

CRITERIA A - CONTENT TYPOLOGY

A broad diversity of types of content can potentially be illegal in certain countries or represent a risk of harm to users. The ease of public access and viral propagation of expression that was previously kept in private also raise new challenges. In a context of lack of sufficiently clear and agreed international definitions, the table below is a non-exhaustive attempt (that might be subject to further refinement) at describing the main issues at stake, to help all actors develop diversified and nuanced approaches to each type of challenge, in the respect of international human rights. It should not be understood as a normative index of content that should be restricted.

Types of Content	Description
Rights of the Child: Art. 24 ICCPR	
<p>Article 24 states that every child shall have, without any discrimination as to race, color, sex, language, religion, national or social origin, property or birth, the right to such measures of protection as are required by his status as a minor, on the part of his family, society and the State. The Declaration of the Rights of the Child states, "the child, by reason of his physical and mental immaturity, needs special safeguards and care..."The Convention on the Rights of the Child defines children as individuals under the age of 18 and Art. 17 requires States Parties to "[e]ncourage the development of appropriate guidelines for the protection of the child from information and material injurious to his or her well-being, bearing in mind the provisions of articles 13 and 18." The Universal Declaration of Human Rights proclaims that childhood is entitled to special care and assistance.</p>	
Child Abuse Material or anything objectionable involving minors	Content which includes sexual or sexually suggestive content involving minors, child abuse imagery or other content posted with the intent to do harm and take advantage of their youth. This may include image privacy rights for children under 13 and up to 18 depending on jurisdiction and context.
Grooming or predation	Online grooming is when a person uses social media to deliberately cultivate an emotional connection with a child in order to sexually abuse or exploit that child.
Right to Privacy: Art. 17 ICCPR	
<p>Article 17 protects the right to respect of privacy, family, home and correspondence, and the protection of honor and reputation. It states that "[n]o one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks." Any restrictions need to be lawful, necessary and proportionate. See</p>	

Also CCPR/C/GC/16 ¹ .	
Breaches of personal information	Personally Identifiable Information (PII), Sensitive Personal Information (SPI) or confidential information which is disclosed without the person's consent. Definitions vary among jurisdictions but generally include any information which reveals a person's identity. This can also include content that facilitates identity theft by posting or soliciting personally identifiable information, sharing PII via external link, sharing private financial info of business, self, or others, and sharing private contact info. "Phishing" for instance is the use of fake email, text messages, or copycat websites to steal PII.
Defamatory personal content	Content which can cause injury to dignity, reputation or personality rights. Platforms generally do not restrict this type of content unless it crosses the threshold to hate speech or incitement. The platforms surveyed restrict allegedly defamatory content only when it becomes coordinated with an intent to harm. The rights of individuals to be protected from "unlawful attacks" on his/her "honor and reputation" must be balanced with the rights of speakers to hold opinions without interference and right to access information. However, political speech or opinion, criticisms of public officials acting in their official capacity or speech which is in the public interest, even if it is considered defamatory, will receive the highest level of protection under international law. Depending upon context and intent, defamatory content may be subject to legitimate restriction under art. 19(a) for respect of the rights or reputations of others.
Coordinated/organized attempts at defamation	Defamatory content becomes coordinated when an individual or organized group spreads the impugned content simultaneously on various platforms.
Defamatory autocomplete suggestions or search results pointing to defamatory content ("Google Bombing," "Googlewashing")	These terms refer to a practice of artificially elevating a particular website in search results by linking it to a search term which may be derogatory or defamatory. Motivations may be personal, political, or just as a prank.
"Right to be Forgotten"	These claims surround requests to delist information that is no longer valid due to passage of time and changing of circumstances, and the continued accessibility of the information constitutes a

	violation of reputational rights.
Impersonation (fake accounts/profiles/pages)	Copying a user's layout, using a similar username, or posing as another person in profiles, pages, comments, emails, or videos. This is often done with the intent to harm an individual or mislead viewers. It also includes bot or other applications for propaganda. This includes impersonation of a channel (i.e. YouTube) and impersonation of an individual (i.e. Facebook or Twitter) or of a company/corporation. Impersonation for satirical or artistic purposes would be protected.
"Deep Fakes"	This has been described by the UK Government as "audio and videos that look and sound like a real person, saying something that that person has never said." Depending upon the context and intent, as well as the jurisdiction (i.e. US First Amendment), this content may be protected.
Sexual objectification	Content that objectifies their targets, including through manipulated photographs and sexually explicit descriptions of their bodies. Photographs are often used without their consent and manipulated so that they appear in pornographic scenes or used in memes.
Unauthorized Dissemination of Intimate Images ("Revenge porn")	Distribution of sexually graphic images without the consent of the subject of the images. The abuser obtains images or videos in the course of a prior relationship, or hacks into the victim's computer, social media accounts or phone. This is usually done with the intent to harass, humiliate injure the person.
Right to Freedom of Expression: Art. 19 ICCPR	
<p>Art. 19 (1): Right to hold opinions without interference</p> <p>Art.19(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. General Comment 34 (CCPR/C/GC/34²) also highlights how States parties have to be proactive in putting in place "effective measures to protect against attacks aimed at silencing those exercising their right to freedom of expression" (para 23).</p>	
Disinformation	Disinformation includes the distribution of false or inaccurate information (i.e. "fake news") for political, ideological or economic gain, either by individuals or bots. Extreme examples aim to influence elections and disrupt democratic processes. It can take the form of "news" articles based partially on fact and misinformation, tweeting, posting or commenting. This type of content often does not cross legal boundaries and must be carefully

	distinguished from opinion and satire which are protected forms of expression.
Medical misinformation	Content which disseminates false or misleading information that could have detrimental effects on individual or public health and safety. Examples include promotion of false cures for illnesses and anti-vaccination advice. This may be done for financial gain or simply out of ignorance rather than with an intent to harm. This type of content has been subject to restriction where it has been deemed a threat to public health.
Sexually explicit content; nudity or porn	Images of explicit sexual activity or fetishes, and nude or partially nude people in sexually suggestive poses. This type of content may be protected for adults depending on the jurisdiction or community guidelines but restricted for children or youth. Artistic, scientific, documentary or educational nudity is protected. Images which identify an individual and are posted without consent, may be subject to restriction.
Content critical of religion (Blasphemy/Apostasy)	Content critical of religion, including opinions or artistic works (i.e. satirical cartoons) is protected under ICCPR Art. 19 but it must be balanced with legitimate restrictions under art. 18 (3) (Freedom to manifest one's religion or beliefs may be subject only to such limitations as are prescribed by law and are necessary to protect public safety, order, health, or morals or the fundamental rights and freedoms of others) and art. 20 if it advocates religious hatred that constitutes incitement to discrimination, hostility or violence. Some States restrict this type of content under Blasphemy/Apostasy laws.
<p>Art. 19(a): Legitimate Restriction for respect of the rights or reputations of others.</p> <p>General Comment 34 (CCPR/C/GC/34 para 35)² highlights that “When a State party invokes a legitimate ground for restriction of freedom of expression, it must demonstrate in specific and individualized fashion the precise nature of the threat, and the necessity and proportionality of the specific action taken, in particular by establishing a direct and immediate connection between the expression and the threat”.</p>	
Defamatory personal content	See Defamatory personal content under Right to Privacy (above). General Comment 34 ² calls on State parties to “consider the decriminalization of defamation” and recommends that, in any case of defamation, “imprisonment is never an appropriate penalty”

	(para 47).
Bullying	Content that purposefully targets private individuals with the intention of degrading or shaming them. An extreme form of this is called "flaming." Does not apply to public figures who are expected to tolerate higher levels or criticism within reason, in so far as the content does not include hate speech or credible threats.
Harassment	Content which is disseminated on multiple occasions to cause an individual stress, humiliation, anxiety or fear of violence. Content may contain targeted swearing, grossly offensive comments, or threats of physical harm or even of death. Depending on intent and context, only speech which is considered hate speech or a credible threat may be restricted. Many states have harassment statutes which are applicable for online violations.
Coordinated/organized Harm	Deliberately sabotaging or invading multiple online spaces for the purposes of harassing a target. Users are currently unable to report this scope and context of the harassment, as each platform will only consider the harassment happening on their own sites.
Cyberstalking	No legal definition but examples include repeated threatening or obscene emails or text messages, spamming, 'flaming' (targeted online verbal abuse), "Baiting" through taunts, or sending menacing unsolicited messages.
Deadnaming	Using/disclosing a transgender person's birth name to harass them, invalidate their identity, and/or inflict emotional stress.
Doxing	Searching for and publishing private or identifying information about a particular individual, often by hacking and with malicious intent. "Dox" is a slang version of "documents." Causing fear, stress and panic is the objective of doxing, even when perpetrators think or say it is "harmless."
Art. 19 (3)(b): Legitimate restriction for the protection of national security or of public order (ordre public), or of public health or morals, if it conforms to the strict tests of necessity and proportionality. See CCPR/C/GC/34 ²	
Violent/graphic content	Content that is that is sensational or gratuitously violent or glorifies violence. This must be distinguished from graphic content which may have educational, scientific or public interest purposes, such as information on historical or current events, which would be protected under international human rights law. Depending on the type of content, it may require age verification or be age restricted.
Promoting or	Content which incites or abets criminal activity and is believed to be

publicizing crime	a credible threat to personal or public safety or property.
Content to organize violence or support violent organizations	Content that makes credible threats of serious physical harm (organized violence, murder, human trafficking) against a specific individual or defined group of individuals or expresses support or praise for groups, leaders or individuals involved in these activities.
Sexual violence and exploitation	Content that depicts, threatens or promotes sexual violence, assault or exploitation.
Abetting self-harm or suicide	Content that promotes harmful behavior, or anything that encourages or suggests self-harm, like mutilation, eating disorders or drug abuse. Content that identifies and negatively targets victims or survivors of self-injury or suicide.
Leaked confidential or secret information	Some states may justify restricting leaked sensitive information on national security grounds. However, information in the public interest which is leaked by "whistleblowers" may receive protection under the UN Convention Against Corruption. Art. 33 recommends states provide protections to "reporting persons" who report "in good faith and on reasonable grounds to the competent authorities any facts concerning offenses" covered by the convention. Art. 32 provides protection of witnesses, experts, and victims.
Lèse-majesté or comments critical of historical personages	A few jurisdictions restrict speech considered offensive to the monarchy or historical personages on the grounds that it is either treasonous or a threat to public order. However, open and critical discussion of political leaders and public figures is protected under international law. Only credible threats to incitement may warrant restrictions.
Prohibition of propaganda for war and inciting national, racial or religious hatred: Art. 20	
Art. 20(2) Advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law	
Hate speech	Hate speech includes serious attacks on people based on their race, ethnicity, national origin, caste, religion, gender identity, sexual orientation, disability, veteran status or medical condition. May also include targeting people based on age, weight, immigration or veteran status. Examples are violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation. This also may include imagery, such as of lynching, or coordinated conduct to discriminate or dehumanize. Live-streaming or posting of archived videos of live events which amplify or incite hate crimes

	<p>may require immediate restrictions (i.e. Christchurch Massacre). ICCPR articles 18, 19, 20 and 26. ICERD art. 4. Although there is no internationally agreed upon definition of or threshold for hate speech, states have an obligation under the ICCPR to prohibit advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. It is tied to a guarantee of equal protection before and of the law. Evaluating hate speech can be very subjective and therefore great care must be taken as overly broad restrictions can infringe on other fundamental rights. See also CCPR/C/GC/11³ and Rabat Plan of Action.</p>
<p>Violent content extremist</p>	<p>Most platforms prohibit content which they define as "extremist". Definitions include content which "incites" violence, "celebrates" terrorist acts, "instructs," solicits or advocates persons or a group of person to participate in the activities of a terrorist group or provides instruction on the making or use of weapons. Some platforms deny accounts to organizations labelled as terrorist or which engage in premeditated violent activities against persons or property as an act of intimidation with a political, religious or ideological purpose. This includes terrorist hoaxes. ICCPR Art. 20 prohibits any propaganda for war. However, any restrictions should expressly exclude content which is disseminated for "educational, journalistic, artistic or research purposes or awareness raising activities against terrorism." Content flagged as "terrorist" may in some instances constitute documentation of war crimes or atrocities, and therefore should not be deleted but shared with the appropriate law enforcement authorities. Although art. 19 (3)(b) provides a legitimate restriction for the protection of national security, there is no agreed upon definition for terrorist content under international law and to that end definitions for terrorist content should be "clear, foreseeable and narrow to prevent unlawful interferences with fundamental rights." See also Art. 6 ICCPR (Right to Life); International Instruments (UN)⁴</p>
<p>Intellectual Property Rights</p>	
<p>Copyright</p>	<p>Copyright is a legal right that protects original creative works such as music, movies, works of art or books. It does not protect facts or ideas.</p>
<p>Trademark</p>	<p>A trademark is a word, phrase, symbol, and/or design (i.e. logos,</p>

	brand names) that identifies and distinguishes the source of the product/services of one party from those of others. It does not expire as some copyrights do.
Regulated Goods and Services	
Content pertaining to regulated or illegal goods and services will vary depending upon the jurisdiction.	
Regulated goods and services	States as well as platforms restrict content which facilitates the purchasing, trading or selling of illegal drugs, illegal services (online gambling, counterfeit documents), stolen goods, firearms or other weapons, based on the laws in the jurisdiction. Some states or platforms may also restrict content which promotes the use of illegal drugs or weapons or provides instructions for manufacturing weapons (i.e. 3D printing).
Sexual solicitation	Content disseminated with the intent to engage in sexual activity for a fee or the functional equivalent of a fee.
Fraud	
Fraudulent content seeks to deliberately deceive individuals for unlawful gain or to deprive them of their rights. Most jurisdictions cover fraudulent activities in their civil or criminal statutes.	
Misleading metadata (title, description, tags, annotations, and thumbnail)	Includes titles, descriptions, tags, annotations, and thumbnails being used to game or trick the search algorithms for online video, rather than being representative of the actual content in the video.
Blackmail/extortion	Content or messages which threaten to reveal embarrassing information or photos of the victim (which may have been collected illegally or with the consent of the victim) unless the victim provides favors, property or money. Sextortion is a form of blackmail where sexual information or images are used to extort sexual favors, money or other demands from the victim.
Scams: trick others for their own financial gain	Content that deliberately tries to mislead users for financial gain or to access PII. Examples are purchasing views, deceptive layouts, artificial subscriptions, serving pop-up ads and re-directs, vote manipulation (manipulating content votes up/down), or "spoofing" brand name logos.
Spam	Content which includes targeted, unwanted, or repetitive content in videos, comments, private messages, or off-domain redirects. It may have the intent to artificially boost views or drive down scores and other metrics through coordinated campaigns (i.e. troll teams).

	It can also include misleading content or behavior, like deceptive design elements or suspicious pop-ups.
Artificial traffic spam: artificially incentivize viewers for engagement	These are targeted messages to incentivize views, called "view count gaming," which tries to make a non-view into a view for financial gain.
Fraudulent accounts	Accounts which are run by bots rather than humans with the intent to spread misinformation and distort debate.

(Sources, as footnotes in table):

¹ CCPR/C/GC/16: <https://www.refworld.org/docid/453883f922.html>

² CCPR/C/GC/11:
<https://www.ohchr.org/Documents/Issues/Opinion/CCPRGeneralCommentNo11.pdf>

³ CCPR/C/GC/34:
<https://bangkok.ohchr.org/programme/documents/general-comment-34.aspx>

⁴ Art 6 - International Legal Instruments:
<https://www.un.org/counterterrorism/ctitf/en/international-legal-instruments>