Digital Policy Lab

Policy Digest #4

18 May 2022

ISD | Powering solutions to extremism, hate and disinformation

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

Section 1 Digital policy developments

Global: Declaration for the Future of the Internet

Type Non-regulatory (non-binding declaration)
Status Signed

On 28 April 2022, a global coalition of more than 60 countries including the EU Member States and the US, endorsed a Declaration for the Future of the Internet. The list of signatories is available here. Its signatories "support a future for the Internet that is an open, free, global, interoperable, reliable, and secure", committing "to protecting and respecting human rights online and across the digital ecosystem". The Declaration notes, "Online platforms have enabled an increase in the spread of illegal or harmful content that can threaten the safety of individuals and contribute to radicalization and violence". Signatories specifically made the commitment that "actions taken by governments, authorities, and digital services including online platforms to reduce illegal and harmful content and activities online be consistent with international human rights law, including the right to freedom of expression while encouraging diversity of opinion, and pluralism without fear of censorship, harassment, or intimidation".

The Declaration expects to contribute to existing processes in the UN system, G7, G20, the Organisation for Economic Cooperation and Development (OECD), the World Trade Organisation (WTO), the Internet Corporation for Assigned Names and Numbers (ICANN), Internet Governance Forum (ICF), and Freedom Online Coalition (FOC).

Canada: Bill C-18 (The *Online News Act*)

Type Regulatory (legislative and regulatory framework)
Status Introduced

On 5 April 2022, the Minister of Canadian Heritage introduced Bill C-18, the *Online News Act,* which would establish a new legislative and regulatory framework to "mitigate bargaining imbalances between digital platforms and news outlets over the way digital platforms allow news content to be accessed and shared on their platforms". Bill C-18 requires "digital platforms" with a "significant bargaining imbalance with news media" to make fair commercial deals with outlets for the news and information shared on their platforms. Such deals would need to provide fair compensation, respect journalistic independence and invest in a diversity of Canadian news outlets, including independent local businesses. The legislation would target all news businesses that meet the eligibility criteria, regardless of media type, including newspapers and news magazines with a digital presence, online news outlets, and private and public broadcasters that produce and publish original online news content, as well as both television and radio broadcasters. The Bill allows media outlets to assemble and bargain collectively for the purposes of negotiation with digital platforms.

The Bill "builds on similar legislation in Australia, by adding greater public accountability and transparency to the process". A backgrounder by the Government of Canada can be found here. Once the Bill receives Royal Assent, the Governor in Council would publish regulations specifying the application of the Act and the criteria for exemptions.

---

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

EU: Digital Services Act (DSA)

Type Regulatory (legislative acts)
Status Provisional agreement adopted

On 23 April 2022, the trilogue negotiations between the Council of the EU, the European Parliament and the European Commission reached a provisional agreement on the DSA. Some of the contours of the agreement have been outlined in press releases by the Parliament, Council and Commission, indicating where compromises were reached on wedge issues. Among the compromises, the DSA will ban profiling for advertising purposes based on sensitive data (e.g., political and sexual orientation) and the data of minors for advertising purposes. It will ban "dark patterns" that deceive or push users into consenting to certain uses of their data, or purchasing products. Very large online platforms (VLOPs) and very large search engines (VLOSEs) with over 45 million users in the EU will be obliged to take risk-based action to prevent the misuse of their systems and undergo independent audits of their risk management systems. VLOPs and VLOSEs must also facilitate the development of voluntary industry standards and codes of conduct. A new article introducing a crisis response mechanism has been added, which "will be activated by the Commission on the recommendation of the board of national Digital Services Coordinators". The mechanism will "decide on proportionate and effective measures to be put in place for the respect of fundamental rights".

The agreement is now subject to formal approval by the two co-legislators, the Parliament and the Council. The text is expected to be finalised before a vote in the Internal Market and Consumer Protection (IMCO) of the Parliament in June, with the plenary vote planned for July, coinciding with the adoption of the Digital Markets Act (DMA). Once adopted, the DSA will be applied in its entirety across the EU and will apply fifteen months or from 1 January 2024, whichever is later, after entry into force. The rules for VLOPs and VLOSEs will apply four months from their designation as such by the Commission. Between 20 and 30 platforms are expected to fall into the VLOP or VLOSE category, and most of them will also be considered gatekeepers under the DMA.

EU: European Data Protection Board's Guidelines on dark patterns

Type Non-regulatory (non-binding guidelines)
Status Consultation closed

On 2 May 2022, the European Data Protection Board closed its consultation on the recently published "Guidelines 3/2022 on dark patterns in social media platform interfaces: how to recognise and avoid them". The guidance offers practical recommendations on how to assess and avoid "dark patterns" that infringe the principles of the General Data Protection Regulation (GDPR). The guidelines are aimed at social media providers, recalling the "obligations coming from the GDPR, with special reference to the principles of lawfulness, fairness, transparency, purpose limitation and data minimisation in the design of user-interfaces and content presentation of their web services and apps".

In the context of the Guidelines, "dark patterns" are considered as "interfaces and user experiences implemented on social media platforms that lead users into making unintended, unwilling and potentially harmful decisions regarding the processing of their personal data". The guidelines provide a non-exhaustive list of concrete examples of dark pattern types for different use cases within the life cycle of a user account.

UK: Online Safety Bill (OSB)

Type Regulatory (legislative and regulatory framework)
Legislative status Introduced in Parliament

The Second Reading debate of the revised OSB on 19 April 2022 demonstrated cross-party consensus on the need for legislation, and despite the condensed time allocated for the session, the Bill passed this stage of the legislative process without a vote. Among the issues raised by Members of Parliament were the lack of clear definitions and the amount of detail that will be left to secondary legislation, risks that the categorisation of companies will allow harms on smaller platforms, the extent that the Secretary of State's powers will negatively affect Ofcom's independence, as well as potential gaps in the Bill, including disinformation, action on "cross-platform harms" and risks from emerging technologies. The Bill moves through to the Public Bill Committee stage for further line-by-line scrutiny and consideration of potential government or opposition amendments. The Committee is accepting written evidence submissions as well as oral evidence, and is scheduled to report by the end of June. This will be followed by discussion of any tabled amendments in both Houses. A factsheet provided by the Department for Digital, Culture, Media & Sport (DCMS) can be found here.

Section 2 Topic-specific snapshot: "Assessing and mitigating risks of algorithmic recommender systems"

Social media platforms use algorithmic systems to curate and personalise online content. As seen in documents leaked by whistleblower and former data engineer at Facebook, Frances Haugen, much of the damage done by harmful content such as hate speech and disinformation relates to its widespread amplification by the platform's algorithms. The leaks showed that algorithms boosting "meaningful social interactions", i.e. engagement, promote misinformation, toxicity, and violent content, while lacking transparency and accountability. Numerous proposed digital policy interventions now incorporate platforms' algorithmic recommender systems and design choices within their scope. A "systemic" approach to platform regulation demands transparency, accountability, and oversight over platforms and their systems. Such proposals include privacy protections and user controls as well as the auditing of policies, processes and outcomes of algorithmic systems to understand the "systemic risks" they pose to society. This section presents summaries of selected initiatives within the remit of regulating algorithms, including ideas from academia and civil society.

EU: Risk-based assessments of recommendation algorithms

The **Digital Services Act** (DSA, see Section 1) recognises the potential harms resulting from the use of algorithmic systems, in particular as regards to the societal risks they pose when disseminating illegal and harmful content, including disinformation.

The current agreement of the DSA obliges very large online platforms (VLOPs) to carry out annual assessments of the risks posed by the "design of their recommender systems" and "other relevant algorithmic systems" to fundamental rights, human dignity, data protection, diversity of expression and media, non-discrimination, protection of minors and consumer protection. Any negative effects of platform services on civic discourse, electoral processes, and public security, on gender-based violence, the protection of public health, minors and serious negative consequences to the person's physical, mental and social well-being, must be assessed by the platforms (Art. 26 & 27). During inspections, the European Commission, auditors and experts appointed by it may require VLOPs to provide access to and explanations of its algorithms and databases (Art. 53 & 57).

Additionally, online platforms will be required to set out in their terms and conditions, "in plain and intelligible language", the "main parameters used in their recommender systems" as well as "any option for the user to modify or influence" those parameters. The main parameters must include the "criteria most significant in determining the information" as well as the "reasons for the relative importance for those parameters". Where several options for modifying the parameters are available, online platforms should make a functionality directly and easily accessible that allows the user to "select and modify any time their preferred option" (Art. 24).

The Greens/EFA Group in the European Parliament previously proposed the **interoperability of recommender systems** to address the concentration of power with a few large social media platforms. The amendments, ultimately rejected, would have enabled users to choose third-party systems for the curation of content in their timelines (so called "middleware").

🇩🇪 Germany: The Interstate Media Treaty

In November 2020, the **Interstate Media Treaty** (Medienstaatsvertrag or MStV) entered into force, introducing rules for not only television, radio and certain telemedia, but also search engines, smart TVs, app stores, social media and video sharing platforms. The treaty defines "media intermediaries" as "any telemedia that aggregates, selects, and presents third-party journalistic/editorial offers, without presenting them as a complete offer."

With regard to algorithmic transparency and to "ensure diversity of opinion", providers of media intermediaries must present "in easily understandable, directly accessible, and continuously available" manner the "central criteria of an aggregation, selection, and presentation of content and the weighting thereof, including information about the functionality of the implemented algorithms". The treaty signals regulators' willingness to take on online platforms and pushes the policy debate beyond questions of content moderation and deletion.

🇺🇸 US: Towards algorithmic accountability

In the United States, several legislative initiatives are on the table to make platform's use of algorithmic systems more accountable and transparent.

On 3 February 2022, Democrats in the Senate and the House introduced the **Algorithmic Accountability Act of 2022** (H.R. 6580 and S.3572), an update of the Algorithmic Accountability Act from 2019. The Bill would direct the Federal Trade Commission (FTC) to require impact assessments that refer to an "ongoing study and evaluation of an automated decision system or augmented critical decision process and its impact on consumers". The Bill defines an "automated decision system" as "any system, software, or process (including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques and excluding passive computing infrastructure) that uses computation, the result of which serves as a basis for a decision or judgment". The Act would apply to entities with more than $50 million USD in average annual gross receipts (turnover), or entities worth more than $250 million USD in equity value that develop or deploy any automated decision system on more than 1 million consumers, households, or consumer devices.

Specifically, the Bill:

- requires the FTC to implement regulations that require covered entities to perform impact assessments and meet other requirements regarding automated decision-making processes, to publish an annual anonymised aggregate report on trends, and to establish a repository of information where consumers and advocates can review which critical decisions have been automated by companies;

- requires entities to maintain documentation of any impact assessments performed, to submit annual reports to the FTC covering any ongoing impact assessments; and to eliminate or mitigate, in a timely manner, any impact made by an augmented critical decision process that demonstrates a likely material negative impact;

- would establish a Bureau of Technology, headed by a Chief Technologist, and to appoint 50 personnel as well as 25 additional personnel to the Division of Enforcement of the Bureau of Consumer Protection.

The **FTC** previously appointed senior advisors on Artificial Intelligence joining from the AI Now Institute at New York University, a group that developed the use of algorithmic impact assessments as a framework for evaluating the effects of algorithmic systems.

Another proposal, the **Algorithmic Justice and Online Platform Transparency Act of 2021** (H.R. 3611), would prohibit the discriminatory use of personal information by online platforms in any algorithmic process. The Bill requires online platforms to describe to users in plain language the categories of personal information it collects or creates for purposes of the algorithmic process, as well as to maintain detailed records, including methods by which the type of algorithmic process weighs or ranks certain categories of personal information, for review by the FTC. The Bill further requires online platforms to publish annual public reports detailing their content moderation practices. Enforcement powers would be shared between the FTC and state attorneys general and individuals would be given a private right of action.

Other Bills introduced in Congress seeking to regulate algorithmic systems focus on reforming Section 230 to specifically address social media harms, rather than addressing algorithmic technologies used by companies more generally. For example, the **Protecting Americans from Dangerous Algorithms Act** (H.R. 2154) would limit a social media company's immunity from liability if it promotes certain types of content on its platform. Specifically, the Bill removes this immunity from a social media company with more than 10 million monthly users if it utilises "an algorithm, model, or other computational process to rank, order, promote, recommend, amplify, or similarly alter the delivery or display of information" that is directly relevant to a claim involving (1) interference with civil rights, (2) neglect to prevent interference with civil rights, or (3) acts of international terrorism.

Further reading: "Social Media Futures: Interventions Against Online Unpleasantness", Oliver Marsh, Tony Blair Institute for Global Change, 19 April 2022.

In this article, Oliver Marsh argues that platforms should take an interventionist approach to help encourage user reflection about online "unpleasantness". Marsh considers "unpleasantness" as behaviours, which are not "harmful" in the manner of hate speech, but can nonetheless still negatively impact upon online environments. Marsh argues that users should be given greater controls over their own experience, while also prompting them to consider how they affect others' experiences. The article considers following alternatives to restricting content:

- · "Positivity prompts" which nudge users away from offensive language could appear on posts, which breach a certain 'toxicity score', algorithmically assigned by software such as Google's Perspective API. Such steps can encourage users reconsidering their language, as well as reduce the toxicity of users' future comments, and of other comments in the conversation.
- · "Explanatory prompts" could summarise reasons why a particular behaviour could cause upset, including less frictional design choices, e.g., a section of a comment box could change colour and pattern as a typed comment became more unpleasant.
- · "Positive modes" for users to indicate that they wish to skew the balance towards more zealous filtering of negative content. This technology could be developed by platforms themselves. Alternatively, it could be developed by external companies as "middleware" — "software and services that would add an editorial layer between the dominant internet platforms and internet users". Behavioural signals from users with 'positive mode' switched on could provide feedback for platforms to further refine negativity filters.

Marsh asserts, "Decisions will need to be made around what behaviour should receive prompts, be filtered in positive mode, and similar questions". Platforms should continually — and in a transparent and collaborative manner — decide whether certain topics are "of social importance" and allow them some exemption to filtering. Civil society and other organisations, Marsh argues, should support developing visions of what "good" discourse might look like.

Further reading: "Amplification and Its Discontents - Why regulating the reach of online content is hard", Daphne Keller, Knight First Amendment Institute, Colombia University, 8 June 2021.

In this essay, Daphne Keller lays out why "regulating amplification" to restrict distribution of harmful or illegal content is difficult. Keller notes that laws assigning liability to platforms for hosting content and laws assigning liability for amplifying it share practical issues, i.e., someone has to decide which content is excluded, whether it is removed from the whole platform or just from features like recommendations. Keller debates three common regulatory models:

1. Illegal speech models that increase platform liability for amplifying illegal speech;
2. Harmful speech models that increase platform liability for amplifying currently lawful but harmful speech;
3. Content-neutral models that increase platform liability for amplifying any speech.

In her debate of the third model, Keller considers legal questions as to whether a particular rule actually achieves content neutrality. In this section, Keller also discusses two promising content-neutral rules: "circuit breakers" to slow the spread of highly viral content, and laws to give users more control over algorithmically ranked content using competition, privacy, or possibly consumer protection rights. Keller argues that better transparency about the parameters of major platforms' algorithms would benefit consumers, lawmakers, and society at large – on its