



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

## Section 1 Digital policy developments

### Australia: Social Media (Anti Trolling) Bill 2021

**Type** Regulatory

**Legislative status** Introduced into parliament

On 1 December 2021, Australia's Attorney-General's Department proposed the [Social Media \(Anti-Trolling\) Bill 2021](#). The Bill is a response to the High Court's decision in [Fairfax Media Publications v Voller \[2021\] HCA 27](#), which ruled that Australians who maintain a social media page (in this case a public Facebook page) may be exposed to liability for defamatory comments posted on that page by others, even if they are not aware of the defamatory material. For the purposes of the general law of the tort of defamation, the Bill proposes:

- A social media service provider, not a social media page owner, is the publisher of a third-party comment posted on that page (a "page" meaning any distinct part of the service);
- A defence in a defamation proceeding for social media service providers, if a complaints scheme meeting prescribed requirements is in place;
- A complaints scheme for a complainant to request the provider to disclose relevant contact details of the commenter (the provider is required to ask the commenter for consent);
- If the provider considers the complaint does not genuinely relate to the potential institution of defamation proceedings, the service is not required to take any action.

### EU: Digital Services Act (DSA)

**Type** Regulatory

**Legislative status** Ordinary legislative procedure: first reading

On 20 January 2022, the European Parliament [voted on amendments to the Digital Services Act \(DSA\)](#) approving a mandate to start trilogue negotiations in late January, with the French Presidency of the Council of the EU, representing Member States, and the European Commission. The amendments:

- Strengthen prohibition of general monitoring obligations to be neither de jure, nor de facto, through automated or non-automated means (Art. 7);
- Reject a de facto exemption for 'media' from content moderation, but create an obligation for intermediary services to respect in their terms and conditions the freedom of expression and freedom and pluralism of the media, as enshrined in the Charter of Fundamental Rights of the EU (Art. 12);
- Prohibit using the structure, function or manner of operation of online interfaces to distort or impair users' ability to make a free, autonomous and informed decision or choice (Art. 13);

<sup>1</sup>We welcome any feedback from DPL members in relation to developments that may have been missed, which can be added to a revised version circulated after the respective session. Looking ahead, we also welcome own submissions from DPL members who wish to be featured in the digest.

- Prohibit targeting or amplification techniques that process, reveal or infer personal data of minors or sensitive personal data referred to in [Article 9\(1\) of Regulation \(EU\) 2016/679 \(GDPR\)](#) for the purpose of displaying advertisement (Art. 24);
- Oblige very large online platforms to provide at least one recommender system which is not based on profiling, as well as an easily accessible functionality on the online interface allowing users to select and modify at any time their preferred option (Art. 29);
- Expand access to data to vetted researchers, vetted not-for-profit bodies, organisations or associations for the sole purpose of conducting research that contributes to the identification, mitigation and understanding of systemic risks (Art. 31).

The first trilogue negotiations took place on 31 January 2022, and will continue over the coming months, alongside parallel negotiations on the Digital Markets Act (DMA).

## EU: Proposal on a regulation on the transparency and targeting of political advertising

**Type** Regulatory

**Legislative status** Awaiting committee decision (Internal Market and Consumer Protection Committee (IMCO))

On 25 November 2021, the European Commission published its proposal for a regulation on the transparency and targeting of political advertising. The proposal is part of the [Commission's 2021 work programme](#), which included as one of its priorities 'A New Push for European Democracy'. The aim is for the new rules to enter into force and be fully implemented by Member States by Spring 2023, one year before the elections of the European Parliament. The proposal stipulates that:

- Paid political advertising must be clearly labelled and provide a set of key information, including an easily retrievable transparency notice containing information on the dissemination period, any linked election, amount spent for the specific advertisement as well for the entire advertising campaign, and the source of the funds;
- Targeting and amplification techniques, which use or infer sensitive personal data (such as ethnic origin, religious beliefs or sexual orientation), will be prohibited; and the use of such techniques will be allowed only after explicit consent from a person concerned;
- Member States will be required to introduce effective, proportionate and dissuasive fines when the rules on transparency of political advertising are breached.

## EU: Revision of the Code of Practice on Disinformation (CoPD)

**Type** Self-regulatory

**Legislative status** Revision

On 2 December 2021, the European Commission [announced](#) that 26 new prospective signatories joined the process of drafting the strengthened Code of Practice on Disinformation (CoPD). The revised Code will outline granular commitments adapted to diverse services. Current and prospective signatories are expected to deliver the strengthened Code by the end of March 2022. In its guidance, the Commission proposes:

- Larger signatory participation with tailored commitments corresponding to the size and nature of services, including the online advertising ecosystem and private messaging services;
- Demonetising of disinformation, including exchanging information on disinformation ads;
- Ensuring integrity of services, including reducing manipulative behaviour;
- Improving the empowerment of users, including accessible, effective tools and procedures to flag disinformation, and appropriate and transparent mechanisms to appeal and seek redress;
- Increasing coverage of fact-checking and providing increased access to data to researchers;
- Creating a more robust monitoring framework based on clear key performance indicators (KPIs) measuring the results and impact of actions by signatories.

## Republic of Ireland: Online Safety and Media Regulation Bill 2022

**Type** Regulatory

**Status** Awaiting House approval

On 12 January 2022, the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media initiated the long-awaited Online Safety and Media Regulation Bill. The Bill:

- Introduces an online safety regime focused on “harmful online content” (see Section 2 below), and binding online safety codes that deal with content display and moderation;
- Establishes a new Media Commission, which will appoint an Online Safety Commissioner to act as the regulator, replacing the existing Broadcasting Authority of Ireland;
- Creates powers of enforcement, including imposing financial penalties of up to €20 million or 10% of annual turnover, issuing content limitation notices to designated online services, and blocking access to certain online services;
- Transposes the revised Audiovisual Media Services Directive into Irish law, including video-sharing platform services, into the regulatory framework for online safety.

## New Zealand: Netsafe Code of Practice for Online Safety and Harms

**Type** Non-government, self-regulatory

**Status** Public feedback period closed

On 2 December 2021, NetSafe – New Zealand’s independent online safety organisation – published the Draft Aotearoa New Zealand Code of Practice for Online Safety and Harms for public feedback by 2 February 2022. The draft Code was developed in consultation with digital platforms including Meta, Microsoft, Google, Twitter, TikTok, and Twitch alongside Netsafe. The self-regulatory best-practice framework introduces:

- A set of guiding principles, commitments, outcomes and measures focused on “harmful content themes” (see Section 2 below);
- Oversight powers for the Administrator and a multi-stakeholder Sub-committee to provide monitoring and oversight of Signatories’ commitments and review of the Code;
- A complaints mechanism that enables the public to report breaches by Signatories of their Code commitments;
- Sanctions including the termination of a Signatory due to repeated non-compliance with the Code, or publicly naming a Signatory for failing to meet its commitments;
- Annual compliance reports outlining actions and measures taken in relation to Signatories’ commitments under the Code, which will be made public and open for scrutiny;
- A biennial review of the Code.

## US: Banning Surveillance Advertising Act of 2022

**Type** Regulatory

**Status** Introduced in the Senate and the House

On 18 January 2022, Reps. Anna Eshoo (D-CA) and Jan Schakowsky (D-IL) in the House and Cory Booker (D-NJ) in the Senate introduced the Banning Surveillance Advertising Act of 2022. The Bill applies to any “advertising facilitator” that “receives monetary consideration or another thing of value to disseminate an advertisement to an individual, connected device, or group of individuals or connected devices; and collects or processes personal information with respect to the dissemination of the advertisement”. The Bill:

- Prohibits any targeting based on personal information that identifies or acts as a reasonable proxy for identifying an individual as a member of a protected class such as race, colour, national origin, religion, sex, age, or disability;

- Allows for contextual advertising that is disseminated based on information that the individual is viewing or with which the individual is otherwise engaging; or for which the individual searched; and is displayed in close proximity to that information. Targeting based on location associated with the individual (or device) is not prohibited;
- Empowers the Federal Trade Commission and state Attorneys General with the authority to enforce the new rules for targeting;
- Empowers individual users to sue platforms, if they break the law, granting up to \$5,000 USD in relief per violation.

## US: Eliminating Abusive and Rampant Neglect of Interactive Technologies Act (EARN IT Act) of 2022

**Type** Regulatory

**Status** Reintroduced in the Senate and the House

On 31 January 2022, Senators Richard Blumenthal (D-CT) and Lindsey Graham (R-SC) [reintroduced the EARN IT Act of 2020](#) in the Senate. A few days later, Representatives Ann Wagner (R-MO) and Sylvia Garcia (D-TX) [reintroduced the Bill in the House](#). In brief, the Bill:

- Amends Section 230 of the Communications Decency Act of 1996 so that interactive computer services cannot use Section 230 as a defence in Court cases involving child sexual abuse materials;
- Establishes a National Commission on Online Child Sexual Exploitation Prevention that would be responsible for developing voluntary best practices, comprised of 19 members of the Department of Justice, Federal Trade Commission, and Department of Homeland Security;
- Excludes end-to-end encryption as “an independent basis for liability” of a service provider when considering evidence of actions or circumstances in Court.

In a previous session of Congress, the Earn IT Act passed the Senate Judiciary Committee in July 2020 by unanimous vote but was not passed by the full Senate or the House. Most recently, in a mark-up hearing on 10 February 2022, the Senate Judiciary Committee [voted](#) to refer the Bill to a floor vote. It is supported [by more than 240 groups](#), including [the National Center for Missing & Exploited Children](#). However, the Bill has drawn significant criticism from [more than 60 human rights organizations](#), including [the American Civil Liberties Union](#) and [the Electronic Frontier Foundation](#), who warn that the Bill significantly undermines encryption technology by exposing platforms to more liability simply for utilizing such technology.

For a comprehensive list of proposed amendments to Section 230 of the Communications Decency Act, please consult the [Slate 230 Tracker](#).

## Section 2 Topic-specific snapshot: “Introducing risk assessments for online platforms: Scope and definitions”

Many recent government proposals adopt platform regulation that address “risks” rather than just the removal of content related to specific harms. In December 2020, the European Commission moved ahead with the Digital Services Act (DSA) to introduce a legal framework for risk assessments of very large online platforms. A year later, the UK government presented its approach in the draft Online Safety Bill (OSB). Other liberal democracies, including the Republic of Ireland and New Zealand, have also published proposals that include “risk-based” mechanisms and instruments.

The overview below outlines services and risks included in selected proposals, ranging from regulatory to self-regulatory approaches. Risk assessments are generally considered to go beyond content-based approaches to safeguard both freedom of expression and online safety. A thoughtful discussion of the proposed definitions, scope and methodologies is still required to reflect on the anticipated outcomes, prevent infringements of fundamental rights, and reduce the risk of abuse of similar approaches by authoritarian regimes.

### EU: Digital Services Act (DSA)

Text in ***bold/italic***: Amendments adopted by the European Parliament

#### Services in scope of risk assessments

“Online platforms”: provider of a hosting service which, at the request of users, stores and disseminates information to the public.

#### Methodology:

Online platforms which provide their services to a number of average monthly active users in the EU equal to or higher than 45 million, calculated in accordance with the methodology set out in delegated acts (“Very large online platforms”).

***A methodology shall take into account: number of active users based on each service individually; active users connected on multiple devices are counted only once; indirect use shall not be counted; active users are assigned solely to the online platform closest to the user; automated interactions are not included.***

#### Risks in scope of risk assessments

#### Categories (“systemic risks”):

- Dissemination of illegal content ***or content that is in breach with their terms and conditions***;
- Any ***actual and foreseeable*** negative effects on fundamental rights, ***including for consumer protection***, to respect for ***human dignity***, private and family life, ***the protection of personal data***, freedom of expression and information, ***as well as to the freedom and the pluralism of the media***, the prohibition of discrimination, ***the right to gender equality*** and the rights of the child);
- ***Any malfunctioning*** or intentional manipulation of the service, including inauthentic use or automated exploitation ***or risks inherent to the intended operation of the service, including the amplification of illegal content, of content that is in breach with their terms and conditions or any other content*** with an actual or foreseeable negative effect on the protection of public health, minors, ***and other vulnerable groups of recipients of the service, on democratic values, media freedom, freedom of expression and civic discourse***, or actual or foreseeable effects related to electoral processes and public security;
- ***Any actual and foreseeable negative effects on the protection of public health as well as behavioural addictions or other serious negative consequences to the person’s physical, mental, social and financial well-being.***

**Risks in scope of risk assessments**  
(continued)

**Methodology:**

- **Effectively and diligently** identify, analyse and assess, **and in any event before launching new services, probability and severity of** significant systemic risks stemming from **the design, algorithmic systems, intrinsic characteristics**, the functioning and use of services;
- **Take into account risks per Member State, in particular to a specific language or region;**
- Be specific to the services **and activities, including technology design, business-model choices;**
- Take into account **whether and** how content moderation **systems, terms and conditions, community standards, algorithmic** systems, recommender systems and systems for selecting and displaying advertisement **as well as the underlying data collection, processing and profiling** influence any of the systemic risks

 Republic of Ireland: Online Safety and Media Regulation Bill 2022

**Services in scope of risk assessments**

Audiovisual on-demand media services and designated “online services”: information society services on which user-generated content is made available;

- Designated online services may include: social media services, public boards and forums, online gaming services, ecommerce services, private communication services, Private online storage services, online search engines, and, internet service providers

**Methodology:**

Designation of online services will have regard to nature and scale of the service, levels of availability and risk of exposure to harmful online content when using the service.


**Risks in scope of risk assessments**

**Categories** (“harmful online content”):

- Offence-specific online content;
- Cyberbullying material;
- Material encouraging or promoting eating disorders,
- Material encouraging or promoting self-harm or suicide;
- Non-offence specific online content subject to a “risk test”: risk to a person’s life or risk of significant harm to a person’s physical or mental health, where the harm is reasonably foreseeable.

**Methodology:**

More categories of harmful online content to be created through a process: Specification will have regard to levels of risk of harm, from the availability of content or of exposure to it, the impact of automated decision-making in relation to content delivery and content moderation.

 New Zealand: Code of Practice for Online Safety and Harms (by Netsafe)

**Services in scope of risk assessments**

Products and services that facilitate user-generated content (including sponsored and shared) and are delivered to end-users based in Aotearoa New Zealand.

- The Code may also apply to other digital products or services where the spread or prevalence of “harmful content online” are a concern.

**Methodology:**

Signatories may choose to specify which products and services — as well as commitments, outcomes and measures — that are most relevant for the purposes of the Code in the Signatory Participation Form, having regard to greatly varying incidence, risk level, and impact of user activity and content.

**Risks in scope of risk assessments**

**Categories** (“harmful content online”):

- Child sexual exploitation and abuse;
- Bullying or harassment;
- Hate speech;
- Incitement of violence;
- Violent or graphic content;
- Misinformation;
- Disinformation.

**Methodology:**

Focus on the Signatories’ architecture of systems, policies, processes, products and tools established to prevent the spread of potentially harmful content. Signatories may make commitments to the Code that best matches their risk profiles, either for the company or for specific products/services.

 UK: Draft Online Safety Bill (OSB)

Text in ***bold/italic***: Joint Committee report recommendations

**Services in scope of risk assessments**

User-to-user services and search services;

- “Category 1 services”: likely to include the largest user-to-user service

**Methodology:**

Ofcom must develop risk profiles for different kinds of regulated services, categorising the services as Ofcom consider appropriate, taking into account (a) the characteristics of the services, and (b) the risk levels and other matters identified in the risk assessment”. These characteristics include the functionalities of the service, its user base, business model, governance and other systems and processes.

***Categorisation of services should adopt a more nuanced approach, based not just on size and high-level functionality, but factors such as risk, reach, user base, safety performance, and business model (risk profiles).***

**Risks in scope of risk assessments**

**Categories:**

- Illegal content: terrorism content, CSEA content, priority illegal content, and other illegal content;
- Content harmful to children;
- Content harmful to adults: material risk of the content having, or indirectly having, a significant adverse physical or psychological impact on an adult of ordinary sensibilities.

**Risks in scope of risk assessments**

(continued)

**Methodology:**

- Level of risk encountering above content, taking into account risks presented by algorithms, and the way that the service indexes, organises and presents search results;
- Level of risk of functionalities facilitating the presence or dissemination of relevant content;
- Different ways in which the service is used;
- How design and operation of the service (including business model, governance and other systems and processes) may reduce or increase the risks identified.

*Ofcom should have the powers to set minimum quality standards of risk assessment, under which service providers will be required to undertake independent audits of their systems, processes and algorithms. Required content of risk assessments should follow the risk profiles developed by Ofcom, which in turn should be based on characteristics of the service, platform design, risk level, and the service's business model and overall corporate aim. End-to-end encryption should be identified as a specific risk factor in risk profiles and risk assessments.*

**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 German Federal Foreign Office for their support for this project.