

**LEGAL AND POLICY  
FRAMEWORKS TO  
COMBATE SEXIST  
HATE SPEECH**

# HATE SPEECH IN HUMAN RIGHTS LAW

- Freedom of expression – Article 10 ECHR (Art 19 ICCPR)
  - The right to hold opinions and to receive and impart information and ideas without interference
  - Can be subject to restrictions that are prescribed by law and necessary in a democratic society for (inter alia) the protection of the rights of others
- Rights of others: equality and non-discrimination (Art 14 ECHR)

# HATE SPEECH IN HUMAN RIGHTS LAW

- Article 17 ECHR
  - Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention.
  - Eg. Publication of anti-semitic material contravened Article 17 and the applicant was not able to complain of a freedom of expression violation: “such a general, vehement attack on one ethnic group is directed against the Convention’s underlying values, notably tolerance, social peace and non-discrimination”
- Liability of internet intermediaries
  - *Delfi v Estonia (2015)*: the rights and interests of others and of society as a whole may entitle Contracting States to impose liability on Internet news portals, without contravening Article 10 of the Convention, if they fail to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties.



**HATE  
SPEECH**

Intense and irrational emotion of opprobrium, enmity and detestation towards an individual or group.

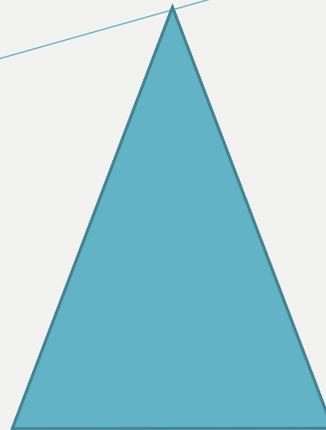
Any expression of hate towards an individual or group defined by a protected characteristic.

Any expression imparting opinions or ideas – bringing an internal opinion or idea to an external audience. It can take many forms: written, non-verbal, visual, artistic, etc, and may be disseminated through any media, including internet, print, radio, or television.

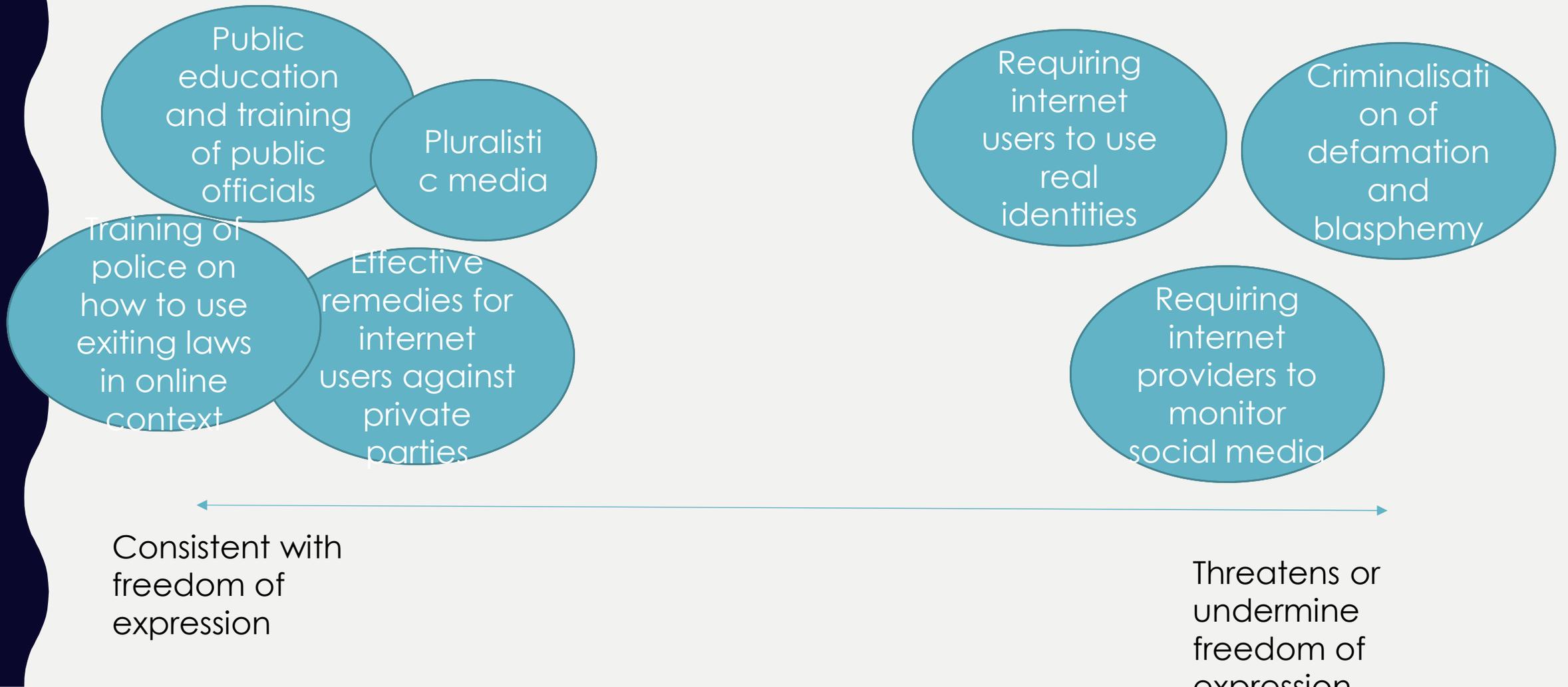
# RESPONDING TO HATE SPEECH: A BALANCING ACT

Right to  
equality,  
freedom from  
discrimination  
and violence,  
dignity

Freedom  
of  
expression



# THE IMPACT OF ANTI-HATE SPEECH LAWS & POLICIES ON FREE SPEECH



## CRIMINAL LIABILITY FOR HATE SPEECH

- Greece
  - Public incitement to hatred on basis of sexual orientation or gender identity
- Netherlands
  - Criminal liability for incitement to hatred, discrimination or violence on basis of sex or sexual orientation
  - Criminal liability dissemination or distribution of expression inciting to hatred on basis of sex or sexual orientation
  - Covers online crime
- Sweden
  - Criminal liability for incitement to hatred, discrimination or violence on basis of sex or sexual orientation

## LEGAL REMEDIES FOR SEXIST HATE SPEECH

## CYBERSTALKING LAWS

- Netherlands
  - Article 285b of the Penal Code: it is an offense to systematically infringe another person's privacy with the intention of compelling the other person to do something, not to do something or to intimidate the victim
- UK
  - Malicious Communications Act 1998, Offence of Sending Letters etc. with Intent to Cause Distress or Anxiety: it is an offense to send an indecent, offensive or threatening letter, electronic communication or other article to another person.
- Germany
  - Cyber-stalking accounts for 40% of all stalking cases in Germany. Offenders can be held criminally liable according to Sections 185, 186, or 187 of the Penal Code.
- Malta
  - Criminal Code, Article 251A Harassment and 251B

## LEGAL REMEDIES FOR SEXIST HATE SPEECH

## REVENGE PORN LAWS

- France
  - Penal Code Article 226: forbids the willful violation of the intimate private life of another by "transmitting the picture of a person who is within a private place, without the consent of the person concerned"
- UK
  - Criminal Justice and Courts Act section 33: offence to disclose private sexual photographs and films without the consent of the individual depicted and with the intent to cause distress.

## LEGAL REMEDIES FOR SEXIST HATE SPEECH

# NEW MODELS OF LEGAL REMEDIES

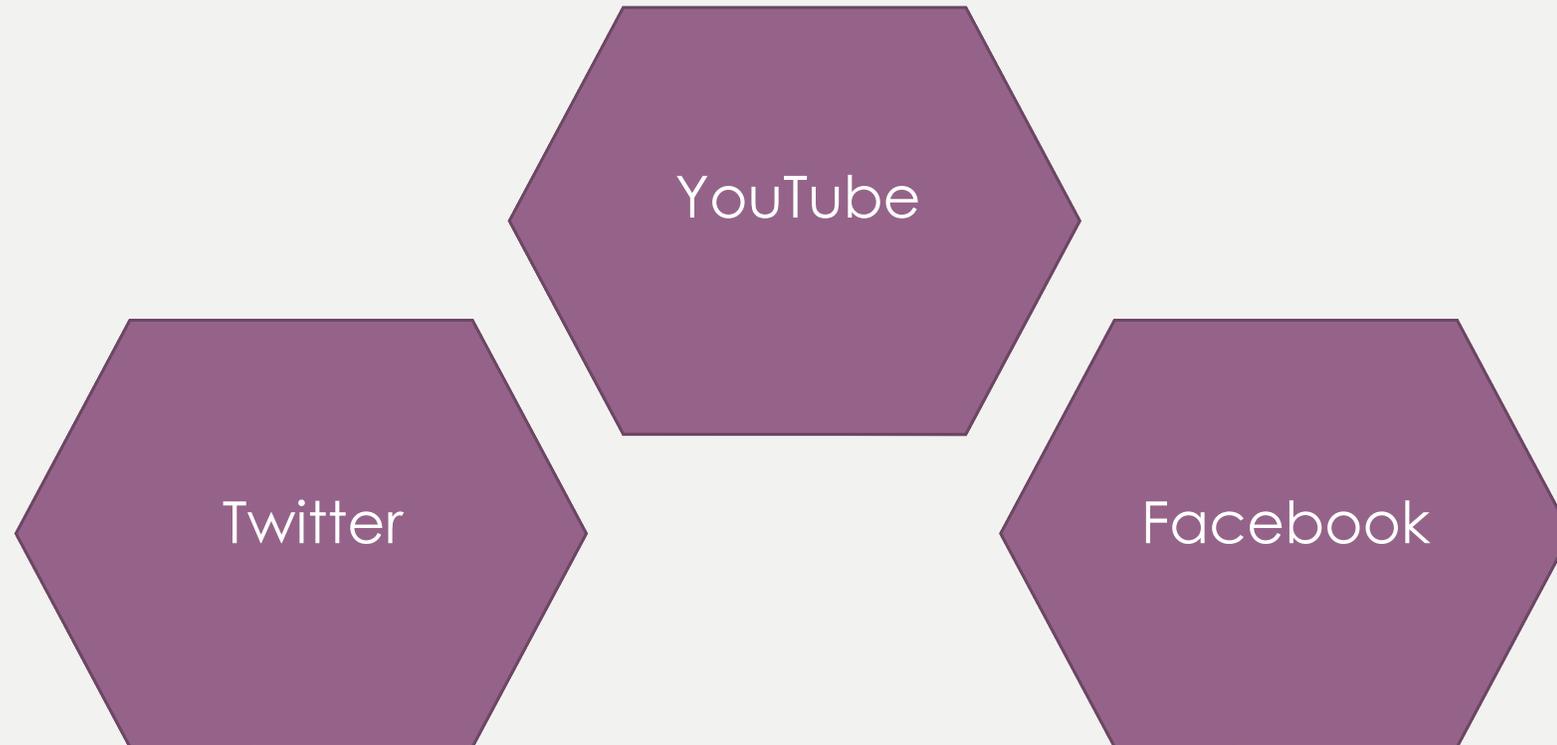
## CANADA – CYBER SAFETY ACT, NOVA SCOTIA

- Individuals subject to cyberbullying (or, in the case of minors, their parents) can apply to a Justice for a protection order
- Definition of cyberbullying: is intended or ought reasonably be expected to cause fear, intimidation, humiliation, distress or other damage or harm to another person's health, emotional well-being, self-esteem or reputation
- Protection orders can require disconnection of internet service
- Also creates the tort of cyberbullying

## US – REVENGE PORN LAWS CALIFORNIA

- SB 255 Electronic Communication Devices: Prohibited Distribution of Personal Information
- Misdemeanour of disorderly conduct by way of distribution of intimate photographs with the intent to cause serious emotional distress
- Amendment added by ACLU: under circumstances where the parties agree or understand that the image shall remain private

# POLICY FRAMEWORKS OF MAJOR INTERNET PROVIDERS



# REDRESS FOR SEXIST HATE SPEECH – FACEBOOK

- User policies recognise that users have a right not to be bullied, harrassed, subject to hate speech, threatened with violence
- Facebook does not permit hate speech, but distinguishes between serious and humorous hate speech
- Reporting can be done at the source of the content (with a 'Report' button)
- The reportee is warned, and the reporting user's identity is concealed
- Reportee may be blocked or have account disabled
- Complaints regarding violence or harmful behaviour against women must qualify as either Graphic Violence or Credible Threat of Violence. Once they are reported to Facebook, the reportee will be blocked from Facebook after the complaint is

# REDRESS FOR SEXIST HATE SPEECH – FACEBOOK

- Reporting can be done at the source of the content (with a 'Report' button)
- The reportee is warned, and the reporting user's identity is concealed
- Reportee may be blocked or have account disabled
- Complaints regarding violence or harmful behaviour against women must qualify as either Graphic Violence or Credible Threat of Violence.
- Once they are reported to Facebook, the reportee will be blocked from Facebook after the complaint is weighed against the Community Standards.

# Abuse Standards Violations

All the items below should be confirmed; anything not on this list can be unconfirmed

## Sex and Nudity

- Any OBVIOUS sexual activity, even if naked parts are hidden from view by hands, clothes or other objects. Cartoons/art included. Foreplay allowed (Kissing, groping, etc.), even for same sex (man-man/woman-woman)
- Naked 'private parts' including female nipple bulges and naked butt cracks; male nipples are ok.
- Pixelated or black-barred content showing nudity or sexual activity as above.
- Naked children, including cartoon versions (able to stand on their own) and older minors - **Escalate if unsure of sexual context (child porn)**
- Depiction of sexual assault or rape in any form.
- Mothers breastfeeding without clothes on.
- Escalate** bestiality, necrophilia, and pedophilia. Confirm cartoon/digital versions BUT **escalate** if content is promoting.
- Digital/cartoon nudity. Art nudity ok.
- People "using the bathroom".
- Blatant (obvious) depiction of camel toes and moose knuckles.
- Sex toys or other objects, but only in the context of sexual activity.
- Depicting sexual fetishes in any form.

## Illegal Drug Use

- Unconfirm all marijuana unless context is clear that the poster is selling/buying/growing.
- Illegal drugs shown NOT in the context of medical, academic or scientific study.
- Note:** Any depiction of marijuana alone (any amount) or implements used for smoking marijuana are ok (unconfirm)

## Theft Vandalism and Fraud

- Praising or displaying crimes that they or their friends committed
- Organizing criminal activity or soliciting illegal services.
- Encouraging others to engage in criminal activity.
- Escalate** based on credibility assessment

## Hate Content

(Valid Name Match not required)

- Slurs or racial comments of any kind
- Attacking based on protected characteristics
- Hate symbols, either out of context or in the context of hate phrases in support of hate groups.
- Showing support for organizations and people primarily known for violence.
- Depicting symbols primarily known for hate and violence, unless comments are clearly against them.
- "Versus photos" or "Vs photos": photos comparing two people side by side.
- Any photoshopped images of people, whether negative, positive or neutral
- Images of drunk and unconscious people, or sleeping people with things drawn on their faces.
- Videos: Street/bar/schoolyard fights even if no valid name match is found. School fight videos are only confirmed if the video has been posted to continue tormenting the person targeted in the video.

## Notes:

- Hate symbols are confirmed if there's no context OR if hate phrases are used
- Humor overrules hate speech UNLESS slur words are present or the humor is not evident.

## Graphic Content

- Content showing Poster's delight in/involvement in/promoting of/encouraging of violence against humans or animals for sadistic purposes (e.g. torture, staged animal fights, animal starvation, obvious neglect, etc.)  
Depicting the mutilation of people or animals, or decapitated, dismembered, charred, or burning humans.  
Poaching of animals should be confirmed. Poaching of endangered animals should be **escalated**
- Urine, feces, vomit, semen, pus, and ear wax. (Cartoon feces, urine and spit are OK; real and cartoon snot is OK)
- Violent speech (Example: "I love hearing skulls crack")
- Photos and digital images showing internal organs, bone, muscle, tendons, etc. Deep flesh wounds are ok to show; excessive blood is ok to show.
- Crushed heads, limbs, etc are ok as long as no insides are showing
- Note:** No exception for news or awareness related content.

## IP Blocks and International Compliance

### Escalated:

- Holocaust denial which focuses on hate speech
- All attacks on Ataturk (visual and text)
- Maps of Kurdistan (Turkey)
- Burning Turkish flag(s)

### Confirmed (unless clearly against PKK and/or Ocalan):

- PKK support and depiction
- Abdullah "Apo" Ocalan-related content

## Self Harm

**Note:** All self harm content should be **escalated**.

- Threat and serious promotion of suicide.
- Supporting people, groups, and symbols that advocates and promoting eating disorders as a lifestyle choice.
- Depicting self-mutilation and groups and people that promote and support it (ex: cutting groups)

## Bullying and Harassment

- Valid name matches no matter what the content is (negative, positive or neutral)
- Contacting other users persistently without prior solicitation or continue to do so when the other party has said that they want no other further contact with the sender.
- Attacking anyone based on their status as a sexual assault or rape victim.

Attacking anyone based on their status as a sexual assault or rape victim

(Escalate as per credibility assessment)

- Credible threats or incitement of physical harm against anyone
- Credible indications of organizing acts of present or future violence
- Any threats of violence against Heads of State (HOS) or Law Enforcement Officers (LEO) should always be **escalated** even if not credible
- Any credible indication of terrorist activity or organized past/future crime.

Attacking based on protected categories

# REDRESS FOR SEXIST HATE SPEECH – TWITTER

- Self-reporting model
- Twitter Help Center – How to report violations, or Abusive Behaviour Policy page
- 3 options when reporting
  - “Someone on Twitter is posting my private information”
  - “Someone on Twitter is being abusive”
  - “Someone on Twitter is sending me violent threats”
- Alternative means of dealing with the problem
  - Blocking the user – no notification sent to blocker
  - Flag either media or individuals tweets or direct messages for violations (not available on apps)

# REDRESS FOR SEXIST HATE SPEECH – YOUTUBE

- “We encourage free speech and defend everyone's right to express unpopular points of view. But we don't permit hate speech (speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status, and sexual orientation/gender identity).”
- There is zero tolerance for predatory behavior, stalking, threats, harassment, invading privacy, or the revealing of other members' personal information. Anyone caught doing these things may be permanently banned from YouTube.
- YouTube adopts a self-reporting model, where users are invited to “flag” content or comments for review by YouTube. Two general types of complaints are possible: a content-related complaint, or a privacy-related complaint

- **Legal remedies**

- Few law specifically designed to target sexist hate speech
- Difficulty in balancing freedom of expression and equality/non-discrimination
- Harassment/stalking laws require different threshold

- **Policy frameworks**

- Requires self-reporting
- Opaque process with no ability to scrutinise
- Often lack of responsiveness

**SO WHERE ARE THE  
GAPS?**