Digital Policy Lab

Policy Digest #8

10 November 2022

ISD Powering solutions to extremism, hate and disinformation

Sponsored by Federal Republic of Germany Foreign Office

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

Section 1 Digital policy developments

EU: Digital Services Act (DSA)

Type Regulatory
Status Signed

On 19 October 2022, the DSA was signed by the European Parliament and the Council of the EU, and was published in the Official Journal of the EU on 27 October, entering into force 20 days after (16 November). The regulation will apply for Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) four months after they are designated as such (the first designations are expected for mid-February 2023). Following the takeover of Twitter by Elon Musk, Thierry Breton, Commissioner for the Internal Market, made it clear that the rules will also apply to Twitter. VLOPs and VLOSEs will have to offer users a system for recommending content that is not based on profiling as well as assess and mitigate systemic risks they create – such risks relate to the dissemination of illegal content, negative effects on fundamental rights, on electoral processes and on gender-based violence or mental health. The Commission will have direct supervision and enforcement powers and can, in the most serious cases, impose fines of up to 6% of the global turnover of a service provider. For rogue platforms refusing to comply with important obligations, it will be possible as a last resort to ask a court for a temporary suspension of their service, after involving all relevant parties.

EU: European Media Freedom Act

Type Regulatory
Status Proposal

On 16 September 2022, the European Commission adopted a proposed Regulation establishing a common framework for media services in the internal market, the European Media Freedom Act, which includes safeguards against political interference in editorial decisions and against surveillance. It will replace the European Regulators Group for Audiovisual Media Services (ERGA) with a new body, the European Board for Media Services, which will comprise representatives from national media authorities and be tasked with advising the Commission on the effective and consistent application of the EU media law framework. The regulation was announced by Commission President von der Leyen in her 2021 State of the Union Address, and builds on the Commission's rule of law reports and the revised Audiovisual Media Services Directive, which provides for EU-wide coordination of national legislation for audiovisual media.

The Commission opened a feedback period until 28 December 2022 (although the eight-week feedback period is being extended every day until this adopted proposal is available in all EU languages). All feedback received will be presented to the Parliament and Council with the aim of feeding into the legislative debate. The Committee on Culture and Education (CULT) has been pre-designated as the committee responsible, with the Committees on Civil Liberties, Justice and Home Affairs (LIBE) and Internal Market and Consumer Protection (IMCO) asked to give an opinion. The Council is moving forward with examining the text at the Audiovisual Working Party but no General Approach is expected under the Czech Council Presidency.

[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: Regulation on the transparency and targeting of political advertising

Type Regulatory
Status Under discussions in the Parliament and Council

In November 2021, to address the challenges posed by online electoral campaigns, the European Commission presented a proposal for a Regulation that aims to build a harmonised set of rules on transparency and targeting of political advertising, and would apply to both online and offline political advertisements. The proposal defines political advertising as the placement, promotion or dissemination of any message that features specific political messages, regardless of whether or not the publisher or disseminator of the message disseminates it on the basis of providing a "service" to a sponsor. On 19 October 2022, a compromise text by the Council's Czech Presidency reinforced this broad definition of advertising. On 28 October 2022, more than 30 civil society organisations signed a public letter addressed to the Czech Presidency criticising, among other things, that the compromise text "dangerously mischaracterises the mere expression of political ideas and civic engagement as political advertising". The letter urges the Czech Presidency to distinguish civic voices from political advertising, defining the latter as *always* involving a service to a sponsor.

Germany: Network Enforcement Act (*Netzwerkdurchsetzungsgesetz*, or NetzDG)

Type Regulatory
Status Issued (Penalty notices)

On 10 October 2022, the German Federal Office of Justice (BfJ) issued two fines against Telegram. The messaging app is accused of violations of the obligations in the NetzDG to provide legally compliant reporting channels and to name an authorised person or institution with an address in Germany, so that German courts and authorities can serve the providers with legally binding documents. The BfJ imposed a fine of 4.25 million EUR for the violation of the obligation to provide legally compliant notification channels, and another fine of 875.000 EUR for failure to designate a domestic agent. Since April 2021, the BfJ made several attempts to serve letters at Telegram's headquarters in Dubai. Despite assistance from the competent authorities, this effort did not succeed. After placing both letters in the Federal Gazette, Telegram commented, but did not refute the allegations. The penalty notices are not yet legally binding. Telegram can file an appeal with the BfJ. If the BfJ does not grant an appeal, it will send the respective files via the public prosecutor's office to the competent district court in Bonn for a court decision.

UK: Online Safety Bill (OSB)

Type Regulatory
Status Report stage

On 26 October 2022, Westminster Hall held a debate on Online Harms, brought by Damian Hinds MP. During the debate, Damian Collins MP, who then held the Tech and Digital Economy post at the Department of Digital, Culture, Media and Sport (DCMS), commented on the timelines of the OSB, "I want to see it complete its Commons stages and go to the House of Lords as quickly as possible." The Bill was due to have its third reading in the Commons on 1 November, but has since been removed from the Commons timetable. While Collins has since left the position, to be replaced by Paul Scully MP, the DCMS Secretary of State Michelle Donelan, appointed by then Prime Minister Liz Truss in September, has remained in her post in new Prime Minister Rishi Sunak's reshuffle. The change of Prime Minister may not lead to major changes in policy approach, as Sunak has expressed similar previous concerns about clauses pertaining to the adult safety duties around 'legal but harmful' content in previous versions of the Bill. In August 2022, Sunak's spokesperson noted during his initial leadership campaign, "Rishi believes the Government has a duty to protect children and crack down on illegal behaviour, but should not infringe on legal and free speech."

UK: Ofcom's first report on video-sharing platforms (VSPs)

Type Regulatory
Status Published (Implementation report)

On 20 October 2022, Ofcom published the findings from the first year of its implementation of the VSP regime, which requires providers to take appropriate measures to protect the general public from "relevant harmful material" as well as under-18 year olds from "restricted material". Ofcom reported on TikTok, Snapchat, Twitch, Vimeo, BitChute, OnlyFans, and smaller (adult) VSPs. The report notes that platforms generally provided limited evidence on how well their safety measures are operating to protect users. It notes concerns that smaller adult sites do not have robust measures in place to prevent children accessing pornography. They all have age verification measures in place when users sign up to post content. However, users can generally access adult content just by self-declaring that they are over 18 years of age. Over the next year, adult sites that Ofcom already regulates must have in place a clear roadmap to implementing robust age verification measures. Ofcom also found that platforms are not prioritising risk assessment processes, which it believes are fundamental to proactively identifying and mitigating risks to user safety. In terms of scope, Ofcom's VSP guidance notes that providers are not required to take all proposed measures, but should "determine whether it is appropriate to take a particular measure, according to whether it is practicable and proportionate to do so, considering factors including the size and nature of the platform; the type of material on the platform and the harm it might cause; and the rights and legitimate interests of users". Hence, it is up to VSP providers to self-assess whether and to what extent the VSP framework and its statutory requirements apply to them, affording VSP providers flexibility in how they protect users.

US: Gonzalez v. Google (Supreme Court case on Section 230)

Type Judicial
Status Proceedings ongoing

On 3 October 2022, the US Supreme Court (SCOTUS) granted a petition for a writ of certiorari in the case of *Gonzalez v. Google*, agreeing to hear a case in relation to the scope of Section 230(1)(c) of the Telecommunications Decency Act. Under Section 230, if a user posts defamation, harassment, or other forms of harmful speech, the individual user can be sued, but the platform (with a few exceptions) cannot. Essentially, Section 230 provides immunity to platforms for the publication of content provided by users.

The question put to SCOTUS is whether this immunity also applies to targeted algorithmic recommendations of content provided by another content provider, or only limits the liability when services engage in traditional editorial functions (such as deciding whether to display or withdraw) with regard to such information. The case was initiated by relatives of Nohemi Gonzalez, a US citizen killed by ISIS terrorists in the November 2015 attacks in Paris. The plaintiffs filed a claim in a California federal district court against Google under the Anti-Terrorism Act, alleging that "by recommend[ing] ISIS videos to users, Google assists ISIS in spreading its message and thus provides material support to ISIS". The district court dismissed the complaint, finding that the claims "fall within the scope of [Section 230's] immunity provision". In 2021, the US Court of Appeals for the 9th Circuit ruled that Section 230 protects such recommendations. Now *Gonzalez v. Google* presents an opportunity for the SCOTUS to weigh in on the scope of Section 230.

At the same time, two other cases, *NetChoice v. Paxton* and *Moody v. NetChoice*, are challenging laws in Texas and Florida that restrict platforms' authority to remove user-generated content. The US Court of Appeals for the 5th Circuit upheld the law in Texas, while the Court of Appeals for the 11th Circuit struck down Florida's similar law, so SCOTUS will very likely weigh in to resolve a circuit split. Former President Donald Trump and 16 Republican-led states already filed amicus curiae briefs (see here and here), urging the justices to find in favour of Florida Attorney General Ashley Moody (R) and supporting the law barring platforms from banning political candidates, among other provisions. The Biden Justice Department has previously defended the constitutionality of Section 230, while Republican states are seeking a new status quo.

US: Blueprint for an AI Bill of Rights

Type Non-regulatory (Principles)
Status Published

In October 2022, the White House Office of Science and Technology Policy published a non-binding "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People", which is grounded in a vision laid out in 2021 by White House policy advisor Dr. Alondra Nelson. The Bill considers greater transparency on how algorithms are created, more accountability for AI-based decision-making, the ability for citizens to complain, data privacy protections, as well as fallback and escalation processes if an automated system fails. The Bill comes amid parallel developments at the EU level. During the next meeting of the EU-U.S. Trade and Technology Council to be held in early December, officials expect to publish a "joint roadmap on AI evaluation and measurement tools for [trustworthy] AI and risk management," according to a leaked document obtained by POLITICO. In the US, long-standing legislation such as Illinois' Biometric Information Privacy Act already give citizens the ability to sue if they believe their data has been misused. Contrary, in the EU, the upcoming AI Act is based on a top-down approach that puts almost all onus on policymakers and regulators to combat the risks associated with AI.

Section 2 Topic-specific snapshot: "Data Access Challenges of the Evolving Online Ecosystem"

This section presents summaries of selected analyses and commentary published by civil society and academia on emerging challenges related to external researcher access to platform data.

Elizabeth Hansen Shapiro, Michael Sugarman, Fernando Bermejo & Ethan Zuckerman. New Approaches to Platform Data Research. *NetGain Partnership*. February 2021.

The report explores the challenges and potential of a wide range of approaches to studying social media platforms, ranging from cooperative to adversarial strategies. The report considers perspectives of academic researchers, journalists and activists, including an overview of the different ways researchers are trying to understand social media data, the obstacles to accessing that data, and a set of recommendations for policymakers and philanthropic funders to increase data access. The report argues that researchers need better research tools — including panel studies and data donation approaches — that can study social media from a *user* point of view as well as from a platform point of view. The report notes that data privacy scandals such as Cambridge Analytica created tensions where platforms invoke user privacy as a reason to restrict access to data. As a result, platforms would need better incentives to enable data access, while privacy advocates would need to take seriously researchers' needs to access data, especially to ensure that platforms are actually implementing the privacy practices advocates are seeking.

The report recommends:

- Legislative action to create a "safe harbor" for researchers to access data, protecting research from some types of prosecution under "anti-hacking" laws;
- Robust dialogue between privacy activists and researchers about legally safe and ethical approaches to accessing platform data;
- Common ethical standards for social media research, especially around data collection and analysis;
- Support for new and ongoing experiments in social media data donation and panel studies, two promising approaches to understanding what users are exposed to across platforms;
- Support for unauthorised indices of platform data, including ongoing work to index the content of platforms with a history of permitting extremist and hate speech;
- Regular audits of platforms conducted either by outside auditing bodies or through a movement towards internal best practice audits;
- Regulatory action that treats large platform companies as common carriers, subjecting companies to stricter oversight and auditing;
- Support for the creation of new types of platforms designed from creation for study and monitoring by outside researchers.

The authors conclude that there will be no single simple solution to the complex problems of allowing researchers increased access to platform data. The authors note that foundations and other funders that support advocacy around technology may bear some responsibility for the tensions between privacy and research rights. In sum, the report calls for a range of approaches, including those that involve platform cooperation and those that assert a right to research without the platforms' explicit permission.

Jakob Guhl, Oliver Marsh & Henry Tuck. Researching the Evolving Online Ecosystem. *Institute for Strategic Dialogue*. July 2022.

In their report, the authors highlight barriers posed by online platforms to researching and mitigating harmful content and behaviours, and review existing research methodologies and tools to address these barriers. The report presents possible future scenarios for the evolving online ecosystem, and proposes a series of recommendations for policymakers, platforms and the research community.

The authors outline barriers to finding content and identifying harmful behaviours online, notably technological features which block or limit access to data, ethical and legal issues faced by researchers, as well as fragmentation of content across platform(s) in a way which impedes efficient and systematic data. For example, limited access may be a side effect of features (e.g., end-to-end encryption), while some forms of content or data are not (yet) as amenable to systematic search and storage, primarily data on audio-visual platforms. The authors emphasise that features aimed at protected, private and secure communication have major upsides from a human rights and privacy rights perspective, so combatting harmful activity on platforms should not come at the price of sacrificing data privacy and protection (such as encryption). Furthermore, accessing data, and particularly the collection and processing of data, can raise ethical issues, such as invasions of privacy or the use of data or content without users' consent. This may lead to contraventions of ethical research practices, platform terms and conditions, or even the law. Lastly, much online content is theoretically accessible without barriers caused by technological structures or ethical and legal issues; however, one still does need to know where to look. Often relevant content is among vast amounts of material that cannot be searched quickly and systematically, for example, via a platform-wide search function or API. In brief, in a fragmented environment, theoretically accessible content cannot be searched quickly or systematically.

Chris Riley & Susan Ness. Modularity for International Internet Governance. *Lawfare*. July 2022.

In their article, Chris Riley and Susan Ness propose that modularity can be a useful approach to improve digital platform accountability through harmonised policies and practices among democracies embracing the rule of law. Through multi-stakeholder, co-regulatory governance, "modules" – discrete mechanisms, protocols, and codes – are developed to enable internationally aligned corporate technical and business practices. The authors propose that modularity involves five steps: problem identification (e.g., lack of data access), module formation (e.g., a group of experts develops standards for vetting procedures), validation (e.g., government approves modules as sufficient), execution (e.g., researchers apply for clearance), and enforcement and analysis (e.g., government ensure compliance).

The authors emphasise need for better and more global alignment of privacy and platform regulations, in particular between the United States and Europe. The EU-U.S. Trade and Technology Council could help reduce this gap, but absent major legislation by the U.S. Congress, the authors say, effective alignment remains impossible. In addition to growing transnational regulatory differences, governance faces constantly evolving user behaviour and online harms. The article argues that governance mechanisms should create incentives for continued investment and assessment, resulting in an improved baseline of behaviour.

Modularity would offer a means of making it easier to adapt expectations for corporate practices through diverse input without the need for legislative change. In the case of auditing requirements in the EU's Digital Services Act, an audit module could be created through collaboration across national borders with auditors, platform policy and integrity workers, and third-party stakeholders including civil society and, where appropriate, government experts. Such process could develop vetting mechanisms and minimum standards for hiring audit firms and conducting audits. The EU, alongside other legal jurisdictions, could then recognise the module in its enforcement, granting the resulting audits legitimacy. Transatlantic effort could focus on common modular researcher vetting processes to be developed by a multi-stakeholder body that could include researchers, platform representatives, and representatives of the European Commission, the US government, or other governments authorising the module.

In sum, cross-border collaboration on common processes and codes of practice through modularity could facilitate greater regulatory consistency across jurisdictions, reducing conflicting requirements and implementation costs, and improving compliance with lower regulatory costs for governments.

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