Digital Policy Lab

Policy Digest #9

7 March 2023

ISD Powering new generations against extremism

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

Australia: eSafety's Report on the Basic Online Safety Expectations

Type Regulatory
Status Published

On 15 December 2022, Australia's eSafety Commissioner released a report assessing online service providers' implementation of the "Basic Online Safety Expectations" with respect to child sexual exploitation and abuse (CSEA). The report summarises responses provided by Apple, Meta (Facebook and Instagram), WhatsApp, Microsoft, Skype, Snap and Omegle pursuant to legal notices issued by eSafety in August 2022 under Australia's Online Safety Act 2021, which requires a response within 28 days under the threat of daily fines of up to $550,000 AUD. The report criticises Apple and Microsoft for not attempting to proactively detect previously identified CSEA material stored on iCloud or OneDrive respectively, despite the availability of technologies to enable this (for example via PhotoDNA detection technology, or the live streaming of CSEA content via Skype, Microsoft Teams, or FaceTime). eSafety also described Apple's decision to drop the proactive scanning of CSEA material on iCloud as "a major step backwards from its responsibilities to help keep children safe from online sexual exploitation and abuse." In December 2022, Apple decided to drop the CSEA-detection tool following widespread criticism from privacy and digital rights groups who were concerned that the capability itself could be abused to undermine the privacy and security of iCloud users. eSafety's report further highlights significant disparities in companies' response times to user reports of CSEA on their services (from four minutes on Snap to two days for Microsoft), the lack of on platform or in app reporting on Apple or Omegle services, and a lack of cross-platform coordination by companies such as Meta that operate multiple services to identify perpetrators with accounts on multiple platforms, or to prevent the creation of new accounts.

EU: Digital Services Act

Type Regulatory
Legislative status In force

The Digital Services Act (DSA), which was published in the Official Journal of the EU on 27 October and entered into force on 16 November 2022, obliges very large online platforms (VLOPs) and very large online search engines (VLOSEs) to conduct risk assessments and mitigation measures on their services, including 'systemic risks' related to the dissemination of illegal content, negative effects on fundamental rights, civic discourse and electoral processes, public security, and gender-based violence or mental health. The DSA establishes a threshold of 45 million average monthly active recipients of the service (users) as the criterion to designate a service as a VLOP or VLOSE. Providers were required to publish their user numbers in a publicly available section of their online interface by 17 February 2023. On 1 February, the European Commission published

[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.

a Q&A to assist platforms and search engines with their publication obligation. The Q&A provides guidance of what "active recipients" mean, whilst noting that only the Court of Justice of the EU is competent to authoritatively interpret EU law. On 17 February, tech companies such as Google, Meta, Microsoft, TikTok, Twitter, or Snapchat reported on their user numbers. While the DSA will apply to all service providers from 17 February 2024, VLOPs and VLOSEs will be subject to the rules four months following their designation as such. Following the receipt of user numbers, Commission officials indicated it could take between two and six weeks for the official designation of VLOPs and VLOSEs to be confirmed.

EU: Proposed European Media Freedom Act

Type Regulatory
Status Awaiting committee decision (Ordinary legislative procedure)

In September 2022, the European Commission adopted the proposed European Media Freedom Act (EMFA). Building on the DSA, the EMFA includes additional safeguards for editorial integrity of content provided online by media service providers that adhere to certain standards. It will also establish a new European Board for Media Services comprised of national media authorities that will organise a dialogue between VLOPs and the media sector to promote access to diverse media offers (see also Policy Digest #8).

On 23 January 2023, the Commission closed a feedback period, the results of which will be presented to the European Parliament and Council, with the aim of feeding into the legislative debate. In Parliament, the Committee on Culture and Education (CULT) has been 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. In the Council of the EU, the Audiovisual and Media Working Party examined the Commission's proposal at meetings held between September and November 2022. In a progress report, presented on 29 November 2022, several Member States asked for clarification about who is covered by the definition of 'editor', while others raised the issue of subsidiarity given the interaction with national criminal procedure. Under the current Swedish Presidency of the Council, Culture Ministers of Member States will take part in an Education, Youth, Culture and Sport Council meeting on 15-16 May 2023, during which negotiation results and conclusions regarding the proposal will be presented.

In a public statement, published on 24 January 2023, civil society organisations called for the rejection of Article 17 of the EMFA, which puts forward the proposal for "media privilege." The statement argues that the "ex-ante notification to self-declared media" establishes "de facto fast-track, non-transparent procedures to certain privileged actors that will have a major negative impact on the right to freedom of expression and information, even opening the door to rogue actors (…)." The statement further criticises that Article 17 entrusts VLOPs with a discretionary power to assess the integrity and reliability of media service providers' self-declaration, which "would grant them even more power to shape the public sphere."

EU: Proposed regulation on the transparency and targeting of political advertising

Type Regulatory
Status Trilogues (Ordinary legislative procedure)

The European Commission proposal for a Regulation on the transparency and targeting of political advertising, presented in November 2021, seeks to introduce a common regulatory framework for sponsored political advertising applying to both online and offline environments in the run-up to elections and between elections. The proposal would impose obligations on providers of political advertising services. Among other obligations, political advertisements would have to be labelled

as such to distinguish them from editorial content and be accompanied by certain information and a transparency notice, referring to the sponsor's identity and contact details, the period of publication, and the amounts spent and their sources. The proposal also seeks to ban targeting and amplification techniques that involve the processing of sensitive personal data.

On 13 December 2022, the Council of the EU adopted its mandate (general approach) for negotiations with the European Parliament. In its approach, the Council seeks to provide greater legal certainty regarding the regulatory scope, including what is to be considered "political advertising." Among other things, the Council makes it clear that the "regulation will not affect the content of political advertisements nor the EU or member states' rules on aspects not covered by the regulation." On cross-border cooperation, the Council "has aligned the new rules with the Digital Services Act to reflect the country-of-origin principle." On 2 February 2023, the negotiating mandate proposed by the IMCO Committee was approved by the European Parliament plenary. The IMCO position streamlines the scope of application, "excluding political views expressed under the editorial responsibility of a media service, and providing a non-exhaustive list of elements to be taken into account to determine whether a message is a political advertisement." The report also provides for the creation of a "European repository for online political advertisements," managed by the Commission and including all online political advertisements and the information provided in their transparency notices. The two co-legislators, the Parliament and the Council, are now commencing the interinstitutional negotiations in the trilogues.

Ireland: Future of Media Commission Report and the Online Safety and Media Regulation Act 2022

Type Regulatory
Status Signed into law

On 18 January 2023, the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media (TCAGSM) published the "Future of Media Commission Report – Implementation Strategy & Action Plan." The Future of Media Commission (distinct from the Media Commission below) was tasked with "addressing how media should serve Irish society, assessing how well the current system meets these goals and considering what changes ought to be made to support print, broadcast, and online media in a platform agnostic fashion." A key recommendation of the Report is the establishment of a new body with a regulatory and developmental function in respect of the media.

The Online Safety and Media Regulation Act 2022, which was signed into law by the President of Ireland on 10 December 2022, provides for the establishment of such a new media regulator, the Coimisiún na Meán - CnM (Media Commission). Two of the initial three Commissioner roles (the Broadcasting Commissioner and the Media Development Commissioner) will play a significant role in delivering on the Government's implementation of the Report. Once established, the CnM will replace the Broadcasting Authority of Ireland (BAI) and oversee the new regulatory regime with the goal of improving online safety. It will also be the national agency (or Digital Services Coordinator, DSC) charged with implementing the obligations for intermediary services under the Digital Services Act (DSA). The CnM is to be responsible for implementing online safety obligations, creating Online Safety Codes for "designated" service providers, and establishing a system for nominated organisations to report concerns about online service providers. It will have the ability to investigate and impose penalties of up to EUR 20 million or 10% of the entity's turnover and compel service providers to restrict access to certain internet services or audio-visual on-demand media services.

UK: Online Safety Bill

Type Regulatory
Status Passed Second Reading

In December 2022, several clauses and schedules in the new version of the Bill as well as a list of proposed amendments from the Department for Digital, Culture, Media & Sport (DCMS) were sent back to the Public Bill Committee for scrutiny. In a two-day committee debate between 13-15 December, the Committee approved all the tabled amendments. On 31 December, Labour's Shadow Secretary of State for Digital, Culture, Media and Sport Lucy Powell committed to attempting to amend the Bill again when Parliament returns to session. The move came after the provisions for an 'Adult Safety Duty' were removed by the Government in November 2022, and replaced with a 'triple shield' to ensure "social media firms will be legally required to remove illegal content, take down material in breach of their own terms of service, and provide adults with greater choice over the content they see and engage with." At the same time, the government has proposed making certain types of "harmful" content illegal through the introduction of new offences, including the promotion of suicide, self-harm or eating disorders. Many previous concerns remained, however. For example, the Bill allows the regulator Ofcom to order platforms to use "proactive technology" to identify and remove content that falls in scope of the Bill, even though proactive monitoring technologies often have high rates of inaccuracy.

On 25 January 2023, Online Safety Bill Second Reading debate took place in the House of Lords. The debate was notable for the cross-party consensus that the Bill was long overdue. There was strong cross-party backing for amendments championed by Baroness Kidron (strengthening children's protections), Lord Bethell (introducing stronger provisions to prevent children accessing pornography), Baroness Morgan (bringing in a Violence Against Women and Girls (VAWG) code of practice) and Baroness Stowell (reducing the powers of the Secretary of State). Other issues that were discussed included: whether the proposed user empowerment tools should be on by default; the lack of risk assessments for adult harms; media literacy provisions; anonymity; categorisation of platforms; and the media exemption. Having passed its Second Reading, the Online Safety Bill moved to Committee stage, which is likely to start towards the end of March. Amendments are already being tabled. The government has already delayed the end of the current parliamentary session from Spring 2023 until Autumn 2023, potentially allowing for more time for the bill to be amended.

In addition, a Cabinet reshuffle announced on 6 February has seen the creation of four new departments across Whitehall. Prime Minister Rishi Sunak announced the formation of the new Department for Science, Innovation and Technology (DSIT), which will combine responsibilities for tech-related policies that were previously split across the now former Department for Digital, Culture, Media and Sport (DCMS) and the Department for Business, Energy and Industrial Strategy (BEIS). Alongside DSIT, the new departments include the "refocused" Department for Culture, Media and Sport (DCMS). The online safety portfolio, including the Online Safety Bill, and the associated DCMS civil servants will move across to DSIT, although the Minister overseeing the Online Safety Bill in the House of Lords will remain at DCMS.

US-EU: Joint Statement of the Trade and Technology Council

Type Non-regulatory
Status Published

On 5 December 2022, the third US-EU Trade and Technology Council (TTC) Ministerial met in Washington DC. The Joint Statement notes that ten TTC Working Groups aim to support "sustainable, inclusive economic growth and development", promote "a human-centric approach to the digital transformation" and ensure "that international norms and the international trade rulebook are respected." The Working Groups on "Data Governance and Technology Platforms" and on "Misuse of Technology Threatening Security and Human Rights" are coordinating to address the "spread of Russian information manipulation and interference, particularly in the context of Russia's aggression against Ukraine, and its impact on third countries, notably in Africa and Latin-America."

The TTC Statement referred to the Joint Statement on Protecting Human Rights Defenders Online, released in advance of the meeting, to promote an "open, free, global, interoperable, reliable, and secure Internet." The Statement also urges "companies to prevent the misuse of their products and platforms, conduct due diligence, take effective action to address all forms of online violence and unlawful or arbitrary surveillance against human rights defenders." It calls on companies to "establish a grievance mechanism for internal and external reporting of misuse" and "provide a safe space for human rights defenders to carry out their work." The co-chairs intend to meet again in mid-2023 in Europe to review the joint work and discuss new ways to expand the transatlantic partnership.

Section 2 Topic-specific snapshot: "Online misogyny and gender-based violence (GBV)"

This section presents summaries of selected proposals, analyses and commentary published by governments, civil society, and academia on the topic of online misogyny and gender-based violence.

Monetizing Misogyny - Gendered Disinformation and the Undermining
(Lucina Di Meco, *#ShePersisted,* February 2023)

In their research, the global initiative #ShePersisted highlights the patterns, impact, and modus operandi of gendered disinformation campaigns against women in politics in Brazil, Hungary, India, Italy, and Tunisia. The report notes that the lawyer Cynthia Khoo coined the umbrella term "technology-facilitated gender-based violence" (TFGBV), which includes a spectrum of activities and behaviours, including both online gender-based violence and gendered disinformation. The presented case studies explore "how gendered disinformation has been used by political movements, and at times the government itself, to undermine women's political participation, and to weaken democratic institutions and human rights." Despite the regional and cultural diversity of the countries analysed, the research identifies several trends and patterns:

1. Gendered disinformation, paired with online abuse and violence, is a pervasive problem faced by women in politics and represents a very significant barrier to their political participation and freedom of expression.
2. Gendered disinformation should be treated as an early warning system for both backsliding on women's rights and the erosion of democratic principles and institutions.
3. Women who come from traditionally marginalised sections of society are made even more vulnerable.
4. Social media platforms have failed to protect their users or to deliver their early promises of being an equalising and democratising force.
5. While there is no silver bullet to address gendered disinformation, there are policies and practices that can support addressing this problem and in-depth, targeted research is needed to test and refine them.

The report emphasises that greater investments and focus need to be directed in two areas: "legislative frameworks and approaches" on the one hand, and "targeted, strategic, solution-oriented research and programs" on the other.

An Intersectional Lens on Online Gender Based Violence and the Digital Services Act
(Asha Allen, *CDT for Verfassungsblog,* November 2022)

Asha Allen, Advocacy Director for Europe, Online Expression & Civic Space at the Center for Democracy and Technology (CDT), argues for the use of intersectional methodology (as presented by Kimberlé Crenshaw) when conducting risk assessments and mitigation measures anticipated in the Digital Services Act (DSA). Notably, the DSA obliges VLOPs to assess and mitigate systemic risks they create, which includes online GBV.

Allen notes that online GBV "exists on a spectrum and can take many forms, including actions that may not rise to the level of illegal conduct, which nevertheless have a chilling effect on women and non-binary people's speech." Risk assessments thereby should apply a holistic approach and be accompanied by effective accountability mechanisms. Allen emphasises that a "comprehensive understanding of how forms of discrimination [based on social categorisations such as race, class, and gender] may intersect will need to be developed," using

intersectionality as an analytical framework. This approach would not only identify specific impacts but their severity and how they may intersect with other fundamental rights violations. A first step would thereby need to have "specific indicators related to the experience of online GBV amongst different marginalised groups, before then assessing how this systemic risk intersects with others." For example, Allen notes that "gendered disinformation is based on misogyny but can simultaneously intersect with discrimination based on racism, ableism, religious identity", and thereby poses risks to fundamental rights, including to human dignity and freedom of expression, to civic discourse and electoral processes, and to person's physical and mental well-being – all of which are specifically identified in the DSA.

Civil society consultations should accompany this intersectional approach. For example, Allen notes that researchers could "request cross-sectional data points as one way of developing comprehensive analyses" and "assessing the impact on marginalised communities." Allen further recommends the establishment of "a formal mechanism by which civil society can actively participate, evaluate and provide recommendations for improved enforcement and implementation."

Proposal for a Directive on combating violence against women and domestic violence
(European Commission, March 2022)

The inclusion of online GBV as a systemic risk in the DSA aligns with the EU's aim to criminalise certain forms of online violence ('cyber violence') in the proposed Directive on combating violence against women and domestic violence, published in March 2022. The European Parliament has repeatedly called on the Commission to propose legislation, including proposals to add gender-based violence as a new area of crime listed in Article 83(1) of the Treaty on the Functioning of the European Union (TFEU).

Noting a fragmentation of legislation and "significant legal gaps" at both EU and Member State level, the proposed Directive recognises various forms of 'cyber violence', including non-consensual sharing or manipulation of intimate material, cyber stalking and cyber harassment. It notes that "cyber violence particularly targets and impacts women politicians, journalists and human rights defenders", with "the effect of silencing women and hindering their societal participation on an equal footing with men" (r. 17). It further asserts, "the use of information and communication technologies bears the risk of easy, fast and wide-spread amplification of certain forms of cyber violence with the effect of creating or enhancing profound and long-lasting harm for the victim" (r. 18).

The proposed Directive contains measures to strengthen victims' access to justice and rights to appropriate protection, including "the removal of online content in relation to offences of cyber violence, and a possibility of judicial redress for the affected users" (Art. 25). Given that the DSA does not provide an EU-level definition of what constitutes illegal content, the Directive aims to complement the DSA by setting minimum rules for the offence of 'cyber violence', which is defined as any act of violence covered by this Directive that is committed, assisted or aggravated in part or fully by the use of information and communication technologies" (Art. 4). Specifically, the Directive sets minimum rules for certain 'computer crime offences': non-consensual sharing of intimate or manipulated material (Art. 7), offences concerning cyber stalking (Art. 8), offences concerning cyber harassment (Art. 9), and cyber incitement to hatred or violence (Art. 10).

Analysis of the Directive on combating violence against women and domestic violence
(Rita Jonusaite (*EU DisinfoLab*), Maria Giovanna Sessa (*EU DisinfoLab*), Kristina Wilfore and Lucina Di Meco (*#ShePersisted*), October 2022)

An analysis by EU DisinfoLab and #ShePersisted of the Commission proposal for a 'Directive on combating violence against women and domestic violence' argues that the text "is limited in scope and does not consider the impact of harmful content, notably gender-based disinformation (GBD)." The authors assert that online violence and GBD are not mutually exclusive and that GBD should be specifically acknowledged in the Directive.

The analysis notes that "disinformation is often the precedent, background, and trigger of violence", while recognising that "disinformation should not be criminalised in the same way as the offences lined up in the Directive." It recommends that GBD should also be "specifically considered when VLOPs identify and mitigate systemic risks under the DSA, paying particular attention to coordinated campaigns." Another overarching recommendation would be to ensure that platforms "introduce policies, remedies, and mechanisms that are tailored from a gender perspective across all aspects of the platform, and that are designed in consultation with those affected."

A specific recommendation would be for platforms to "introduce a rapid response system where victims can flag GBD in a manner tailored to prevent them from reliving their traumatic experience and is addressed with priority (Article 16 of the Directive) to avoid further harm." The authors conclude that "the Directive must display a sound understanding of the networked nature of disinformation and extremism as reproduced online."

Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the 'Manosphere'
(Elise Thomas & Kata Balint, *ISD,* April 2022)

In 2022, ISD documented how YouTube's algorithms contributed to promoting misogynistic, anti-feminist and other extremist content to Australian boys and young men. Using experimental accounts, the research tracked the content that YouTube and the 'YouTube Shorts' feature routinely recommended to boys and young men.

This short-term, qualitative study involved analysing algorithmic recommendations provided to 10 experimental accounts. As the study progressed, each account was recommended videos with messages antagonistic towards women and feminism. Following the recommendations and viewing and liking the suggested content resulted in more overtly misogynist 'Manosphere' and 'incel' content being recommended.

The study found that while the general YouTube interface recommended broadly similar content to topics the accounts originally engaged with, 'YouTube Shorts' appeared to operate differently. 'Shorts' seemed to optimise more aggressively in response to user behaviour and show more extreme videos within a relatively brief timeframe. On 'Shorts', all accounts were shown vastly similar and sometimes even the same specific content from right-wing and self-described 'alt-right' content creators. Moreover, the algorithms did not appear to make any distinction between the underage and adult accounts in terms of the content served.

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.