Digital Policy Lab

Policy Digest #16

14 November 2024

ISD Powering new generations against extremism

ALFRED LANDECKER FOUNDATION

Policy Digests offer an overview of recent digital policy developments in Digital Policy Lab (DPL) member countries, including regulatory and non-regulatory initiatives aiming to combat online harms such as disinformation, hate speech, extremist or terrorist content. In addition to general updates, each Policy Digest provides a snapshot of topic-specific schemes relevant to the upcoming DPL session.[1]

Section 1 Digital policy developments

European Union: European Commission launches public consultation on draft Delegated Act on data access provided for in the Digital Services Act

Type Regulatory (consultation)
Status Published

On 29 October 2024, the European Commission opened a public consultation on their draft Delegated Act on data access under the Digital Services Act (DSA). The Delegated Act intends to clarify conditions under which vetted researchers can access non-public data from very large online platforms (VLOPs) and very large search engines (VLOSEs) to enhance platform transparency and accountability, as per Article 40 of the DSA. The framework outlines the conditions, relevant procedures, and advisory mechanisms related to this access. Key points include the requirement for VLOPS to compile a detailed "data inventory", recognition of the data needed to study systemic risks, provisions for researchers to propose flexible access methods, and the establishment of a "DSA Data Access Portal" to serve as a central interface for researchers, regulators, and platforms to simplify data access requests and enhance transparency. The consultation will remain open until 26 November 2024. The Commission intends to implement the regulation in the first quarter of 2025.

European Union: Council Declaration on Combating Antisemitism with a focus on online offences

Type Declaration
Status Adopted

On 15 October 2024, the Council of the EU issued a declaration addressing antisemitism across the EU. The declaration calls on Member States to establish national strategies aimed at combatting antisemitism and to appoint special envoys or coordinators. The declaration emphasises that online hate crimes should be prosecuted similarly to those committed offline, in line with relevant legal frameworks, and that online platforms must comply accordingly and put adequate detection and mitigation measures in place as laid out in the DSA and the Code of Conduct on tackling illegal hate speech online.

European Union: Court of Justice of the European Union ruling concerning the processing and analysis of personal data for targeted advertising under the GDPR

Type Litigation
Status Decision

On 4 October 2024, the Court of Justice of the EU ruled on processing personal data for targeted advertising under the GDPR. The case concerned Meta and privacy advocate Maximilian Schrems, focusing on the legality of using personal data,

[1] We welcome any feedback from DPL members regarding additional developments, as well as own submissions from DPL members who wish to be featured in the digest.

particularly on sexual orientation, for targeted advertising. Schrems publicly disclosed his sexual orientation during a panel discussion but did not share it on Facebook. Meta was found to have collected user data both on and off its platform via cookies and social plug-ins for targeted advertising. The Austrian Supreme Court sought clarification from the Court of Justice on whether Schrems' public disclosure allowed Meta to process sensitive data under the GDPR. The court determined that the GDPR's principle of data minimisation limits the unrestricted aggregation, analysis and processing of personal data for targeted advertising, regardless of its source. It also stated that public disclosure of sexual orientation does not authorise Meta to process related personal data unless compliant with the GDPR.

European Union: European Commission investigation into YouTube, Snapchat, and TikTok on DSA compliance to transparency of recommender systems

Type Regulatory (investigation)
Status Information request

On 2 October 2024, the EC requested information from YouTube, Snapchat and TikTok on their compliance with the DSA concerning their recommender systems, with the deadline of 5 November 2024. YouTube and Snapchat must detail the parameters of their algorithms used to recommend content and how these contribute to systemic risks, including those affecting electoral processes, civic discourse, and users' mental health. Moreover, they must lay out their measures on how they protect minors and mitigate the spread of illegal content. TikTok must provide information on measures taken to prevent manipulation by malicious actors and address risks concerning elections, civic discourse and media pluralism. Once the platforms' responses are submitted, they are assessed by the EC who will determine the next enforcement measures.

European Union: European Commission closes consultation on enforcement guidelines on the protection of minors online under the DSA

Type Regulatory (guidelines)
Status Closed consultation

On 30 September 2024, the EC concluded its consultation on guidelines designed for online platforms to strengthen the protection of minors online under the DSA. The guidelines focus on the DSA's requirement to provide a high level of privacy, safety and security for minors on online platforms. The final guidelines will suggest best practices and recommendations to protect minors from risks and encourage platforms to consider a risk-based approach to online harm and proactively carry out regular risk assessments. The guidelines are planned for adoption in the first quarter of 2025.

European Union: European Parliament publishes impact assessment results to complement the proposed EU AI Liability Directive

Type Regulatory (guidelines)
Status Published

On 19 September 2024, the EU Parliament published findings from its supplementary impact assessment on the proposed EU AI Liability Directive to revise non-contractual civil liability rules concerning AI to better protect individuals from complex AI-related harm. The proposal intends to broaden liability to include general-purpose and high-impact AI systems and relevant software, introducing a mixed liability framework that employs fault-based and strict liability models. It also suggests shifting from an AI-specific directive to a broader software liability legislation to prevent market fragmentation and establish clearer legal standards across the EU.

European Union: EU AI Office closes consultation on the General-Purpose AI Code of Practice

Type Regulatory (consultation)
Status Closed

On 18 September 2024, the EU AI Office concluded its consultation on trustworthy general-purpose AI models as part of an effort to establish a Code of Practice, anticipated by Article 56 under the EU AI Act, to outline the AI Act's provisions for providers of general-purpose AI models. Consequently, the consultation addressed issues related to transparency and copyright obligations, the taxonomy of systemic risk, risk assessment, and mitigation strategies. The Commission will publish a summary of the consultation results and aims to finalise the Code of Practice by April 2025.

G7: G7 Data Protection and Privacy Authorities issue two statements: one on fostering trustworthy AI and a second one on AI and children

Type Voluntary (joint statement)
Status Published

On 11 October 2024, the G7 Data Protection Authorities (DPAs) issued two separate statements highlighting their commitment to fostering trustworthy AI and protecting children's fundamental rights and freedoms regarding AI. One statement emphasises the role of DPAs in addressing AI-related risks by fostering trustworthy AI. It focuses on integrating data protection principles such as fairness, transparency, and accountability into AI governance, including bias and discrimination. It also called for enhanced collaboration between jurisdictions and regulatory bodies to promote responsible AI development. The other statement specifically addresses AI risks for children, advocating for age-appropriate safeguards and incorporating privacy by design within transparent AI models. It stresses the importance of protecting children from online manipulation, discrimination, and the misuse of personal data. It also emphasises the need to improve digital literacy within educational frameworks and reinforce international collaboration among data protection authorities to safeguard children's rights.

France: Regulatory Authority for Audiovisual and Digital Communication (ARCOM) adopts technical standard on age verification systems for access to online pornographic content

Type Regulatory (technical standard)
Status Adopted

On 9 October 2024, ARCOM, France's Digital Services Coordinator (DSC), established a technical standard for age verification systems used by online services distributing pornographic material to safeguard minors from harmful content. The standard mandates that these systems effectively distinguish minors and adults, incorporate anti-fraud measures, and adhere to the GDPR's provisions on personal data handling. Additionally, it requires the use of a double anonymity mechanism to protect user identities and offers at least two methods for validating age. Regular audits and their findings must be made public to enhance user awareness regarding privacy safeguards. This standard is part of France's Security and Regulation of the Digital Space law (SREN) which adjusts the DSA and Digital Markets Act (DMA) to French law. It came into effect in May 2024 and provides ARCOM with the authority to issue a compliance notice if standards are not met. Continued non-compliance could lead to a financial penalty.

Germany: Bundestag passes motion concerning freedom of expression on social media platforms

Type Motion
Status Passed

On 18 October 2024, the Bundestag passed a resolution No. 20/13364 on no restriction of freedom of expression on social media platforms with the call to abolish the DSA. The resolution was brought forward by the far-right political party Alternative für Deutschland (AfD, Alternative for Germany) and criticises the tendency of online services to suppress users' legally permissible posts on social media platforms under platform policies aimed at combating hate speech or misinformation. The motion urges the federal government to repeal the DSA and halt financial backing to organisations that assist platforms in removing posts protected by freedom of expression laws and to investigate and eliminate any potentially anti-competitive deletion practices. Additionally, it stipulated that the government should refrain from appointing "trusted flaggers" as laid out in the DSA.

Ireland: Irish Data Protection Commission (DPC) fines LinkedIn €310 million for breaching GDPR data processing rules for advertising and behaviour tracking

Type Litigation
Status Decision

On 22 October 2024, the Irish DPC An Coimisinéir Cosanta Sonraí issued a final decision regarding Microsoft Linkedin's processing of personal data for behavioural analysis and targeted advertising, following a complaint lodged with the French DPA Commission nationale de l'informatique et des libertés (CNIL). The inquiry determined that LinkedIn did not lawfully process member data, failing to validly rely on consent, legitimate interests, or contractual necessity under the GDPR. As a result, LinkedIn received a reprimand and an order to comply, along with administrative fines amounting to €310 million for infringing on various GDPR provisions. This is the fifth largest penalty that the Irish DPA has imposed under the GDPR and the sixth largest overall by any EU authority since the regulation was established in 2018. LinkedIn has 30 days to contest the ruling.

Ireland: Irish Media Commission published Online Safety Code

Type Regulatory
Status Adopted

On 21 October 2024, Ireland's media commission Coimisiún na Meán released the finalised Online Safety Code, which imposes binding regulations on video-sharing platforms headquartered in Ireland to safeguard individuals, particularly children, from harmful content. The Code includes provisions to ban the uploading and sharing of dangerous material, such as cyberbullying, self-harm promotion and any form of incitement to hatred or violence. It also mandates age assurance measures to shield children from pornography and gratuitous violence, alongside necessary age verification protocols, as well as offering parental controls to protect the well-being of children under 16 years old. The Code is scheduled to be implemented in Q4 after receiving approval from the EU.

Netherlands: Dutch authorities adopt guidelines on responsible AI development under the EU AI Act

Type Regulatory
Status Adopted

On 16 October 2024, the Dutch government released a guide on the implementation of the EU AI Act, outlining rules for responsible AI development and use to safeguard public safety, health and fundamental rights. The guide proposes compliance measures such as a risk assessment to classify AI systems as prohibited, high-risk, or other categories, and clarifies whether the assessed AI system applies to the EU AI Act rules, and whether the organisation is an AI provider or user as laid out in the Act. The guide details the gradual implementation of the regulation, which is expected to be fully in effect by mid-2027, with specific AI systems encountering restrictions starting from February 2025.

Netherlands: Consultation: Guidance on Manipulative, Deceptive and Exploitative AI systems

Type Regulatory (consultation)
Status Closed

On 27 September 2024, the Dutch DPA initiated a consultation to collect input on the EU AI Act's prohibitions concerning certain AI systems, focusing on manipulative, deceptive, and exploitative practices under Article 5 of the EU AI Act, which will take effect on 2 February 2025. The consultation closed 17 November 2024. After this call for input, a summary will be published. Other calls for input on different parts of the AI Act are due to follow including a call on AI systems for emotion recognition in the workplace or in education first.

South Korea: Proposed legislation against producing, possessing and distributing intimate AI-generated images

Type Legislative
Status Awaiting formal adoption

On September 26, 2024, South Korea passed a bill which criminalizes the possession and watching of sexually explicit deepfake images and videos. Pornography material and the distribution of pornography is already illegal in the country. Following this bill, anyone who watches, saves, or purchases deepfake pornography faces a fine up to the equivalent of $22,600 USD or three years in jail while anyone producing deepfake pornography faces a fine up to the equivalent of 37,900 USD or up to seven years in jail. The bill is awaiting approval by President Yoon Suk Yeol before it can be enacted.

United Kingdom: Ofcom launches consultation on Online Safety Act fees and penalties

Type Regulatory (consultation)
Status Published

On 24 October 2024, Ofcom launched its first consultation regarding the implementation of a new fees and penalties regime under the Online Safety Act (OSA). The consultation outlines proposals for defining qualifying worldwide revenue, which will be used to determine both the fees and penalties imposed on regulated services under the OSA. It addresses possible exemptions from the fee regime, the Statement of Charging Principles, and the information providers are required to submit for fee notifications. Stakeholders are invited to respond by 9 January 2025. Regulated services' first online services risk assessments are due on 31 March 2025 and mark the first step to enforcing the Act.

United Kingdom: Data (Use and Access) Bill introduced to the House of Lords

Type Regulatory (proposal)
Status Under deliberation

On 23 October 2024, the UK government introduced the Data (Use and Access) Bill to the House of Lords. The proposed legislation aims to reform data protection laws by incorporating elements from the previous Data Protection and Digital Information Bill. The new Bill aims to facilitate the secure sharing of 'smart data', enhance digital verification services, simplify data protection principles and broaden data access for online safety researchers (Clause 123 creates a power for the Secretary of State to put in place a framework for researchers to access data held by tech companies to conduct research into online safety matters, which would be put in place through secondary legislation). It introduces a revised definition of personal data and adjustments to the role of Data Protection Officers and Data Protection Impact Assessments, while also relaxing some accountability measures. Additionally, it established the concept of "vexatious" data subject access requests, requires the Information Commissioner (ICO) to consider government strategic priorities, enhances the ICO's enforcement powers.

United States: The U.S. Senate Select Committee on Intelligence held a hearing on "Foreign Threats to Elections in 2024 – Roles and Responsibilities of U.S. Tech Providers"

Type Hearing
Status Complete

On September 18, 2024, the U.S. Senate Select Committee on Intelligence held a hearing on "Foreign Threats to Elections in 2024 – Roles and Responsibilities of the U.S. Tech Providers". The hearing hosted representatives from Alphabet, Meta, and Microsoft to discuss the companies' roles in platform security, disinformation, foreign threats, and content moderation. Democratic Committee Chairman Mark Warner questioned the witnesses on how misinformation and disinformation campaigns are reaching American users, specifically citing Russia's doppelganger campaign which imitates legitimate news organizations to push pro-Kremlin narratives. Republican Vice Chairman Marco Rubio probed the representatives about the companies' content moderation policies, particularly on how Meta determines what is false information. Meta representative Nick Clegg, President of Global Affairs, detailed Meta's use of independent fact checkers to verify information and promised to send a list of all fact-checking organizations that Meta uses.

United States: Kids Online Safety Act (KOSA) passed out of House Committee on Energy and Commerce

Type Legislation
Status Awaiting vote in the House of Representatives

On September 18, 2024, the House Committee on Energy and Commerce held a markup of the Kids Online Safety Act (KOSA), where committee members debated and amended the proposed legislation. The bill creates a duty of care provision for online platforms that minors are likely to use, requiring them to remove harmful content for minors, gives minors the ability to opt-out of algorithmic recommendations, and disables addicting features. KOSA passed the Senate in July 2024 with a vote of 91-3. Notably, there are significant differences between two versions of KOSA in the House and the Senate. The House version changed the duty of care provision to exclude specific mental-health related rules. The future of the bill is unknown with some House Democrats voicing their intent to vote against the bill given the changes.

United States: The Biden-Harris Administration released a National Security Memorandum (NSM) on Artificial Intelligence

Type Presidential action
Status Issued

On October 24, 2024, the Biden Administration released the first-ever National Security Memorandum (NSM) on Artificial Intelligence (AI), titled "Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfil National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence". The memo directs the U.S. government to use AI to advance the American national security mission while protecting civil liberties, human rights, and privacy. Also, the memo aims for the US to lead the world in the development of safe AI and promoting international governance on AI. The memorandum will guide federal agencies' policy objectives and requires periodic transparency reporting from named agencies.

Global: The Global Online Safety Regulators' Network (GOSRN) publishes regulatory index

Type Regulatory network (joint index)
Status Published

On 24 October 2024, the GOSRN published its first regulatory index, outlining various approaches adopted by nine online safety regulators across five continents. It serves to enhance collaboration amongst online safety regulatory bodies and provides a detailed comparison of the eight regulators' mandates: Australia, Fiji, France, Ireland, Republic of Korea, Slovakia, South Africa, and the UK. It outlines the specific online harms they address, the entities they regulate, and their enforcement capabilities. This initiative aims to promote consistency in online safety regulation.

Section 2 Topic-specific snapshot: "Addressing Hate Speech Across Jurisdictions"

This section summarises selected analyses and responses published by government agencies, civil society organisations and academia on hate speech across jurisdictions with a focus on the EU, UK, US, Canada and Australia.

A Safer Digital Space: Mapping the EU Policy Landscape to Combat Online Disinformation and Hate Speech,
Democracy Reporting International, 2023

This report provides an overview of the relevant laws and policies to combat online disinformation and hate speech in the EU. It analyses 27 case studies – each EU member state – and their approaches to combatting the spread of online disinformation and hate speech. It examines different approaches applied by member states to address these issues and to what extent the DSA has influenced them.

Key insights include:

- The status quo of national approaches to disinformation and hate speech:
  - Most EU countries use already existing hate speech laws and apply them to online hate speech. Some countries have introduced stricter measures with a focus to the online sphere, such as Spain which introduced stricter consequences on publicly supporting terrorism online.
  - However, protected characteristics are considered differently which results in varying definitions on hate speech.
  - Until DSA implementation, countries have had different requirements to online platforms and how quickly they need to act on a notice. In Germany, for example, this was defined under the 2018 national platform regulation, the Network Enforcement Act (NetzDG).
- How the DSA will impact national approaches:
  - Existing national platform regulations such as the NetzDG will be amended or replaced.
  - The recitals of the DSA outline how illegal behaviour can be harmonised across the EU given varying national legislation.
- How recent events have influenced the disinformation and hate speech landscape:
  - The COVID-19 pandemic and Russia's invasion of Ukraine have highlighted the importance to tackle online disinformation campaigns and hate speech to safeguard democratic societies. In response, member states have issues legislative and non-legislative measures.
- On collaboration efforts:
  - The report finds that over the recent years, some member states have acted as experimental grounds for innovative approaches to involve citizens and the private sector in collaboratively tackling disinformation and hate speech.

Navigating hate speech and content moderation under the DSA: insights from ECtHR case law,
Therese Enarsson, 2024

Enarsson examines the challenges of regulating hate speech online and the crucial role of content moderation in safeguarding individuals and democratic values while preserving freedom of expression. The author notes the complexity of creating a regulatory framework that balances corporate interests with social responsibilities. The article extends to a discussion on the broad definition of what constitutes illegal content in the EU under the DSA and the Code of Conduct on countering illegal hate speech online, which uses a definition based on the Council Framework Decision aimed at addressing racism and xenophobia, including public actions that incite violence or hatred against individuals or groups due to their race, colour, religion, or origin. However, given the voluntary nature of the Code, and the absence of a definition under the DSA, platforms are not required to act on hate speech as defined under the Council Framework Decision.

Enarsson suggests that the 2022 Recommendation from the Committee of Ministers offers a broader perspective by including incitement to violence or discrimination based on age, disability, sex, gender identity, and sexual orientation. Additionally, the European Convention on Human Rights (ECHR), which goes beyond EU member states, and related case law, should provide an interpreting framework of hate speech under the DSA. The analysed case law provides insights into the differentiation between free speech and hate speech and the often specific context courts need to consider to come to a judgement, considering the European Court of Human Rights (ECtHR)'s rulings on hate speech as an "ad hoc" approach so far.

In the context of the DSA, Enarsson concludes a lack of clear guidance for Very Large Online Platforms (VLOPs) regarding their responsibilities in protecting users' fundamental rights while moderating hate speech. The author points out that the absence of specific definitions may lead platforms to rely on their own Terms of Service (ToS) rather than national laws, which could adversely affect their moderation practices and potentially curtail freedom of expression due to over-censorship. Although the EU has established a comprehensive framework with the DSA, its effectiveness may be undermined by these ambiguities. Ultimately, VLOPs are expected to navigate these complex regulatory challenges and integrate fundamental rights into their moderation systems and ToS, leaving user safety in the digital space uncertain.

Online Content Moderation – Current challenges in detecting hate speech,
European Union Agency for Fundamental Rights, 2023

This report on online content moderation addresses the difficulties of identifying and removing hate speech on social media. It points out the lack of a cross-jurisdictional definition of online hate speech and the opaque nature of content moderation systems, complicating efforts to understand and combat online hate effectively. The report examines four platforms (Reddit, Telegram, X, and YouTube) in Bulgaria, Germany, Italy, and Sweden, but FRA did not have access to data from Facebook and Instagram. During the specified period, nearly 350,000 posts and comments were collected using specific keywords, with approximately 400 random posts evaluated by human coders from each country to assess hatefulness. The analysis of posts and comments from January to June 2022 revealed significant issues:

- Widespread online hate: Over half (53%) of 1,500 posts examined were deemed hateful by human reviewers.
- Misogyny: Women were the predominant targets across all platforms and countries studied, with most abusive content including denigrating language and involving harassment and incitement to sexual violence.
- Negative stereotyping: Individuals of African descent, Roma, and Jews faced frequent negative stereotyping.
- Harassment: Nearly half (47%) of hateful posts were direct harassment.

To address online hate, the report suggests that the EU and online platforms should:

- Create a safer online environment, focusing on protected characteristics such as gender and ethnicity in moderation efforts. Major platforms like X and YouTube should address sexist online hate in their risk assessments under the DSA.
- The EU and member states should offer clearer guidance and rules on what constitutes illegal online hate.
- Ensure all forms of online hate are captured by establishing and funding a network of trusted flaggers, with adequate training for police, content moderators, and flaggers.
- Test technology for bias to protect against discrimination, as highlighted in FRA's previous reports on 'bias in algorithms' (2022) and 'AI and fundamental rights' (2020).
- Strengthen access to data for independent research, to assess the effectiveness of hate speech detection and its impact on fundamental rights.

Online safety and social media regulation in Australia: eSafety Commissioner v X Corp,
Marcus Smith, Mark Nolan, and John Gaffey, 2024

The article summarizes the Australian approach to online safety by exploring theoretical legal issues and the eSafety Commissioner v X Corp case. Australia is notable for its significant regulation of social media companies, having three laws which legislate on online hate speech and videos of extremist violence: the Online Safety Act, Division 474 Subdivision H of the Criminal Code provides takedown orders of abhorrent violent material, and Division HA of the Criminal Code which attempts to regulate Nazi imagery online and/or in public. Furthermore, the Online Safety Act delegates authority to the eSafety Commissioner to provide takedown notices to companies, which they must comply with within 24 hours and create online safety guidelines and reporting requirements.

In the case of the eSafety Commissioner v X Corp, eSafety sent takedown notices to X after a video of Bishop Mar Mari Emmanuel being stabbed during his sermon was spread on the platform. In total, eSafety issued 109 takedown notices, 65 of which were on X. In response, a debate ensued whether the videos needed to be removed from the platform for all users was necessary for compliance, which eSafety believed, or if geo-blocking the videos for users in Australia was enough, which was the opinion of X. The case went to the Australian Federal Court, which ruled that requiring the videos be removed for all X users was not in scope of the Online Safety Act.

The Legal Aspects of Hate Speech in Canada,

Lex Gill, 2020.

This report summarizes the legal landscape of Canada in relation to hate speech and the internet. In Canada, hate speech is governed in both criminal and administrative contexts. Regarding the former, key sections of the criminal code regulate hate speech in Canada: section 319(2) which prohibits promoting hatred against "identifiable groups" other than in private conversations, section 318 which criminalizes advocating or promoting genocide, section 319(1) which bans inciting hatred against a group which would likely lead to a breach of peace, and section 430(4.1), which bans mischief in relation to property that is motivated by bias. Importantly, the Canadian government also protects an individual's right to freedom of expression and belief to a reasonable limit, drawing the line at violations of the aforementioned criminal code. Moreover, the rise of the internet has led to increasing challenges with addressing online hate speech while maintaining an individual's right to freedom of expression. The report details a few strategies that could be implemented by the government to address online hate speech, including platforms self-regulating and counter speech and public education. While the report does not make conclusions on the best way to mitigate online hate speech, the report lays out Canada's legal landscape on hate speech and discusses the positives and negatives of potential efforts that the Canadian government could use.

The Boundaries of Internet Speech,

Tyler Hoguet, Julia Englebert, and Carson Turner, 2024

The article reviews different arguments on how to regulate online hate speech in the United States. Some scholars believe that the courts need to widen its scope on the type of hate speech that it regulates, which is currently only violent speech that poses "an immediate threat". Other scholars believe that Section 230 of the Communications Decency Act needs to be amended. Section 230 protects platforms from the liability coming from third-party content posted on the respective platforms. Democratic Senator Mark Warner has proposed legislation to reform Section 230 in the last two congressional sessions, but the bills have failed to garner widespread support. Other scholars raise concerns that over-regulating online hate speech will violate individuals' First Amendment right to free speech and could be exploited to censor speech from the political opposition.

Countering and Addressing Online Hate Speech: A Guide for policy makers and practitioners,

United Nations, 2023

In 2019, the United Nations Secretary-general created the United Nations Strategy and Plan of Action on Hate Speech as part of the United Nations' goal to address hate speech globally. While there is no internationally accepted definition of hate speech, the UN Strategy and Plan of Action defines hate speech as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor."

As a result, the United Nations Office on Genocide Prevention and the Responsibility to Protect held round-table discussions over a period of three years with members of the United Nations Working Group on Hate Speech, civil society partners, and technology and social media companies on how to mitigate online hate speech. This report summarizes the round-tables' findings and gives individualized recommendations to governments, technology and social media companies, civil society, and the United Nations on how to curb online hate speech. Some of the report's recommendations for governments include:

- Formulate legislation that addresses hate speech holistically, both online and offline by creating programming that teaches media literacy, inclusion, and builds up societal resilience to online incitement to hate.
- Institute requirements that mandate technology and social media companies to be more transparent on content moderation, algorithms, and data gathering and use.
- Engage with a variety of stakeholders, including technology and social media companies, civil society, and affected communities, to foster dialogue to help shape laws and policies to address online hate speech in accordance with international human rights law.
- Ensure that legislation does not impede an individual's right to freedom of expression or the freedom to seek and receive information.

About the Digital Policy Lab
The Digital Policy Lab (DPL) is an inter-governmental working group focused on charting the regulatory and policy path forward to prevent and counter disinformation, hate speech, extremism and terrorism online. It is comprised of a core group of senior representatives of relevant ministries and regulators from key liberal democratic countries. The DPL aims to foster inter-governmental exchange, provide policymakers with access to sector-leading expertise and research, and build an international community of policy practice around key regulatory challenges in the digital policy space. We thank the Alfred Landecker Foundation for their support for this project.