Digital Policy Lab

Policy Digest #20

19 May 2025

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

European Union: Preliminary findings indicate TikTok's ad repository violates the Digital Services Act

Type Regulatory (compliance review)
Status Preliminary findings issued

On 15 May 2025, the European Commission stated that it has informed TikTok of its initial assessment indicating that the company has not met the requirements of the Digital Services Act (DSA) regarding the publication of an advertisement repository. This repository should enable researchers and civil society to identify deceptive advertisements and misinformation, particularly in the context of elections. The Commission noted deficiencies in the information TikTok provides about ads, including their content, targeted users, and who pays for them, as well as limitations in public access to this information. TikTok now has the opportunity to respond to these preliminary findings and review the relevant investigation documents. If the Commission's views are validated, it may lead to a non-compliance decision, potentially resulting in fines of up to 6% of TikTok's global annual revenue and additional oversight to ensure compliance.

European Union: Belgian court declares transparency and consent framework illegal, having decisive effect across Europe

Type Litigation
Status In effect

On 14 May 2025, the Belgian Court of Appeal ruled that the transparency and consent framework (TCF), is illegal. This accountability framework, aimed to facilitate with compliance with provisions of the ePrivacy Directive and GDPR, has been used by major technology companies such as Google, Microsoft and Amazon for data processing. This decision confirms findings from 2022 that highlighted multiple infringements by the TCF. As for next steps, the ruling will be implemented immediately across Europe, compelling the TCF and associated parties to re-evaluate their compliance with data protection laws and improve their practices accordingly.

European Union: European Commission launches consultation on enhanced online protection guidelines for minors

Type Regulatory (consultation)
Status Open

On 13 May 2025, the European Commission opened a public consultation on guidelines to enhance online protection for minors under the DSA. These guidelines focus on online privacy, safety, and security for child-accessible platforms, proposing

[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.

measures like age assurance mechanisms to limit children's exposure to inappropriate content, default private settings for children's accounts, and adjusted recommender systems prioritising user feedback. Recommendations also suggest children to block and mute users and require explicit consent for group participation to mitigate cyberbullying risks. The consultation is open until 10 June 2025, with final guidelines expected by the end of Q3 2025. Additionally, the Commission is developing an age-verification app to safely check the age of users ahead of the EU Digital Identity Wallet launch in late 2026, with technical specifications available on GitHub. A proposal for a Digital Fairness Act is also underway to address other digital issues for minors not covered by the DSA.

European Union: European Commission launches consultation on format and specifications for political advertisement labels and notices under implementing regulation

Type Regulatory (consultation)
Status Open

On 30 April 2025, the European Commission launched a public consultation on a draft implementing regulation that outlines the format, template, and technical specifications for the labels and transparency notices accompanying political advertisements. This draft regulation is founded on the Regulation on the transparency and targeting of political advertising (Regulation (EU) 2024/900), which seeks to create consistent standards for the labelling and presentation of such advertisements. These requirements will apply to various media formats, including print, audiovisual, and digital platforms. For electronic advertisements, the draft stipulates that transparency notices must be provided in machine-readable formats, containing comprehensive information on targeting, sponsorship, financing, and ad delivery. The consultation is open until 28 May 2025 and will inform the implementation of the regulation to be in force from 10 October 2025.

European Union: Apple fined EUR 500 million and Meta fined EUR 200 million for breaching Digital Markets Act (DMA)

Type Regulatory (enforcement)
Status Decision issued

On 23 April 2025, the European Commission ruled that Apple and Meta violated the Digital Markets Act (DMA), resulting in fines of €500 million for Apple and €200 million for Meta. Apple failed to allow app developers to inform customers about alternative purchasing options outside its App Store, which limited consumer choice and access to cheaper offers. Meanwhile, Meta's 'Consent or Pay' model did not provide users with an appropriate alternative for less personalised services, infringing on their rights. Both companies must comply with the Commission's decisions within 60 days to avoid further penalties.

Germany: Berlin upholds access restrictions to adult websites citing minors' protection and legal non-compliance

Type Litigation
Status Decision issued

On 29 April 2025, the Berlin Administrative Court (Verwaltungsgericht Berlin) dismissed urgent applications from Aylo Freesites Ltd., a Cyprus-based operator of major adult websites such as Pornhub and YouPorn, challenging access-blocking orders issued by the Berlin-Brandenburg Media Authority (Medienanstalt Berlin-Brandenburg) to a local internet service provider. The Court determined that the content provider did not have a valid interest in halting the blocking measures, given

its continued distribution of unrestricted pornographic content despite previous prohibitions and its failure to implement legally mandated age-verification systems. The ruling highlighted the protection of minors as a primary concern and deemed the ongoing neglect of legal responsibilities as non-compliance. This decision is subject to appeal at the Higher Administrative Court of Berlin-Brandenburg (Oberverwaltungsgericht Berlin-Brandenburg).

Germany: Regional courts provide guidance on Meta's user data processing for AI training

Type Regulatory (guidance)
Status In effect

On 15 and 17 April 2025, the Hamburg Data Protection Commissioner (Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit) and the Saxon Data Protection Commissioner (Sächsische Datenschutz- und Transparenzbeauftragte) respectivley issued guidance concerning Meta's upcoming use of personal data from adult European Facebook and Instagram users to train its artificial intelligence systems, set to begin at the end of May 2025. The data, including public posts and photos, will support tools such as Meta's language models and AI chatbot on WhatsApp. Before the guidance, users were informed of their right to object to this use of their data, with no action needed if they agree. However, both authorities warned that objections must be submitted by users before the end of May, as once the data is used for training, it cannot be extracted from the AI systems.

Ireland: TikTok fined EUR 530 million for transferring European users' personal data and breaching transparency rules

Type Regulatory (enforcement)
Status Decision issued

On 2 May 2025, the Irish Data Protection Commission (An Coimisiún um Chosaint Sonraí, DPC), the Lead Supervisory Authority for TikTok, fined TikTok EUR 530 million for transferring European Economic Area (EEA) user data to China and breaching transparency requirements under the General Data Protection Regulation (GDPR). Besides the EUR 530 million fine, the DPC mandates TikTok to rectify its processing within six months. Additionally, the decision includes an order to halt TikTok's data transfers to China if TikTok is not compliant within this period.

Italy: AGCOM introduces new age verification regulations for online platforms

Type Regulatory
Status In effect

On 18 April 2025, the Italian Communications Regulatory Authority (Autorità per le Garanzie nelle Comunicazioni, AGCOM) issued Resolution 96/25/CONS, which established new age verification requirements for online users. This resolution mandates that video-sharing platforms and websites providing content in Italy implement age verification systems. These systems must be developed in consultation with relevant stakeholders and reviewed by the European Commission, and they are to use certified third parties to carry out a two-step identification and authentication process. The guidelines emphasise data minimisation and user anonymity, allowing platforms a six-month period to comply. The rules are designed in accordance with principles of proportionality, accessibility, and non-discrimination, with flexibility for future adjustments following EU directives.

Switzerland: Federal Data Protection and Information Commissioner issues key guidelines on AI data processing under new Data Protection Act

Type Regulatory (guidance)
Status In effect

On 8 May 2025, the Federal Data Protection and Information Commissioner (FDPIC) published guidance confirming that the Data Protection Act (DPA), which has been in effect since 1 September 2023, applies to data processing activities involving AI. This guidance follows Switzerland's ratification of the Council of Europe Convention on Artificial Intelligence and Human Rights in March 2025. The FDPIC highlighted that the DPA is designed to be technology-neutral and requires transparency in AI operations, including the disclosure of their purposes, functionalities, and data sources. Furthermore, the DPA grants data subjects the right to object to automated processing and to request human evaluation of automated decisions. High-risk AI applications are required to undergo data protection impact assessments, while those that compromise privacy, such as large-scale real-time facial recognition, are banned.

United Kingdom: Ofcom launches Online Information Advisory Committee

Type Regulatory (administrative)
Status In effect

On 28 April, Ofcom announced the establishment of its Online Information Advisory Committee. In accordance with Section 152 of the Online Safety Act (OSA), Ofcom is mandated to set up and oversee an advisory Committee. This Committee is chaired by Lord Allan of Hallam who was appointed by Ofcom. Ofcom Board appointed Elisabeth Costa (Chief of Innovation and Partnerships at the Behavioural Insights Team (BIT)), Jeffrey Howard (Professor of Political Philosophy & Public Policy at University College London), Will Moy (Chief Executive of the Campbell Collaboration), Mark Scott (Senior Resident Fellow at the Atlantic Council's Digital Forensic Research Lab's (DFRLab)) and Devika Shanker-Grandpierre (Strategic Advisor).

United Kingdom: Ofcom publishes major safety updates, including draft Codes to protect children online

Type Regulatory
Status Published

On 24 April 2025, Ofcom published multiple updates aimed at enhancing online safety for children. Key developments include the publication of new Guidance on Content Harmful to Children and the release of the draft Protection of Children Codes of Practice for both user-to-user services and search services. Additionally, updated guidance on highly effective age assurance for part 3 services has been issued, outlining the expectations around age verification processes. Ofcom has also revised its Children's Access Assessments Guidance and introduced the new Children's Risk Assessment Guidance and Children's Risk Profiles, providing frameworks for evaluating potential online harms. To support consistent understanding across the sector, Ofcom has published a Children's Register of Risks Glossary, supporting service providers in identifying and mitigating risks to young users. Services are required to carry out a child access assessment by 16 April 2025, and a child risk assessment, if necessary, by 24 July 2025, for any areas of the service that are not age-restricted.

United States: Congress passes TAKE IT DOWN ACT to criminalise non-consensual intimate imagery

Type Legislative
Status Passed (pending Presidential signature)

On 14 May 2025, the U.S. House of Representatives overwhelmingly passed the bipartisan TAKE IT DOWN Act by a vote of 409–2. The legislation, which had previously cleared the Senate with unanimous consent in February, now proceeds to the President to be signed into law. The TAKE IT DOWN Act criminalises the distribution of non-consensual intimate imagery (NCII), including AI-generated material such as deepfake pornography. It also introduces a requirement for social media platforms and similar online services to remove such content within 48 hours of receiving a victim's notice, with enforcement to be overseen by the Federal Trade Commission (FTC).

United States: Federal Trade Commission issues updated rules on children's privacy

Type Regulatory
Status Final (implementation pending)

On 22 April 2025, the FTC published updated rules under the Children's Online Privacy Protection Act (COPPA), marking the first significant revision since 2013. The updated framework prohibits targeted advertising to children without verifiable parental consent and imposes limits on how long companies may retain children's data. In addition, the definition of 'personal data' has been expanded to include biometric information, while 'contact information' now explicitly includes telephone numbers. The updated regulations, which were finalised in January and formally published in the Federal Register in May 2025, are set to take effect on 23 June 2025. However, companies will have until 22 April 2026 to ensure full compliance.

Section 2 Topic-specific snapshot: "Aligning Safety by Design Principles Across Regulatory Frameworks"

This section summarises selected analyses and responses published by government agencies, civil society organisations and academia on safety by design principles across regulatory frameworks.

Explainer: Safety by Design,
Professor Lorna Woods, 2024

This explainer examines the role of Safety by Design (SbD) within the UK's Online Safety Act, highlighting how the Act reframes regulatory expectations by placing responsibility on service providers to prevent harm through design choices. Drawing on legal interpretation, policy precedent, and comparisons to similar "by design" approaches, Prof. Lorna Woods points out that design decisions such as algorithms, user interfaces, and engagement incentives fundamentally shape user experience and risk exposure. It argues that without proactive safe design, reactive moderation tools are likely to fall short.

By looking at Clause 1 of the Act, the author highlights how the legislation positions SbD not as an abstract principle, but as a guiding objective for compliance. It therefore provides an interpretive benchmark: platform duties must be understood with reference to the design and operation of services, including systemic risks embedded in algorithms, user interfaces and monetisation strategies. Rather than treating harmful content in isolation, the Act's emphasis on SbD recognises that safety outcomes are influenced by upstream decisions about how platforms are designed and built. This shift aligns with broader "by design" frameworks, such as Privacy by Design and Security by Design, by calling for proactive mitigation of foreseeable harms, especially those affecting children and vulnerable users.

The author outlines several key principles and recommendations:

1. A core premise of SbD is that safety must be embedded from the earliest stages of product development, not treated as an afterthought. Platforms should anticipate how different user groups, including children, those with disabilities, or people with lower digital literacy may interact with their services, and design with these needs in mind. Rather than relying on users to protect themselves, platforms are expected to take meaningful, preventative steps to reduce risk through interface design, functionality limits, and default settings that prioritise user well-being.

2. Systemic and ongoing Risk Assessments: From initial concept to post-launch updates, platforms must identify, test for and respond to potential harms. This includes using methods such as abusability testing and red-teaming to uncover design vulnerabilities, especially those that could be exploited at scale. Importantly, platforms must also monitor how real-world use reveals new risks, and be prepared to pause, modify or withdraw features that are shown to cause harm.

3. It is on platforms to critically reassess incentive structures and develop alternative success metrics that reflect long-term user well-being, not just short-term attention. Business models should not make safety subordinate to profitability.

Towards digital safety by design for children,
OECD Digital Economy Papers, 2024

This OECD report provides a comprehensive overview of how digital products and services can be proactively designed to protect children online. It argues that just as physical environments for children are developed with safety in mind (e.g., car seats, playgrounds), digital spaces should embed child protection features from the earliest stages of design and development. This approach shifts the burden of safety away from children and parents toward service providers and regulators.

This report outlines growing international legal developments across OECD member countries, including the EU's DSA, Australia's OSA, and the UK's OSA. In different ways, all these regulatory frameworks impose obligations such as age verification, risk assessments, and content moderation to protect minors. However, the authors warn of the risks of regulatory fragmentation and calls for a harmonised, child-rights-based approach. Key components of safety by design include age assurance mechanisms, child-friendly information, proactive harm detection, and privacy-preserving defaults. It also stresses the importance of empowering children through accessible complaint systems, inclusive design, and participation in decision-making processes, exemplified by initiatives like Australia's eSafety Youth Council.

By presenting a variety of case studies, the report demonstrates that different types of digital services, regardless of whether they are specifically designed for children. It also highlights, that labelling a service as "not for children" is not sufficient, as children often use platforms not originally intended for them.

In sum, the report advocates for embedding safety as a core design objective in the digital environment to ensure children can benefit from technology without undue risk. To facilitate these steps, it supports policymakers, service providers, and regulators with concrete guidance:

1. **Child-centered design:** Products and interfaces should be developed with children's developmental needs, capacities, and vulnerabilities in mind; not simply adapted from adult services. Services should be safe by default, without requiring users or parents to opt into protections.

2. **Privacy by design:** Children's data should be minimally collected, stored securely, and not used for profiling or targeted advertising.

3. **Child-friendly information:** Terms of service and privacy policies must be understandable and usable for children, including visual or interactive formats. Children must be able to report problems easily, with clear follow-up procedures and support.

4. **Participation of children:** Children should be consulted and included in design and policy decisions that affect their digital lives.

Technology, gendered violence and Safety by Design: An industry guide for addressing technology-facilitated gender-based violence through Safety by Design,
eSafety Commissioner Australia, 2024

This guide outlines how digital platforms and technologies can be weaponised to perpetrate technology-facilitated gender-based violence (TFGBV). To support companies in preventing such abuse, these guidelines provide by embedding safety, accountability, and user empowerment into the design and governance of digital products and services. The guide highlights how TFGBV is distinct in its scale, speed, and persistence, often executed anonymously, cheaply, and across platforms. It emphasises that tech companies must take proactive responsibility, by applying Safety by Design (SbD) principles throughout the product lifecycle. Depending on their design and use, AI and algorithmic systems can either exacerbate or help reduce TFGBV. The guide highlights the following examples:

- An estimated 96% of all deepfake videos online are non-consensual sexual content, with 99% depicting women. Free apps can "nudify" images, exposing significant risks for user privacy, consent, and digital safety. To address this issue, the authors outline the need for platform policies (e.g. YouTube's AI content disclosure rules) and shared attribution standards (e.g. C2PA).
- While generative AI is used to create deepfake pornography and gendered disinformation, it also holds promise in automating abuse detection and moderating harmful content. According to this guide, Tinder's "Does This Bother You?" feature increased user reporting by 46%, and Bumble's Private Detector uses AI to blur unsolicited sexual images before users view them.
- The report further highlights case examples where algorithms have promoted harmful content or figures, reinforcing rape myths and gender stereotypes. The guide recommends algorithmic auditability and restrictions on content discoverability, as seen in TikTok's filtering of hate-related search terms.

The guide calls for embedding the three SbD principles, 1) service provider responsibility, 2) user empowerment, and 3) transparency, into every layer of digital product development. It recommends:

- Assigning internal accountability teams for TFGBV,
- Establishing fast, visible reporting and redress mechanisms,
- Conducting routine risk assessments for abuse pathways, and
- Publicly reporting on TFGBV enforcement and innovations.

Building a safe and accountable internet: CCDH's refreshed STAR framework,
Center for Countering Digital Hate (CCDH), 2024

The STAR Framework, developed by the Center for Countering Digital Hate (CCDH), offers a structured approach to regulating social media platforms, based on four core principles: Safety by Design, Transparency, Accountability, and Responsibility. It responds to what the report identifies as a growing body of evidence that design and governance choices made by platforms enable the spread of harm, including disinformation, abuse, and hate speech. The report underscores that many harms persist not due to a lack of tools, but because safety measures are often underused, difficult to access, or inconsistently enforced. For instance, parental controls and reporting mechanisms are rarely utilised, raising questions about whether they are fit for purpose, raising questions about whether they are fit for purpose.

The report also notes that some commonly cited metrics, like content "prevalence," can obscure the reach and impact of harmful material, especially when content with wide exposure still represents a small fraction of total posts. Even content viewed millions of times can be dismissed as insignificant because it represents a small percentage of total posts. This can conceal the effects of tangible harm and offline consequences.

The four core principles of the STAR framework, as listed above, are backed with actionable proposals. These include independent audits of recommender algorithms, and default safety settings, especially for vulnerable users. To strengthen transparency, it recommends standardised public safety reports, improved researcher access to platform data, and whistleblower protections. Regulators should have powers to investigate, fine, and enforce changes, while legal reforms, such as conditional platform immunity, would link liability to demonstrated safety efforts. Internally, companies should appoint senior safety leads, and governments can support safer design through incentives and cross-sector collaboration.

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 Alfred Landecker Foundation for their support for this project.