Reset.

Written evidence submitted by Reset (OSB0203)

Online Safety Bill (2)

September 2021

Introduction

One of the Committee's areas of focus is understanding how the UK's draft Online Safety Bill compares to other equivalent initiatives globally. To support that analysis, this document sets out an international comparison of online safety regulations in the UK, EU, Ireland, Australia, Canada and Germany. It also offers a summary of regulations in the US which cover algorithmic harm. In doing so, we hope to provide the Committee with an insight into how the draft Bill fits into the global online safety environment.

Headline comments

- **Content in scope**: The unequivocal inclusion of "legal harms" is the main provision that makes the UK's Online Safety Bill world leading. The DSA and the Australian legislation include some legal harms but through less explicit language than that in the OSB. On disinformation specifically, it is unclear whether disinformation would be covered by OSB, but it is likely to be in scope of the DSA and possibly the Australian regulation. Most other regulations focus just on illegal content.

- **Systems vs content takedown**: the OSB and the DSA have the most distinct focus on *systems* to reduce harm compared to the provisions in other regulations. They are much less prescriptive than other laws. This is in part because many of the other regulations narrow in on illegal content and so focus on takedown/deletion. Whereas the OSB requires services to "minimise" the presence of illegal content and take it down "swiftly" when it is reported, some other laws - particularly those in Germany, Australia and Canada - set out strict takedown requirements such as specific time frames to remove content. In its systems approach, the OSB requires services to account for risks of certain functionalities, such as algorithms, in spreading such content. There is similar language in the DSA and the bill in Ireland, but the OSB is the most explicit of all in focusing on systems and processes. However, there is a tension with "content that is harmful to adults", where the UK Bill defers content management to companies' terms of service. In this regard, the DSA has the edge over the OSB.

- **Services in scope**: the OSB covers a broad range of services in scope, but exempts ISPs which have obligations in many of the other regulations. Other regulatory initiatives include a more discrete and explicit definition of "social media" (DSA, Australia, Germany) than that in the OSB, which captures many firms outside of social media. The draft OSB seems to go further than any other regulation by potentially including private messaging in scope.

- **Definition of harm**: the OSB appears unique in its fixed definition of individual harm. The EU, Canadian and Irish regulations include a societal element of harm either by including electoral processes, social cohesion or other definitions. They seem to go further on tackling collective harm than the OSB. The Australian bill has potentially the lowest threshold of harm, including in the adult cyberbullying definition content which is "menacing, harassing or offensive" and intended to cause serious harm.

- **Powers of the regulator**: broadly similar powers, including fines and investigatory powers. The DSA mandates algorithmic audit and there is some language in the OSB which appears to give Ofcom similar powers but which is less explicit.

- **Independence of the regulator:** the OSB is unique in allowing political involvement in this agenda, via the SoS provisions. All other regimes create or promote regulatory independence. In some cases, Online Safety Commissioners are appointed to oversee regimes.

- **Transparency**: transparency underpins all regimes. The frequency of reporting varies, with the OSB less frequent than the DSA. The DSA and US regulations mandate data sharing with researchers, which is not a provision in the OSB, and the DSA mandates that transparency reports must be published to the public and not only to the regulator. It is unclear in the OSB whether the same transparency requirements apply.

- **Advertisements**: the DSA and NetzDG have transparency requirements which apply to ads, as do the US regulations which mandate certain standards for ad libraries.

- **Journalism and News:** the OSB includes the broadest and most explicit exemption for news media. It is the only one which includes a duty to protect news content. Where similar exemptions for news media are in other bills, they are much less defined and do not include freedom of expression provisions.

- **User identification:** no regulations explicitly mandate age-verification (although it is alluded to in the OSB) and none include provisions to remove anonymity. In fact, a German law includes provisions to protect anonymity online.

Reset.

SUMMARY PAGE

	UK - Online Safety Bill	EU - Digital Services Act	Ireland - Online safety	Australia - Online Safety	Canada - Online safety	Germany - NetzDG (and others)	US - AJOA and Social Media DATA Act
SYSTEMS VS TAKEDOWN	Systems + Takedown	Systems + Takedown	Systems + Takedown	Takedown	Takedown	Takedown	Systems
CONTENT IN SCOPE	Illegal and legal List of harms to be added later but unclear whether disinformation is in scope.	Illegal and, indirectly, legal Disinformation included indirectly	Illegal and legal Disinformation out of scope	Illegal and legal Disinformation out of scope Intimate images in scope	Illegal Disinformation out of scope Hate speech and intimate images in scope.	Illegal Intimate images in scope	Not a content agenda - focused on data transparency and algorithmic processes/bias
SERVICES IN SCOPE	Services which host or facilitate UGC, apart from news media outlets. Private messaging in scope.	Intermediary services e.g. ISPs and online platforms Private messaging out of scope	Broad range of platforms and services inc press publications which enable UGC Private messaging in for criminal content	Social media services, Relevant electronic service and ISPs (Tight definition of “social media”)	Social media Private messaging out of scope	Social media	Broad range of platforms and sites

Reset.

DEFINITION OF HARM	Individual harm - physical and psychological	No set definition Focus on rights. Includes societal harm.	Varied. Threshold could be considered lower than that in OSB	Largely Individual. Includes "offensive" material to adults	Societal and individual	Criminal law	Algorithmic discrimination against protected characteristics.
POWERS OF REGULATOR	Fines Information gathering powers Language seems to allow algorithmic inspection	Fines Information gathering powers Algo audit mandatory	Fines Information gathering powers. No algo audit	Fines Offers public facing complaint mechanisms, Investigation, Audit	Information gathering powers Inspection powers No algo audit	Fines	Data access and algo audit
INDEPENDENCE OF REGULATOR	Independent however OSB keeps provisions for political agenda setting	Independent as well as EC oversight of large platforms	Independent Creates Online Safety Commissioners	Independent	Independent Creates Digital Safety Commissioner	Independent	Independent Co-reg task force
TRANSPARENCY	Annual transparency reports No data sharing provisions	Six monthly transparency reports (publicly published) Data access for	Periodic transparency reporting	Periodic transparency reporting.	Transparency reporting inc data on takedown volumes and processes.	Transparency reporting.	Data access for researchers Transparency about algo processes

		pre-vetted researchers					
NEWS MEDIA	Out of scope (explicit) Distinct provisions to express free expression of the press	Out of scope (implied)	Included as the Bill also sets up a media regulator. No harm reduction obligations.	Out of scope (implied)	Out of scope (explicit)	MStV makes explicit provisions for protecting news media on platforms	N/A
ADVERTISEMENTS	Out of scope	In scope (transparency requirements)	Out of scope	Out of scope	Out of scope.	Included in MStV (transparency requirements)	Included in Data Act which focuses on transparency
USER IDENTITY	AV in Anonymity out	Not included	Not included	Not included	Not included	AV covered in JMStV Anonymity protected in TMG	N/A

Reset.

UK - Online Safety Bill

	DETAILS	COMMENTARY
APPROACH (SYSTEMS VS CONTENT TAKEDOWN)	Combination of systems approach as well as a takedown/content regime. **SYSTEMS APPROACH (Clauses 9 and 10)** The Bill creates three categories of harm, each of which has different risk management requirements. The categories are: - Illegal content - Services likely to be accessed by children - Content that is harmful to adults (harmful but not illegal). For each category of harm, companies must carry out risk assessments and adhere to "safety duties". The risk assessments for all categories of harm must account for the systems which promote harmful content, including algorithms; functionalities disseminating content; how the design and operation of the service (including the business model) may influence risk.	The approach in the draft Bill differs from that of the original Online Harms White Paper which set out plans for a single duty of care which services in scope would have to introduce to protect users. The ambition was to focus less on content and more on upstream preventative measures to mitigate harm. Clause 11 is the most content-focused of the three duties, deferring the content management to platforms/services.

	The safety duties for <u>illegal content</u> are duties to (Clause 9): 1. Take proportionate steps to **mitigate** and effectively manage the risk of harms to individuals 2. **Use systems and processes to minimise** the presence of illegal content on their platform and "swiftly take down" such content when it is reported. The safety duties for <u>services likely to be accessed by children</u> are to (Clause 10): 1. Take proportionate steps to ***mitigate*** and effectively manage the risk and impact of harms to children in different age groups 3. ***Prevent*** children of any age from encountering certain content 4. **Protect** children in age groups judged to be at risk of harm from encountering harmful content. <u>The above duties apply to **all services in scope.**</u> **CONTENT APPROACH (Clause 11)** The safety duties for <u>content that is harmful to adults</u> **apply only to Category 1** companies and include duties to (Clause 11): 1. Specify in the terms of service how content will be dealt	

	with by the service 2. Ensure that their terms of service are clear, accessible and applied consistently. There is no requirement in Clause 11 for Category 1 companies to use "systems and processes" as there are in Clauses 9 and 10.	
CONTENT IN SCOPE	**User generated content** The Bill focuses on user-generated content (UGC) uploaded or shared on a service. The Bill itself does not list specific types of content in scope but rather defines categories of harmful content (illegal; content on services likely to be accessed by children; content that is harmful to adults). The regulator, Ofcom, will later produce a list of harms in each category. In addition, there are powers reserved for the Secretary of State to define primary priority areas of harm and "in special circumstances" (Clause 112) to instruct the regulator to address specific threats on the grounds of health or public safety or national security. **Content of democratic importance (Clause 13)** There are specific provisions in the Bill for Category 1 companies to "protect content of democratic importance". Such content includes news publisher content or "content that appears to be specifically intended to contribute to democratic political debate in the United Kingdom". For this content, services have a duty to use "systems and processes" to ensure the democratic	Because the approach is not to list the harms in scope at this stage, it is unclear exactly which harmful content will be included. There is much campaigning for content in scope to be extended to include online scams and fraud and paid-for advertisements, as well as to explicitly commit to including certain types of harm such as racist abuse, violence against women and girls, and disinformation. Disinformation is singled out as a harm to be considered via an expert advisory committee that is to publish a report on how the Bill and Ofcom should tackle disinformation. The report must be published within 18 months of the committee's establishment.

	importance of this content “is taken into account” when making decisions about how to treat such content “especially decisions about whether to take it down”. **Content out of scope** Content not in scope of the Bill includes: (a) emails (b) SMS messages (c) MMS messages (d) comments and reviews on provider content (e) one-to-one live aural communications (f) paid-for advertisements (g) news publisher content	
SERVICES IN SCOPE	Any business in the world that is accessible by people in the UK and hosts user- generated content, allows users to create content, or allows users to interact with one another is in scope. This will include the likes of video games, instant messaging platforms and online marketplaces. The Bill separates services into three categories: ● **Category 1:** Companies with user-to-user services where the risk of harm is considered to be the greatest. This is *expected* to include social media firms such as Facebook, Twitter and YouTube. The definition is not purely based on size of service, but also functionality and features it offers which may increase the risk of harm.	The intention in tiering the services in scope is to avoid overburdening small services/companies. There has been much push back from the startup sector about the unintended consequences of the Bill which the Government has tried to correct. Which companies fall into what category is still tbd, but the inclusion of functionalities as a factor in determining categories, rather than just size, means it might not just be large platforms

	• **Category 2A:** Companies with search services such as Google. These companies have to comply with fewer obligations, notably not having to account for "legal but harmful" content. • **Category 2B:** Companies with user-to-user services where the risk of harm is lower than Category 1 companies. "Private" communications are in scope of the Bill. Clause 64.4.b gives the regulator the power to use technology to investigate CSEA material on public or private services. In addition the definition of content in Clause 137 "means anything communicated by means of an internet service, **whether publicly or privately**, including written material or messages, oral communications, photographs, videos, visual images, music and data of any description".	in Category 1. This accounts for the prevalence of harm on emergent services and accounts for the algorithmic promotion of content as a factor in amplifying harm. The inclusion of private communications would mean a revisiting of E2E encryption, allowing the government to inspect private communications where appropriate. This is widely considered a major privacy breach and will be a hot topic of debate as the Bill progresses. In comparison, the DSA has an additional focus on intermediary services, such as ISPs, which feature less heavily in the OSB.
DEFINITION OF HARM	Harmful content that is not illegal will have to fit the **definition of "having, or indirectly having, a significant adverse physical or psychological impact" on an adult or a child "of ordinary sensibilities" (clauses 45 - 46).** There are no explanatory notes setting out how the standards in this clause ("ordinary sensibilities") should be interpreted, and the threshold for "significant" is also yet to be defined. The definition intentionally narrows down to individuals so as to exclude societal and	The focus on individual vs collective/societal harm is at odds with the White Paper and differs from the DSA which, in Article 26, accounts for harm to electoral processes and infringement of human rights.

	democratic harms. **The Government has explicitly stated that it does not intend for this bill to tackle democratic harm,** which was included in the 2019 White Paper**.**	
POWERS OF REGULATOR	The regulator will have a range of powers such as: ● Information gathering powers (cl70); Investigations powers (cl75) including power to require interviews (cl76) powers of entry and inspection (cl77). ● Enforcement powers including directions for improvement (cl80), notices of non-compliance, and fiscal penalties like civil fines (cl85) up to 18 million of 10% of worldwide revenue, and business disruption measures (cl.91). ● **Algorithm audit capabilities *appear* to be provided for as draft OSB requires companies** "design and assess the service [...] including with regard to (i) algorithms used by the service," (cl30(2)). Those algorithms can be audited by Ofcom who have the power to require information (cl72) (including generating material), and commissioning external reports by skilled persons (cl74) to audit on Ofcom's behalf. There are a wide range of purposes to use those inspection powers including "assisting OFCOM in identifying and assessing a failure, or possible failure," (cl74(a)).	The inclusion of language which gives Ofcom powers to inspect the algorithms of services is encouraging, but needs clarifying. Ofcom does not seem to think the language in the draft Bill gives it unequivocal powers for algo audit whereas many others in the community think it does. Ofcom does not appear to have the power to push back if services' risk assessments (undertaken as part of their duties) are inadequate. Many are calling for Ofcom to be given this power and for risk assessments to have minimum standards. Including this power would bring the OSB closer in line with the DSA.
INDEPENDENCE OF REGULATOR	The regulator will be Ofcom which is a statutory body independent of government, with a unitary board, Chairman and	The powers given to the Secretary of State are unprecedented not

	Chief Executive; with a number of sub-committees/boards on specific issues. While Ofcom is regarded as, and has shown itself to be, independent of government, **there are some aspects of the draft Bill which call into question how independent Ofcom will be able to be in fulfilling its functions under the act.** The draft bill includes provisions for the Secretary of State to: 1) direct OFCOM to make amendments to the code to reflect Government policy (cl 33); 2) set strategic priorities which OFCOM must take into account (cl 109 and cl 57); 3) set priority content in relation to each of the safety duties (cl 41 and 47).	only in the UK but also as compared to other online safety regulations. They undermine the independence of the UK's regime and cause unnecessary uncertainty for companies in scope. More details about the powers here.
TRANSPARENCY	**Regulated companies are required to provide annual transparency reports (cl49)** responding to a notice provided by Ofcom setting out what has to be included within the transparency reports (cl49(4)) including information about the incidence of illegal content, how terms of services are applied, systems and process in place for user reporting, risk management, among many other things. Ofcom then themselves must provide transparency reports (cl100) setting out conclusions from those compelled transparency reports. Ofcom must prepare a report about researchers' access to information (cl101) and may from time to time produce reports about online safety matters (cl102).	The transparency provisions in the OSB are less detailed and prescriptive than those in the DSA. The DSA also requires the **public** publication of transparency reports, not just published to the regulator. **The DSA also includes a requirement for platforms to share data with accredited, or 'pre-vetted' researchers - a transparency provision which is not in the OSB.**
ADVERTISEMENTS	Not in scope - **paid for advertisements explicitly out of scope.**	Whereas the OSB puts all paid ads out of the scope, **there are**

		detailed provisions in the DSA on transparency requirements regarding adverts, and there is an ongoing discussion in the European Parliament to impose more restrictions on different targeting techniques that could be used.
JOURNALISM AND NEWS	Among the services **not in scope are news publisher sites *(including when their content is reshared on social media*** and comments on online news sites (clauses 39 - 40). This means that **news publishers do not need to apply content duties on their sites**, in an understandable attempt to avoid press regulation. The definition of "news publisher content" includes news content and commentary as well as "gossip about celebrities, other public figures or other persons in the news". However, **the definition of "news publisher" is sufficiently broad as to potentially include anyone who sets up an eligible news website in the UK.** The exemption extends to when "a link to a full article or written item originally published by a recognised news publisher" is posted on a Category 1 service. This may mean that **any posts on social media which include a link to a news site are exempt from services' safety duties.** Another layer of provisions are the **carve-outs for journalism and political debate (clauses 13-14)**. These apply only to Category 1	The OSB goes further than equivalent pieces of regulation to carve out protections for news outlets. The provisions for content of democratic importance are less explicit in other regulations, although the DSA does account for the Charter of Fundamental Rights including free expression.

	companies and are further attempts by the Government to avoid over-reach and infringing on freedom of expression, which on the surface appear very encouraging. They require the riskiest platforms to have **distinct processes for accounting for "content of democratic importance" and "journalistic content",** including expedited complaints procedures if journalistic content is considered to have been inappropriately treated.	
USER IDENTITY	**No reference to anonymity.** The draft requires companies to account for and **mitigate against the risks to children "in different age groups"**. There are also provisions for "services likely to be accessed by children". **The implications of such language is that services will need to know the age of their users, and therefore apply age verification measures. (Clause 10).** The Bill also revokes provisions in the 2017 Digital Economy Act which demanded mandatory age-verification checks to be included on all commercial pornography websites, with those which failed to comply to be hit with fines. These requirements were never implemented but would now be officially revoked by the OSB.	

EU - Digital Services Act

	DETAILS	COMMENTARY
APPROACH (SYSTEMS VS TAKEDOWN)	Combination of systems approach as well as a takedown/content regime. **TAKEDOWN APPROACH FOR ILLEGAL CONTENT** The DSA seeks to harmonise the notice and takedowns mechanisms across the EU. The DSA requires the implementation of an easy to access, user-friendly mechanism which allows users to submit electronic notices (Article 14). Additionally, online platforms must provide a complaint and redress mechanism (Article 17) as well as an out-of-court dispute settlement system (Article 18). They must also give priority to notifications of entities that have been qualified as so-called trusted flaggers by the authorities (Article 19) and suspend repeat infringers (Article 20). Providers of intermediary services are obliged to act upon orders received from national judicial or administrative authorities to take down illegal content (Article 8).	The DSA is more explicit than the OSB about the sort of redress mechanisms and systems services must deploy to remove illegal content.

	SYSTEM APPROACH The DSA foresees specific rules for very large online platforms when they reach more than 10% of 450 million users in Europe. Such platforms have additional obligations, including assessing the systemic risks stemming from the functioning, use and potential misuses of their services. Three categories of systemic risks should be assessed in-depth: - the dissemination of illegal content (Article 26.1.a) - negative effect for the exercise of fundamental rights, as protected by the Charter of Fundamental Rights, including the freedom of expression and information, the right to private life, the right to non-discrimination and the rights of the child (Article 26.1.b) - intentional manipulation of their service, including by means of inauthentic use or automated exploitation of the service, with an actual or foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security (Article 26.1.c). Very large online platforms must then put in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified (Article 27). These platforms are then subject to yearly audits by independent organisations on the assessment of their obligations, including the mitigations measures (Article 28).	

CONTENT IN SCOPE	**From a liability perspective**, the DSA focuses on illegal content, which it defines as "any information, which, in itself or by its reference to an activity, including the sale of products or provision of services is not in compliance with Union law or the law of a Member State, irrespective of the precise subject matter or nature of that law" (Article 2.g). Illegal content should be understood as "information, irrespective of its form, that under the applicable law is either itself illegal, such as illegal hate speech or terrorist content and unlawful discriminatory content, or that relates to activities that are illegal, such as the sharing of images depicting child sexual abuse, unlawful non-consensual sharing of private images, online stalking, the sale of non-compliant or counterfeit products, the non-authorised use of copyright protected material or activities involving infringements of consumer protection law" (recital 12). **From a risk assessment perspective**, the DSA covers illegal content as well as legal but harmful content such as disinformation and hate speech. The term 'harmful content' is not explicitly mentioned but is encompassed in the risk assessment obligations of Article 26.1 b and c (see above).	**Includes illegal and, indirectly, legal content including disinformation. In opposition to the OSB, the focus is on taking down illegal content and using risk assessments for legal content. OSB requires "systems and processes" for illegal content, but defers to T&Cs for legal but harmful content.**
SERVICES IN SCOPE	The DSA applies to **providers of intermediary services**, and in particular intermediary services consisting of services known as 'mere conduit', 'caching' and 'hosting' services, irrespective of their place of establishment or residence, in so far as they provide services in the Union. The DSA also distinguishes, within the broader category of **providers of hosting services**, the subcategory of **online**	**The focus on intermediary services differs from the UK draft Bill, which excludes ISPs and other intermediaries from the legislation.** Blocking access to sites which fail to comply with the regulations is a power for the UK regulator but there are no

	platforms, which store information provided by the recipients of the service at their request, but also disseminate that information to the public. What is excluded: - **Dissemination of information within closed groups** consisting of a finite number of pre-determined persons such as **messaging and email services.** - Where the dissemination to the public is merely a minor and purely ancillary feature of another service, such as **comments sections in an online newspaper.**	obligations on the ISPs themselves to comply with the safety duties. **There is no suggestion in the DSA that encrypted content would be in scope of the regulations, which differs from the OSB.**
DEFINITION OF HARM	No set definition but, as per Article 26, the risk assessments must account for the infringement of certain rights (26.1.b) and for negative effects on individuals and society (26.1.c) : - the dissemination of illegal content through (Article 26.1.a) - negative effect for the **exercise of fundamental rights,** as protected by the Charter of Fundamental Rights, including the freedom of expression and information, the right to private life, the right to non-discrimination and the rights of the child (Article 26.1.b) - intentional manipulation of their service, including by means of inauthentic use or automated exploitation of the service, with **an actual or foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security** (Article 26.1.c).	The focus on fundamental rights in the DSA is in part reflected in the UK's draft Bill in Clause 12, but the DSA is more explicit on which rights must be protected. **The inclusion of harms to electoral processes and public security in the DSA are not in the UK draft Bill.**
POWERS OF REGULATOR	The DSA grants national Digital Services Coordinators a range of powers: - **Investigation powers**, including to carry out on-site	Similar powers to the OSB. **The audits defined in Article 28**

	inspections, interview staff members and require the production of documents and information (Article 41.1) - **Enforcement powers**, including to order the cessation of infringements, impose interim measures, levy fines (up to 6% of global annual turnover) as well as periodic penalty payments (up to 5% of average global daily turnover), and accept binding commitments (Article 41.2 and 42) As part of the supervision, investigation, enforcement and monitoring of very large online platforms, the DSA grants the same powers to the **European Commission**. The Commission becomes the sole regulator when very large online platforms infringe the DSA. Very large online platforms (VLOPs) are subject to yearly audits to assess their obligations under the DSA. These audits must be performed by independent organisations (Article 28).	**also mandate algorithmic audit performed by independent organisations.**
INDEPENDENCE OF REGULATOR	Oversight and enforcement of the DSA is attributed to **Member States** which will have to appoint at least one national authority as a **Digital Services Coordinator** (DSC). DSCs can be existing national authorities. When exercising their powers they must act with complete independence and remain free from any external influence, whether direct or indirect, and must not seek or take instructions from any other public authority or any private party (Article 39.2). Member States must ensure that their DSC has adequate technical, financial and human resources to carry out their tasks (Article 39.1). The DSA establishes the '**European Board for Digital Services**' (the Board), an independent advisory group composed of the DSCs. It	Much greater independence is awarded to DSCs than to Ofcom, which is subject to steering and guidance from the Secretary of State. The role of the Commission in handling VLOPs has raised some eyebrows. The regulatory model is seen as highly centralised, with the Commission as the sole regulator with strong powers vis-à-vis VLOPs. Questions are raised

	will advise the DSCs and the Commission on the consistent application of the DSA and the efficient cooperation between DSCs. The DSA provides for an enhanced supervision procedure in case of infringements from very large online platforms. DSCs and the Board can request the European Commission to intervene and exercise its investigatory and enforcement powers in such cases or the Commission can choose to do so on its own initiative.	whether the institution is sufficiently resourced to take on this supervisory role.
TRANSPARENCY	**Providers of intermediary services** must include in their **terms of service** information on any restrictions that they impose in relation to the use of their service, including information on policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review (Article 12). They also must publish, at least once a year, clear, easily comprehensible and detailed **reports on content moderation** they engaged in during the relevant period, including: - number of orders received from Member States' authorities - number of notices submitted by users, actions taken pursuant to the notices, and the average time needed for taking the action; - content moderation engaged in at the providers' own initiative including numbers and types of measures taken; - number of complaints received through the internal complaint-handling system including the basis for those complaints, decisions taken in respect of those complaints,	Greater guidance and specificity in the DSA vs the OSB on what must be included in the Terms of Service and transparency reports. Transparency reporting in the **DSA must be every six months, whereas only annually in the OSB.** **The requirement for platforms to share data with accredited researchers is in the DSA but not in the OSB.**

	the average time needed for taking those decisions and the number of instances where those decisions were reversed (Article 13). In addition, **online platforms** must also include in the reports: - the number of disputes submitted to the out-of-court dispute settlement bodies, the outcomes of the settlement and the average time needed for completing the procedures; - the number of suspensions imposed; - any use made of automatic means for the purpose of content moderation, including a specification of the precise purposes, indicators of the accuracy of the automated means in fulfilling those purposes and any safeguards applied. **Very large online platforms** have additional transparency requirements. They must publish their transparency report every six months. In addition, they must publish and make available a report setting out the results of the risk assessment and the related risk mitigation measures identified and implemented, as well as the yearly audit report and implementation report (Article 33). Other related transparency requirements include: - explanation of parameters used in recommender systems in their terms of service (Article 29), - providing access to data to vetted researchers (Article 31).	
ADVERTISEMENTS	Under the DSA, **online platforms that display advertising on their online interfaces have transparency requirements.** They must	**Whereas the OSB puts all paid ads out of the scope, there are**

	ensure that the recipients of the service can identify, for each specific advertisement displayed to each individual recipient, in a clear and unambiguous manner and in real time: (a) that the information displayed is an advertisement; (b) the natural or legal person on whose behalf the advertisement is displayed; (c) meaningful information about the main parameters used to determine the recipient to whom the advertisement is displayed (Article 24). **Very large online platforms (VLOPs) have additional transparency requirements regarding online advertising.** They must compile and make publicly available through application programming interfaces a repository containing: - the content of the advertisement; - the natural or legal person on whose behalf the advertisement is displayed; - the period during which the advertisement was displayed; - whether the advertisement was intended to be displayed specifically to one or more particular groups of recipients of the service and if so, the main parameters used for that purpose; - the total number of recipients of the service reached and, where applicable, aggregate numbers for the group or groups of recipients to whom the advertisement was targeted specifically (Article 33)	**detailed provisions in the DSA on transparency requirements regarding adverts.**
JOURNALISM AND NEWS	Not covered.	
USER IDENTITY	Not covered.	

IRELAND - Online Safety and Media Regulation Bill

	DETAILS	COMMENTARY
APPROACH (SYSTEMS VS TAKEDOWN)	Combination of systems approach as well as a takedown/content regime. The Online Safety and Media Regulation bill is yet to be passed and is currently at the pre-legislative scrutiny phase. The general scheme of the bill seeks to **transpose the amended Audiovisual Media Services Directive** [Directive (EU) 2018/1808] into Irish law and establish a Media Commission (MC) which will regulate audiovisual media services (including **designated online services).** The bill provides for the MC to create rules and online safety codes, as yet unwritten but provided for under **Head 50A**, to be observed by (i) audiovisual media services, (ii) sound media services and **(iii) designated online services (as designated by the MC)**; and to issue guidance materials and advisory notices in relation to harmful online content and age-inappropriate online content.	A much broader piece of regulation which aims to implement the AVMSD as well to create an online harms agenda. Ireland will of course eventually comply with the DSA so these should be seen as preemptive online safety measures. The Bill creates an Online Safety Commission, and Commissioners, who will be responsible for creating binding online safety codes which include provisions for reducing the spread and amplification of harmful content. The services which fall under this scope, determined by the MC, will face sanction if they fail to

	The idea is that the codes **will reduce the spread and amplification of "harmful online content" because designated online services** will be required to develop measures to meet the principles set out in the codes that apply to them. **The MC will assess whether these measures are working through information requests, investigations and audits. On the basis of these the MC can then issue directions, through compliance and warning notices, to online services mandating them to take specific steps to improve their compliance with the codes.**	observe these codes.
CONTENT IN SCOPE	This Bill states that the content it seeks to have either removed or, in certain circumstances, blocked, includes material which is already subject to criminal law and cannot be legally disseminated. The bill states that this includes child sexual abuse material; content containing or comprising incitement to violence or hatred; and/or public provocation to commit a terrorist offence (**Head 49)**. The bill also refers to content which encourages and/or promotes **eating disorders**, and content which encourages and/or promotes **self-harm and/or suicide**. It also seeks to regulate content to help prevent and/or **stop cyberbullying. This includes, "material which is likely to have the effect of intimidating, threatening, humiliating or persecuting a person to which it pertains and which a reasonable person would conclude was the intention of its dissemination"**.	Illegal and legal content included. Intentions are to keep the harms in scope flexible.

	Disinformation does not feature in the scheme of this bill and is not included as a category of harmful online content.	
SERVICES IN SCOPE	● Audiovisual media services, including audiovisual broadcasting services and on-demand audiovisual media services; ● Sound media services, including sound broadcasting services; ● The range of services which the MC will potentially be able to designate as a "designated online service" will emanate from **a pool of online services which facilitate the dissemination of or access to user-generated content**. ○ These services include, but are not limited to, **video-sharing platform services (for example YouTube or TikTok for the whole EU, as the revised Directive follows the internal market country of origin principle of the EU, meaning that any service established in Ireland will be regulated by Ireland on behalf of the whole EU)**, social media services; public boards and forums; online gaming services; e-commerce services, where they facilitate the dissemination of or access to user-generated content; private communication services; private online (cloud) storage services; **press publications, where they facilitate the dissemination of or access to user-generated content;** online search engines; and internet service providers. ○ *In respect of private communication services and private online (cloud) storage services, **the Bill states that the MC's powers will "be explicitly**	Broad range of services in scope, including press publications which enable UGC. Because this implements the AVMSD, regulations extend to broadcast media. It is unclear how the MC will regulate private communications for criminal content.

	limited to matters relating to content which it is a criminal offence to disseminate". An explanatory note under **Head 58** states that "it is not intended to penalise individuals who unwittingly create small-scale On-Demand Audiovisual Media Services (ODAVMS) where the risk of harm from such services remains low. Instead the regulator **will take a risk-based approach to the regulation of small-scale services.**	
DEFINITION OF HARM	**Head 49A** of the scheme of the bill provides that "harmful online content" includes: (a) material which it is an criminal offence to disseminate under Irish [or European Union law], **(b) material which is likely to have the effect of intimidating, threatening, humiliating or persecuting a person to which it pertains and which a reasonable person would conclude was the intention of its dissemination,** **(c) material which is likely to encourage or promote eating disorders and which a reasonable person would conclude was the intention of its dissemination, and,** **(d) material which is likely to encourage or promote [self-harm or suicide] or provides instructions on how to do so and which a reasonable person would conclude was: (i) the intention of its dissemination and (ii) that the intention of its dissemination was not to form part of philosophical, medical and political discourse.**	The Irish Bill explicitly states that, on the one hand, it does not want to create a static definition of harm and prefers to rely on lists of harmful content; however on the other hand sets a threshold which includes "**material which is likely to have the effect of intimidating, threatening, humiliating or persecuting a person to which it pertains and which a reasonable person would conclude was the intention of its dissemination".** The defining language could be considered a "lower threshold" of harm than the UK draft OSB.

	In addition, "harmful online content" specifically does *not* include: (a) "material containing or comprising a defamatory statement, (b) material that violates data protection or privacy law, (c) material that violates consumer protection law, and (d) material that violates copyright law." **Head 49B** provides for the **MC to propose to include or exclude further categories of material from the definition of harmful online content, publish these proposals, invite submissions from interested parties and, subsequently, bring the proposals to the minister and recommend they be adopted by the Government**. The Minister may then, by regulation, include or exclude the proposed categories of material from the categories considered to be harmful online content. **Head 49C** provides that "age inappropriate online content" means material "which may be unsuitable for exposure to minors and that they should not normally see or hear and which may impair their development, taking into account the best interests of minors, their evolving capacities and their full array of rights, and includes: (a) "material containing or comprising gross or gratuitous violence, (b) material containing or comprising cruelty, including mutilation and torture, towards humans or animals, and,	

	(c) material containing or comprising pornography." An explanatory note under **Head 49A** states: "It is not proposed to define harmful online content as a singular concept as it has not been possible to arrive at a suitable, broad, and principle based description of the meaning of this phrase. Instead, it is proposed to enumerate definitions of categories of material that are considered to be harmful online content."	
POWERS OF REGULATOR	The core powers of the Commission, under **Head 11**, are listed as, but not limited to the power to: ● issue notices and warnings, ● to devise, implement, monitor and review codes, including codes of practice, ● to conduct investigations and inquiries, and for the necessary powers to be conferred on the MC to conduct such investigations and inquiries, ● to appoint authorised officers to carry out investigations and to confer such authorised officers such powers as are necessary to fulfil their duties, ● to impose administrative financial sanctions, subject to court confirmation, and the power to enter into settlement arrangements, ● to prosecute summary offences, ● to convey licenses to television broadcasting services, ● to operate a registration system for on demand audio-	The powers granted to the MC are akin to those granted to Ofcom although they extend outside of the online harms regime as it is being established as a new media regulator. As per the next section, the Bill also confirms the creation of an Online Safety Commissioner. This is not included in the UK OSB. Head 53 also gives the MC powers to require platforms to take down individual pieces of content. No clear inclusion of algo audit powers.

	visual media services. In respect of the MC's powers to demand information from online services, **Head 50B** provides that the MC "**may request information from any designated online service regarding their compliance with any online safety code**" and that such a service "shall comply with information requests". Under **Head 15B**, an authorised officer will have the power to search and inspect a premises if they have "reasonable grounds for believing documents, records, statements or other information relating to [a relevant regulated activity] is being kept"; and to inspect such documents or obtain them through legal channels and other means.	
INDEPENDENCE OF REGULATOR	**Head 8** provides that the MC "shall be independent in the performance of its functions". However, as stated above, **Head 49B** would permit the executive to widen or narrow the meaning of harmful content. **Head 10** concerns the specific functions of appointed Commissioners but fails to provide for **the role of the Online Safety Commissioner.** An explanatory note, under Head 10, states: ". . . it should be noted that it is intended that the Commission will formally delegate functions to Commissioners and staff as appropriate. While the delegation of functions is	The appointment of an Online Safety Commissioner is a key facet of the Irish bill, although details are limited. There is some provision for political influence in the Irish Bill as with the UK OSB.

	ultimately a matter for the Commission itself, this provision is desired from a policy perspective as the Minister wishes that individual Commissioners can take responsibility for clearly delegated functions. **This is particularly relevant in the case of the Online Safety Commissioner." This is the only express reference to the Online Safety Commissioner in the general scheme of the bill.**	
TRANSPARENCY	Under **Head 13**, regulated entities will be required to provide periodic reports on their compliance or otherwise with codes. After receiving the report, the MC can, under **Head 15E**, impose a sanction; take no further action; can cause for further investigation to be carried out; hold an oral hearing (the rules for which will be made by the MC); and make a decision in respect of any sanction. The decision, and reasons for the decision, will be made in writing to the regulated entity "as soon as is practicable" and, if necessary, the sanction to be imposed and the reasons for the sanction. However, **Head 53** provides that the **MC may issue compliance notices to online services, which may require the removal of material posted by individuals**. **Head 53 (2)** only provides that the MC ***may*** invite users or uploaders to make a submission in respect of the material at the centre of a notice. **Head 35 (Reporting by Commission)** provides the reporting duties	Transparency reports as per OSB (but not annually). Reporting to the Minister by the MC. Concerns that Head 53 fails to provide procedural safeguards against an interference with the right to freedom of expression in respect of any decision by a State body to remove material. There are no means for individuals to challenge a decision of the MC.

	of the MC to the minister. These include the provision that, no later than June 30 every year, the MC will prepare and submit to the Minister an annual report on its activities in the immediately preceding year, which will be laid before each House of the Oireachtas [parliament]. **Head 66** provides that the MC will report to the Minister on an annual basis on the operation of Heads on European Works (defined under **Head 63**) quotas and prominence. **Head 72 (3)** provides that the MC will review the effect of a media code or rule "from time to time as it sees fit", and shall prepare a report and furnish the report to the Minister.	
ADVERTISEMENTS	**Head 62** provides that the MC will prepare media codes to be observed by media service providers providing audiovisual media services and sound media services. These include provisions to "protect the interests of the audience" but are not applicable to online platforms or services.	Codes regarding the impact of advertising on children are in the spirit of HFSS regulations in the UK. The codes do extend to broader harm caused by ads other than the provision to "protect the interests of the audience". They do not apply to online platforms.
JOURNALISM AND NEWS	Impartiality requirements included as part of the broadcast function of the MC, and in the transposition of the AVMSD. No provisions for online platforms vis a vis news and journalism.	There are broad news and journalism provisisons in the Irish Bill but only as they relate to establishing a new media regulator.

USER IDENTITY	Not covered.	

AUSTRALIA - Online Safety Bill 2021

	DETAILS	**COMMENTARY**
APPROACH (SYSTEMS VS CONTENT TAKEDOWN)	The Bill has a takedown focus, with the start of a systemic approach through the creation of a co-regulatory 'Basic Online Safety Expectations' code. The Bill is yet to be passed, but widely expected to not change, and consultations around the Basic Online Safety Expectations have just begun. The Bill sets up three types of 'takedown' requirements, for: ● Social media providers to respond to take down notices to remove the five types of content in scope (below). Services have either 24 or 48 hours to respond. ● Internet service providers to: ○ take down links where they connect to 'class 1 extreme pornography/gore' that is in scope within 48 hours ○ block to the domain names, URLs and IP addresses that provide access to the abhorrent violent materials for up to 3 months. ● App distribution services (google play etc) to prevent Australians downloading an app where the app provides for 'class 1 extreme pornography/gore' that is in scope within 48 hours	Generally a much heavier takedown focus than the OSB with detailed provisions on timing and rationale for deleting content.

	The Bill also creates the office of the eSafety Commissioner, to administer the cyber-bullying, cyber-abuse and non-consensual image schemes, as well as administering the online content scheme. The Bill starts to look at a systemic approach, in that it paves the way for the eSafety Commissioner to develop a set of Basic Online Safety Expectations (BOSE) for social media services, relevant electronic services and designated internet services. This set of expectations includes requirements for service providers to take all reasonable steps to: - Ensure the safety of end users - Minimise the five types of content on their service (set out below) - Minimise pornography served to child users (Category 2) - Establish an easy to use complaints mechanism for the five types of content in scope - Establish easy to use complaint mechanisms - Respond to requests from the eSafety Commissioner about how many take down notices they got, and how long they took to respond The BOSE are currently being developed with industry, following Australia's unique co-regulatory approach.	
CONTENT IN SCOPE	Five broad types of content are: 1. <u>Cyber bullying targeted at an Australian child (section 6)</u>	**The inclusion of cyberbullying of adults, and the associated definition, is a different approach to that taken in the OSB.** The

	Defined as when: a. An ordinary reasonable person would conclude that it: i. was intended to have an effect an Australian child; *and* ii. "is likely that the effect is seriously threatening, seriously intimidating, seriously harassing or seriously humiliating an Australian child" Where a take-down request is made for content that is considered cyber-bullying, a service has to remove this within **48** hours, and the end user who posted the material must refrain from posting more and must apologise to the child. <u>2. Cyber abuse - for 18+ (section 7 & 8)</u> Material is defined as cyber abuse if it is: a. Provided on a social media service, relevant electronic service or designated internet service *and* b. An ordinary reasonable person would conclude that: **i. it was intended to seriously harm an Australian adult *and*** ii. **the material is menacing, harassing or offensive.** Material is considered offensive if an ordinary reasonable person needs to consider them offensive, considering the 'standards' of the time, the literary/artistic/education merit of the material and the general character of the material (e.g. is it scientific etc) Where a take-down request is made for content that is considered	focus is on takedowns and the threshold for harm appears to be lower in the Australian Bill ("the material is menacing, harassing or offensive").

	cyber-bullying, a service has to remove this within **48** hours. <u>3. Non consensual intimate images (summary from sec 15 & 16)</u> Intimate images depict people's genitalia during a 'private act' (in a state of undress, using a toilet, showering, bathing, engaged in a sexual activity). It's also an intimate image if for religious reasons a picture of a person without particular attire would be distressing. These are considered non-consensual if the are: a. Provided on a social media service, relevant electronic service or designated internet service *and* b. The person in the image did not consent to the provision of the intimate image Where material is considered non consensual, and a removal notice is issued a service has **24** hours to take this down, and the end user who posted it may face a civil penalty <u>4. Class 1 materials (summary from sec 106 & 109)</u> Materials that are, or are likely to be, Refused Classification (RC) rated in Australia's Film and Game Classification system. Where material is considered Class 1,and a removal notice is issued a service has **24** hours to take this down. Link deletion notices can also be issued requiring a service provider to remove the link within **24** hours. App deletion notices can also be issued and an app distributor has **24** hours to prevent Australian users from being able to download the app.	

	5. Aboherrent violence (summary from section 9) Is defined in Australia's Criminal Code, and is material that depicts terrorist acts, murder or attempted murder, torture, rape or kidnap. If material is identified as abhorrent violence, a blocking notice may be issued. This calls for a block to the domain names, URLs and IP addresses that provide access to the materials, and lasts up to three months.	
SERVICES IN SCOPE	Three categories of digital services are in scope. 1. Social media services Defined as services that: - Sole purpose is to enable online social interaction between end users (advertising is not a 'sole purpose') *and* - Service enables end users to interact with each other *and* - Services allow end users to post materials 2. Relevant electronic service Defined as a service that allows end users to communicate with each other, including: - Email - Instant messaging - SMS services - MMS - Chat service - Multiplayer games	**Definition of social media service is more clearly focused at online networking platforms**. Less risk of catching other tech companies. However, broader provision to include email, SMS etc means more services likely to be included in Australian provisions than in the OSB.

	3. <u>Designated internet service</u> Internet services providers. Beyond this blocking requests can also be issued to App distributors (Google play, Apple App store etc).	
DEFINITION OF HARM	Individual. Includes consideration for if material "is menacing, harassing or offensive".	No single definition of harm.
POWERS OF REGULATOR	<u>Complaints and investigation system</u> The eSafety Commissioner has broad powers to investigate complaints where they relate to materials in scope. They administer the take down, blocking and removal systems, investigating complaints made, and can follow up on compliance. <u>Systems and audit</u> There is an expectation that services will consult with the eSafety Commission as they decide what is considered 'reasonable steps' to ensure end users safety, as per the BOSE requirements. The Commission can request and receive periodic audits from providers around the BOSE, and request non-period reports too. They can also request statements regarding compliance with the BOSE (which must be answered in 30 days) They can also issue requests for information about the number of	Less rigorous and formal investigatory powers than the OSB allows. Focus on takedown and complaint mechanisms. No clear algorithmic audit powers.

	complaints and notices from service providers. They can issue formal warnings and penalties for breaches.	
INDEPENDENCE OF REGULATOR	The eSafety Commissioner is independent of government.	No clear powers for political involvement.
TRANSPARENCY	Broadly, if they are requested to, companies will have an obligation to send the eSafety Commissioner ● Periodic reporting around BOSE ● Additional reporting around how a company is meeting the obligations in the BOSE. ● Information about the company's complaints and handling of takedown and removal requests under the Bill	Reporting is less regular and formal than OSB.
ADVERTISEMENTS	There are no specific provisions around advertising.	Advertising not included but not explicitly out of scope.
JOURNALISM AND NEWS	There is no carve out for news and journalism.	Focus is on social media so no required exemption.
USER IDENTITY	No, although eSafety Commissioner has been tasked with developing a 'road map' to age verification to enable the Basic Online Safety Expecation that services will minimise children's access to category 2 pornography. This is currently underway.	

GERMANY - NetzDG

	DETAILS	**COMMENTARY**

APPROACH (SYSTEMS VS TAKEDOWN)	**Content/Takedown** The Act is intended to counteract increasing hate crime and other criminal content. It entered into force on 01.10.2017. The Act was criticized throughout the legislative process with regards to its conformity in terms of constitutional and European Union law. The NetzDG is primarily aimed at providers of social networks and, due to the considerable threat of fines, constitutes administrative offense law and thus criminal law. Apart from reporting obligations (**Systems**), providers of social networks are required to · Take immediate note of complaints · Take down manifestly unlawful content within 24 hours of receiving the complaint. · Remove or block access to all unlawful content within 7 days of receiving the complaint In addition, providers of social networks must also immediately notify the person submitting the complaint and the user about any decision, while also providing them with reasons. In the newest amendment, a counterproposal procedure was introduced: providers of social networks must have an effective and transparent procedure by which both the complainant and the other user can obtain a review of a decision.	A takedown and redress approach. Aimed to improve processes and transparency around reporting and taking down illegal content.
CONTENT IN SCOPE	Only illegal content with regards to criminal law including: · Dissemination of propaganda material of unconstitutional organisations · Preparation of serious violent offence endangering	Heavy focus on illegal content. Notably includes intimate images which is not known to be a priority for the OSB but is also

	state · Instructions for committing serious violent offence endangering state · Disturbing public peace by threatening to commit offences · Revilement of religious faiths and religious and ideological communities · Dissemination, procurement and possession of child pornography · Violation of intimate privacy by taking photographs or other images	included in the Australian and Canadian Bills.
SERVICES IN SCOPE	The Act applies to social networks (Definition: "This Act shall apply to telemedia service providers which, for profit-making purposes, operate internet platforms which are designed to enable users to share any content with other users or to make such content available to the public (social networks)." The Act does not apply to · platforms offering journalistic or editorial content · platforms which are designed to enable individual communication or the dissemination of specific content. · Social networks that have fewer than two million registered users in the Federal Republic of Germany New amendment: Video Sharing Platforms included as well.	Limited to large social networks and video sharing platforms. Narrower scope than OSB.
DEFINITION OF HARM	All illegal content refers to the German Criminal Code.	As defined by criminal law.
POWERS OF REGULATOR	<u>Mediation</u> The Federal Office of Justice may recognize institutions organized under private law as conciliation bodies for settlements between	

	complainants or other users and social network providers. <u>Supervision</u> The Federal Office of Justice is monitoring compliance with the Act which also includes provisions of fines.	
INDEPENDENCE OF REGULATOR	Federal Office of Justice	Independent administrator.
TRANSPARENCY	A key obligation of social network providers is to report on the application of the NetzDG. The semi-annual report, which must be prepared in German, is intended to provide information about the effectiveness of the NetzDG. The report is published in the Federal Gazette and on the website of the respective network. The German government also used the transparency reports in its evaluation of the NetzDG. The 2021 amendment partially specifies the reporting obligation, but also expands it. Noteworthy: **the new duty to provide information regarding the type, basic features of the mode of operation and scope of any procedures used for the automated detection of content**. Although the automatic, complaint-independent checking and deletion activities of the providers are not subject to the procedural obligations of the NetzDG, they must now be reported. Only "basic information in a generally understandable form" is owed. Business secrets do not have to be disclosed. In addition, the provider must also deliver explanations of the general terms and conditions (e.g. community standards) in the transparency reports as well as a presentation on the	Transparency reports are published and evaluated by the regulator. Automated decision making has recently been included as a facet of the transparency reports.

	compatibility of these standards with the law on the use of general terms and conditions.	
ADVERTISEMENTS	No special regulation regarding advertisements. Due to different legislative competencies, advertising is regulated in the State Media Treaty (**MStV**). **MStV:** Advertisements must be clearly recognisable as such and clearly separated from other content. No subliminal techniques may be used in advertising. In the case of political, ideological or religious advertising, the advertiser must be clearly indicated in an appropriate manner.	Not included in scope although there are other provisions in German law to ensure adverts are transparent about the advertiser behind the content - particularly in political ads.
JOURNALISM AND NEWS	No special regulation regarding journalism and news. Due to different legislative competencies, journalism and news are regulated in the State Media Treaty (*Medienstaatsvertrag*). On the contrary, the NetzDG does explicitly not include platforms offering journalistic or editorial content. **MStV:** **Section 93 MStV - Transparency** Media intermediaries (Google, Facebook & Co.) must keep the following information easily perceptible, immediately accessible and permanently available in order to ensure diversity of opinion: ● The criteria which decide on the access of a content ● The criteria for aggregation, selection and presentation of content and their weighting, including information on the algorithms used **Section 94 MStV - Freedom from discrimination**	Like the OSB, NetzDG puts news media out of scope. **The MStV is one of the only other pieces of legislation in this comparison which makes special provision for news publishers and the treatment of their content on social media.** Social media firms must provide transparent information to publishers about how their content is surfaced; *and* they must not discriminate against journalistic content. **This is one of the closest provisions to Clauses 12-14 in the OSB.**

	In order to ensure diversity of opinion, media intermediaries must not discriminate against journalistically and editorially designed offers on whose perceptibility they have a particularly high influence.	
USER IDENTITY	No regulation. Age restrictions regarding harmful content are regulated in the State Media Treaty on Minors (*Jugendmedienstaatsvertrag - JMStV*). **Section 4 JMStV:** Certain offers are not allowed on the Internet (e.g. pornographic content.) if the provider does not ensure that the offers are only accessible to adults. Anonymity on the Internet is guaranteed in the Telemedia Act (*Telemediengesetz - TMG*). **Section 13 TMG:** The service provider must enable the use of telemedia anonymously or under a pseudonym, insofar as this is technically possible and reasonable.	Certain AV provisions elsewhere in German legislation. Anonymity is enshrined as a protection in the TMG - going much further than other pieces of legislation.

CANADA - proposed legislative framework to address harmful content online

	DETAILS	COMMENTARY
APPROACH (SYSTEMS VS TAKEDOWN)	**Takedown approach for online communication service providers** which is intended to capture major platforms and exclude products and services that would not qualify as online communication services, such as fitness applications or travel review websites. Regulated entities would have to take all reasonable measures to **make harmful content inaccessible within 24 hours of being flagged,** and do whatever is reasonable and within their power to monitor for the regulated categories of harmful content on their services, including through the use of automated systems based on algorithms. **<u>Content/ Takedown Approach</u>** The Act is a legislative and regulatory framework for social media, setting "new rules" that **oblige platforms to remove harmful content from their platforms within 24 hours** of being flagged while also providing procedural transparency to users and victims. The Act requires OCSPs (Online Communication Service Providers) to "take all reasonable measures" (including automated filtering) to identify and block the five categories of harmful content (see scope section). The Act sets two types of takedown requirements	Heavy takedown focus, restricted to illegal content. Systems and processes also included however through transparency reporting and better redress mechanisms.

	for: <u>Online Communication Service Provider (OCSP)</u> to remove five categories of harmful content in scope (see below) within 24 hours of being flagged <u>Internet Service Providers (ISPs) to:</u> • block access in Canada as a last resort with a court order, for platforms that persistently do not comply with orders to take down child sexual exploitation and terrorist content **<u>Systems Approach</u>** ● Transparency, reporting and preservation requirements for explicitly harmful (read: illegal) content, including child sexual exploitation, hate speech, and content that may threaten national security ● Procedural fairness for users, victims, and advocacy groups The Act also creates a new **Digital Safety Commission** composed of the: 1) Digital Safety Commissioner of Canada, 2) Digital Recourse Council of Canada, and 3) an Advisory Board.	
CONTENT IN SCOPE	Five categories of harmful content in scope (*as defined under the amended Canadian Human Rights Act and under the Criminal Code)* 1. Hate speech 2. Child sexual exploitation content 3. Non-consensual sharing of intimate images 4. Incitement to violence content	Illegal content only, akin to NetzDG.

	5. Terrorist content The Act provides exemptions for private communications and telecommunications, including messaging services (Whatsapp, Facebook messenger, etc.) and telecommunications companies (Rogers, Telus, Bell, etc.) The Act likewise provides exemptions for non-OSCPs (i.e. websites that provide services and products)	
SERVICES IN SCOPE	The Act applies to Online Communication Service Providers (OCSP) (**definition "a service that is accessible to persons in Canada, the primary purpose of which is to enable users of the service to communicate with other users of the service, over the internet. It should exclude services that enable persons to engage only in private communications.**" Services out of scope: The Act does <u>not</u> apply to: ● Private communications and telecommunications ● Products and services that are not OCSPs	
DEFINITION OF HARM	Individual and societal **(damage to societal cohesion**; vulnerable groups). Individual: to be aligned with the definition of hate speech outlined in Bill C-36: ● Hate speech is defined as "**content of a communication that expresses detestation or vilification of an individual or group of individuals on the basis of a prohibited ground of discrimination"** ● The content of a communication does not express	Broader definition of harm to include societal cohesion and vulnerable groups. Protected characteristics are included in the definition of hate speech. Both of these inclusions are not in the UK OSB. The Canadian Bill

	detestation or vilification, for the purposes of subsection (9), solely because it expresses mere dislike or disdain or it discredits, humiliates, hurts or offends.	probably goes the furthest in including societal harm.
POWERS OF REGULATOR	Legislation would **create a new Digital Safety Commission** of Canada to support three bodies that would operationalize, oversee, and enforce the new regime: the **Digital Safety Commissioner of Canada** (to administer, oversee, and enforce the new legislated requirements), the Digital Recourse Council of Canada (provide independent and binding decisions on whether or not content qualifies as harmful content as defined in legislation and should be made inaccessible), and an Advisory Board (provide both the Commissioner and the Recourse Council with expert advice to inform their processes and decision-making). **<u>Digital Safety Commissioner</u>** *Information gathering powers; information sharing; inspection powers; research powers; outreach responsibilities* Oversees and improves online content moderation by: ● Administering and enforcing obligations; ● Engaging with and considering the particular needs of and barriers faced by groups disproportionately affected by harmful online content such as women and girls, Indigenous Peoples, members of racialized communities and religious minorities and of LGBTQ2 and gender-diverse communities and persons with disabilities ● Supporting platforms in reducing harmful content affecting peoples in Canada. **<u>Digital Recourse Council</u>**	Like Ireland, creates a new regulator and Digital Safety Commissioner with responsibility for overseeing the agenda.

	Decision-making powers; inspection powers ● Provides independent recourse through a digital tribunal system ● Makes binding decisions on content removal **<u>Advisory Board</u>** *Research powers; consultative powers; recommendation powers* ● Provides expert advice and guidance to the Commissioner and the Recourse Council ● Brings expert, equity-deserving, and Indigenous interests to social media regulation	
INDEPENDENCE OF REGULATOR	Digital Safety Commissioner (Independent) ● Oversees and enforces the Act ● Sets norms ● Builds a basis of research	Arms-length regulation via Commissioner.
TRANSPARENCY	Baseline transparency requirements would require providers to disclose Canada-specific data on the volume and type of content dealt with at each step of the content moderation process, as well as information on how regulated entities develop, implement, and update their guidelines for the kinds of content they prohibit. Regulated entities would also be required to publish transparency reports on the Canada-specific use and impact of their automated systems to moderate, take down, and block access in Canada to harmful content. **<u>For Online Communication Service Providers:</u>**	Data transparency requirements appear clearer than those in the UK OSB.

	The Act sets out reporting requirements (module 1B.14) on a scheduled basis for OCSPs to the Digital Safety Commissioner on Canada-specific data about: ● the volume and type of content moderated; of harmful content on their OCS; content that was accessible to persons in Canada in violation of their community guidelines; ● resources and personnel allocated to their content moderation activities; ● their content moderation procedures, practices, rules, systems and activities (including automated decisions) and how they monetize harmful content **For Digital Safety Commission:** The Act requires the Digital Recourse Council and Digital Safety Commissioner provide reports on their activities for the fiscal year to the Minister of Canadian Heritage. **For Platforms:** Platforms must report to law enforcement and CSIS of certain forms of harmful content, including that which suggests an imminent risk of serious harm to any person or property Platforms must also report "prescribed" content of criminal concern to law enforcement and/or CSIS (depending on type of content) To comply with the Act, platforms must provide law enforcement with the content *and* any additional public-facing information as set out in the Governor-in-Council regulations	

	<u>Mandatory Reporting Act</u> The Act requires Internet Service Providers to report certain information in their mandatory reporting when a child pornography offence has taken place.	
ADVERTISEMENTS	**There are <u>no specific provisions</u> around advertising,** although the Act sets out that OCSPs must generate and provide reports on a scheduled basis to the Digital Safety Commissioner on Canada-specific data about their monetization of harmful content. Advertising provisions are covered by other mechanisms within Canada, including the Competition Act. Two years ago (Bill C-76) the government introduced a mandatory election advertising archive for platforms during the election campaign, as well as limits on third party online spending.	Ads not in specifically but perhaps included via the transparency over harms monetization. Ad libraries included in other regs.
JOURNALISM AND NEWS	There are no specific provisions around news and journalism in the online harms legislation. However, Canada's recent Journalism Labour Tax Credit law includes income tax measures to support journalism organizations producing original news content, and the government has opened charitable status to journalistic organizations.	Alternative provisions for journalism outside of this framework.
USER IDENTITY	There are no specific provisions about age verification or self-anonymity.	Like most other regs, user ID not included.

Reset.

US - Algorithmic Justice Act (AJA) & Social Media DATA Act (introduced by Rep Trahan & Rep Castor)

	DETAILS	**COMMENTARY**
APPROACH (SYSTEMS VS TAKEDOWN)	AJA - The bill takes a systematic approach to establish a safety and effectiveness standard for algorithms, such that online platforms may not employ automated processes that harm users or fail to take reasonable steps to ensure algorithms achieve their intended purposes.	An intervention to protect citizens/users from algorithmic bias. Algorithmic harm rather than content harm. Similar to considerations made by the CMA.
CONTENT IN SCOPE	N/A	This is not a content moderation bill.
SERVICES IN SCOPE	AJA - This bill covers online platforms, which include any public-facing website, online service, online application, or mobile application which is operated for commercial purposes and provides a community forum for user generated content, including a social network site, content aggregation service, or service for sharing videos, images, games, audio files, or other content.	Applies more broadly than just social media/UGC sites.
DEFINITION OF HARM	Harm is defined as algorithmic discrimination on the basis of race, age, gender, ability and other protected characteristics. Enforcement actions can be brought by Federal Trade Commission, Department of Justice, states or individuals.	Harm is based on protected characteristics in the AJA.
POWERS OF REGULATOR	AJA - FTC would review detailed **records of platform algorithmic processes**, in compliance with key privacy and data de-identification standards. Also, an inter-agency task force comprised of the FTC, Department of Education, Department of	Joint regulatory task force empowered to investigate algorithmic processes.

	Housing and Urban Development, Department of Commerce, and Department of Justice, would be able to investigate the discriminatory algorithmic processes employed in sectors across the economy. Social Media DATA Act establishes a working group within the Federal Trade Commission that would develop a set of best practices around social media research.	FTC also housing a group to establish best practice for social media research and (below) will improve access to data for researchers.
INDEPENDENCE OF REGULATOR	An inter-governmental task force.	
TRANSPARENCY	AJA - The Bill would **increase transparency into websites' content amplification and moderation practices.** Online platforms would be required to: ● describe to users in plain language the types of algorithmic processes they employ and the information they collect to power them ● maintain detailed records describing their algorithmic process ● publish annual public reports detailing their content moderation practices. ● adopt notice requirements for algorithmic processes; ● adopt a five-year data retention obligation of algorithmic processes; ● draft rules for de-identification of personal information Social Media DATA Act - Large ad platforms with more than 100 million monthly active users would be **required to give**	Greater transparency requirements and a big push to improve access to data for academic researchers.

	researchers affiliated with academic institutions access to these databases, which would include ads from any advertiser spending more than $500 a year on the platform.	
ADVERTISEMENTS	Creates transparency requirements for advertising practices. Social Media DATA Act would force large social media platforms to give researchers and the Federal Trade Commission access to more detailed ad libraries. Those libraries would include, among other things, a description of the audience that was targeted, information about how many people interacted with the ad and details about whether the ad was optimized for awareness, traffic or some other purpose.	Ad libraries need to be boosted and updated with more granular information about relevant ads.
JOURNALISM AND NEWS	N/A	
USER IDENTITY	N/A	

8 October 2021