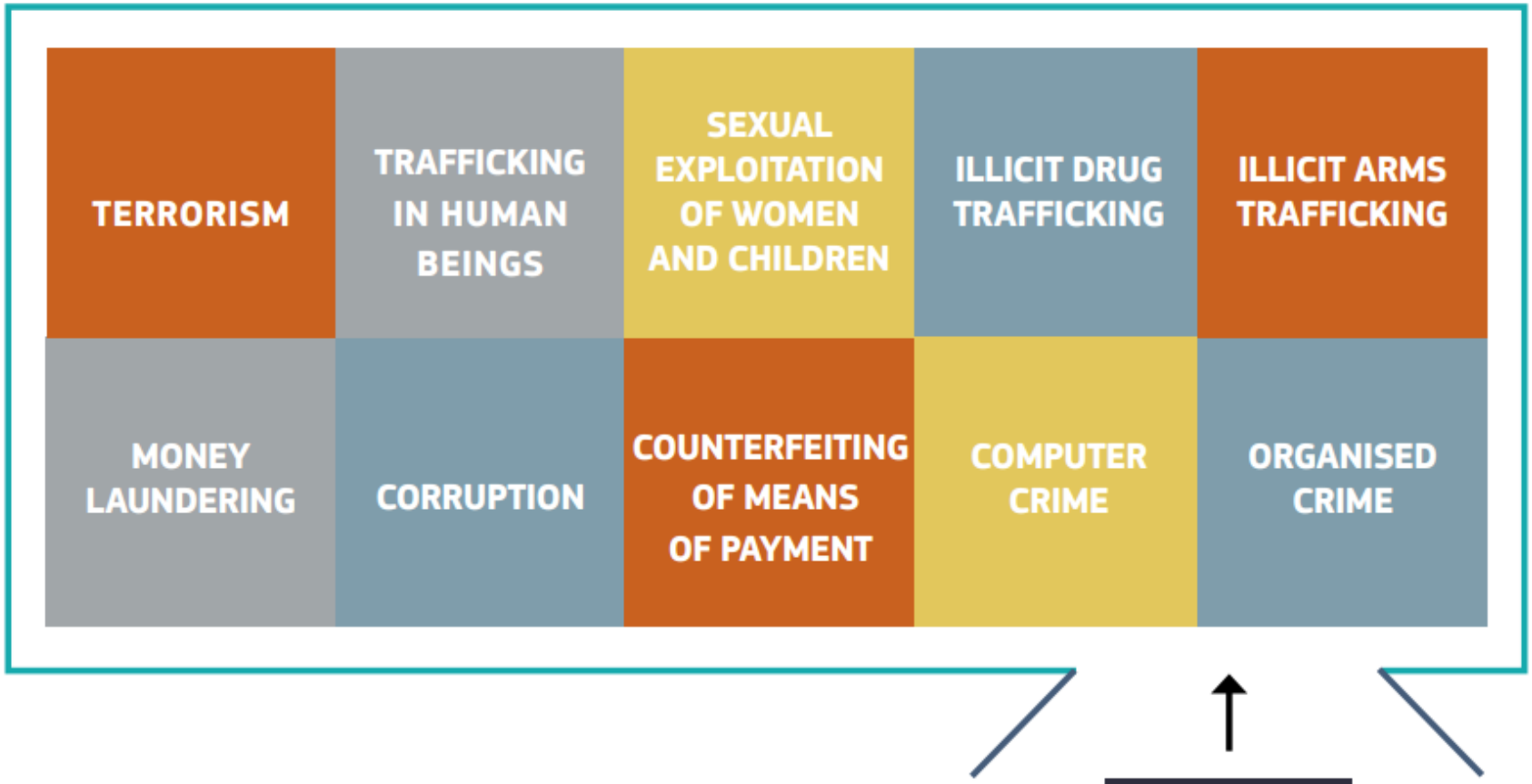
The Commission’s Proposal to Extend Article 83(1) TFEU

•1. The European Parliament and the Council may, by means of directives adopted in accordance with the ordinary legislative procedure, establish minimum rules concerning the definition of criminal offences and sanctions in the areas of particularly serious crime with a cross-border dimension resulting from the nature or impact of such offences or from a special need to combat them on a common basis.

•These areas of crime are the following: terrorism, trafficking in human beings and sexual exploitation of women and children, illicit drug trafficking, illicit arms trafficking, money laundering, corruption, counterfeiting of means of payment, computer crime and organised crime.

•On the basis of developments in crime, the Council may adopt a decision identifying other areas of crime that meet the criteria specified in this paragraph. It shall act unanimously after obtaining the consent of the European Parliament.

FASTLISA



To extend the list of the EU crimes a two-step process is needed:

1) the Council decides to extend the list with the European Parliament's consent
2) the Commission makes a legislative proposal.

FASTLISA

Q&A

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
DIPARTIMENTO DI SCIENZE GIURIDICHE

UAB
Universitat
Autònoma
de Barcelona

InfAI®
Institute for Applied Informatics

Comune di **Ravenna**

Ajuntament
de Santa Coloma
de Gramenet

PRO ARBEIT
Kreis Offenbach
Kommunales Jobcenter

1506
UNIVERSITÀ
DEGLI STUDI
DI URBINO
CARLO BO

tree.
communication & media

FASTLISA