



Addressing hate speech in the media: the role of regulatory authorities and the judiciary

INTERNATIONAL CONFERENCE

Organised by the Council of
Europe in partnership with
the Croatian Agency for
Electronic Media

6-7 November 2018
Zagreb, Croatia

Hate speech, fake news and AI

Elfa Ýr Gylfadóttir
Media Commission, Iceland



FJÖLMÍÐLANEFND

Hate speech, fake news and AI - The role of the national regulatory authorities in a digital environment

Elfa Ýr Gylfadóttir

6. November 2018

Overview

- Introduction – digital technology and social media
- Hate speech and other forms of offensive language in a digital environment
- AVMS Directive: co-regulation is encouraged to protect the public from incitement to violence and hatred on video-sharing platforms
- What are the Platforms (Facebook, YouTube etc.) doing in tackling information disorder and hate speech?
- The challenges for NRAs?



Digital technology and social media

- Communication does not just consist of distributing correct information to people. When communicating, certain values and narratives are also being mediated. (James Carey)
- Information having an emotional impact is more likely to be circulated. Research shows that emotions such as anger and fear have a specially strong impact on people. Also evoking the feeling that a group is superior to another group.
- Information which connects people to its "tribe" (**We** against **Others**)



The magical formula behind a ”successful” narrative

- **Evokes strong emotional response**
- **Is visually appealing**
- **Has a strong narrative**
- **Is repeated over and over again...**

People who are circulating information – correct or wrong – know this magical formula

It is more likely that images and videos are circulated in the algorithms of social media than texts

A new MIT research shows that fake news circulate **six times** more likely to be distributed than correct information and its impact is deeper and lasts longer - <http://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>

AVMSD on video-sharing platforms

Member States should ensure that video-sharing platforms put in place measures to:

- protect the general public from incitement to violence or hatred and content constituting criminal offences (public provocation to commit terrorist offences, child pornography and racism or xenophobia)
- Implementation of the new regime via co-regulation will be encouraged: the proposed rules provide basic requirements and partners who share responsibility and contribute to fulfilling the objectives
- Services such as YouTube will fall under the scope of the revised Directive. Audiovisual content shared on social media services, such as Facebook, will also be covered by the revised Directive.



Challenges linked to social media and video sharing platforms

Google



You Tube



FJÖLMÍDLANEFND

What is Facebook doing?

„**Dark posts**“ are ads that do not show up on advertisers' Facebook sites. Advertisers can buy different ads for different target groups

“**A/B testing**“ Different versions of the same ad are distributed to learn about which ones are the most engaging by the users. (The Trumps team tried out 40-50.000 different versions of ads/ posts every day before the presidential elections)

„**FB Learner Flow**“ An AI product which is used to predict what users will do or buy in the future. Gives advertisers an opportunity to intervene and influence users before they change their minds or swap to another product



“FakeTube” propaganda (video sharing platforms?)

- A new research published by Jonathan Albright at Columbia university shows how fake videos are being distributed in a massive scale on YouTube by using AI
- By the means of using fake news, „A/ B testing“ and personality tests, new propaganda forms are being developed. They are circulating fast in social media and other media
- AI is increasingly used to make YouTube videos. They are in a vast majority fake news or “news” circulated for propaganda purposes
- AI is used to detect the most current news on Facebook and Twitter. AI is used to find video content and pair it with texts and a computerised voice over. Some of those videos include hate speech
- One „FakeTube” factory in Albrights research produced and uploaded 80.000 videos in 19 different YouTube channels in just a few days
- According to Albright’s reserach – such factories are uplodaina AI generated fake videos every 3 minutes on YouTube



Deep Fake



Technology is emerging fast which can change sounds and images in real time

- What should we believe when we cannot trust our senses?



FJÖLMIDLANEFND

What are the challenges?

- Professional media has financial challenges
- Troll factories trying to mislead and distort democratic discussions
- The use of bots and AI to distribute fake news / propaganda / hate speech
- Echo chambers and filter bubbles. Media and social media usage that does not contradict our beliefs and is manifesting values and world views
- People with views that are not generally accepted get empowered on the internet. They become part of a “community” or a “tribe”. Hate speech is not challenged



Challenges for NRAs in a digital environment

- Crossborder content – jurisdictional problems?
- Are traditional regulatory frameworks effective in a new environment?
- Press councils and advertising self-regulatory standards. Are self- and co-regulation effective means for regulating social media / audio-visual platforms?
- What is at stake for democracy if video-sharing platforms do not comply?
- Do all platforms want to comply? (GAB?)
- Technical solutions. Are algorithms and AI necessary to detect hate speech and information disorder?
- Increased co-operation between European NRAs is needed



FJÖLMÍÐLANEFND

Thank you !

www.mediacommission.is