Digital Policy Lab

Policy Digest #17

2 December 2024

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: Bill to ban children under the age of 16 from using social media announced

Type Legislative
Status Announced

On 7 November 2024, Australia's Prime Minister Anthony Albanese and Communications Minister Michelle Rowland announced their intention to introduce a bill to Parliament later this year, banning children under the age of 16 from using social media. The legislation does not include carve-outs for young children to use social media with parental consent or a grandfather clause, meaning current under-age users must cease using social media if enacted. Under-age children and their parents would not be penalised for violation of the law, but rather the social media platforms themselves. The bill has received criticism from platforms and digital rights groups which believe that a ban is too harsh.

Canada: Canadian Artificial Intelligence Safety Institute announced

Type Non-regulatory (Institute established)
Status Announced

On 12 November 2024, the Canadian government launched its Canadian Artificial Intelligence Safety Institute (CAISI). The Institute will promote the advancement of AI safety by working with key stakeholders, so governments understand and mitigate the risks from the deployment of AI systems. CAISI is a founding member of the International Network of AI Safety Institutes, so it will collaborate with international partners to create guidance on AI safety.

Canada: TikTok ordered to close its Canadian offices under the Investment Canada Act

Type Dissolution Order
Status Announced

On 6 November 2024, Canada ordered TikTok to close its Canadian offices in accordance with the Investment Canada Act, which enables the government to review foreign investments that could harm Canada's national security. The closure comes after Canada banned TikTok on all government devices last year. The shutdown does not affect Canadians' ability to use the app.

[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.

European Union: European Union Commission fines Meta €797.72 million for engaging in practices that unfairly advantage Facebook Marketplace

Type Litigation
Status Decision

On 14 November 2024, the European Commission imposed a fine of 797.72 EUR million against Meta for violating EU antitrust regulations by linking its online platform, Facebook Marketplace, to its social media platform, Facebook, resulting in unfair trading conditions on other online ads service providers.

European Union: European Commission receives the first draft of the General-Purpose AI Code of Practice

Type Voluntary Code of Practice
Status In consultation

On 14 November 2024, the European Commission received the first draft of the General-Purpose AI Code of Practice, written by independent experts. This draft, facilitated by the European AI Office, marks the initial deliverable in a four-stage drafting process that started on 30 September 2024 and will continue through four rounds until April 2025. The draft Code takes into account feedback from providers of general-purpose AI systems and examines international approaches. After a plenary session on 22 November, stakeholders had until 28 November to provide input during dedicated working group meetings. Following the feedback, the first draft will be reviewed based on the drafting principles which emphasise alignment between measures, KPIs, risks, and the size of AI model providers. The guidelines aim to ensure simplified compliance for SMEs and startups and provide exemptions for open-source providers. Additionally, the principles highlight the need for clear, yet adaptable requirements in response to technological advancements.

European Union: EU AI Office launches stakeholder consultation on guidelines on the definition and implementation of "unacceptable risk" AI systems under the AI Act

Type Regulatory (consultation)
Status Published

On 13 November 2024, EU AI Office launched a multi-stakeholder consultation concerning upcoming guidelines on defining AI systems and the management of AI practices deemed to pose unacceptable risks under the AI Act. These guidelines are meant to be published in early 2025 and aim to assist national authorities and AI providers in adhering to the AI Act's regulations ahead of their enforcement date on 2 February 2025. The AI Office invites contributions across sectors, including AI system providers, businesses, national authorities, academia, research institutions, and civil society. This consultation seeks additional practical examples to enhance the clarity on practical applications and use cases of the legal concepts defined in the Act. The consultation is open for submissions until 11 December 2024.

European Union: EU Commission and Consumer Protection Cooperation (CPC) Network notify Apple over alleged geo-blocking practices

Type Regulatory (Notification)
Status Response period

On 12 November 2024, the EU Commission and Consumer Protection Cooperation (CPC) Network notified Apple about several potentially illegal geo-blocking practices across its media services and requested the company to align its practices

in accordance with EU law. The network, led by consumer authorities from Belgium, Germany, and Ireland, found that Apple discriminates against consumers based on their residence, limiting their online access and payment options. For example, users can only access the services tailored to their registered country, are restricted to specific payment methods, and cannot download apps available in other EU/EEA countries. Apple has one month to respond to these findings and suggest measures to rectify the identified issues. Failure to comply may result in enforcement actions from national authorities.

European Union: EU Commission and Consumer Protection Cooperation (CPC) Network notify Temu over alleged consumer protection law infringements

Type Regulatory (Notification)
Status Response period

On 8 November 2024, the EU Commission and Consumer Protection Cooperation (CPC) Network notified the online marketplace Temu regarding various practices that contravene EU consumer protection laws. Following an investigation, the network, led by consumer authorities from Belgium, Germany, and Ireland, Temu is urged to rectify issues such as misleading discounts, pressure selling tactics, forced gamification and insufficient information on consumer rights. Additionally, concerns about fake reviews and inadequate contact options were raised. Temu has one month to respond to these findings and suggest measures to rectify the identified issues. Failure to comply may result in enforcement actions from national authorities.

European Union: EU enacts the right to reliable information under the European Freedom Act (EMFA)

Type Regulatory (enforcement)
Status Enacted

On 8 November 2024, Article 3 of the European Freedom Act (EMFA) came into force, marking the legal recognition of European citizens' right to access diverse and reliable information sources. This adds to the regulation which also sets out a framework on how to combat misinformation and foreign information manipulation and interference (FIMI). EU Member States must implement measures to safeguard editorial independence, protect source confidentiality and foster pluralism in media. After adoption by the EU in April 2024, the new rules of the EMFA regime fully apply by 25 August 2025.

European Union: EU Commission adopts Implementing Regulation concerning the transparency reporting obligations under Digital Services Act (DSA)

Type Regulatory
Status Adopted

On 4 November 2024, the EU Commission adopted an Implementing Regulation which establishes standards on the format, content and reporting timelines for transparency reporting by intermediary service providers under the Digital Services Act (DSA). It mandates that all relevant providers disclose their content moderation practices, including specific details such as on the number of content removals, the effectiveness of automated moderation systems, account terminations and insights into their moderation teams. Very large online platforms (VLOPs) and search engines (VLOSEs) must submit reports biannually, and all other providers must provide annual reports. Starting from 1 July 2025, the providers must collect data according to the standards and first reports are expected in early 2026.

South Korea: Personal Information Protection Committee releases deepfake sexual crime response plan

Type Regulatory (guidance)
Status Complete

On November 6 2024, South Korea's Personal Information Protection Committee (PIPC) released an interagency deepfake sex crime response plan, which strengthens investigative capabilities into deepfake sexual crimes. It also requires online platforms to block and remove harmful content within 24 hours. The response plan was released amidst Korea's push for stricter laws and enforcement against the creation, viewing, and purchasing of sexual deepfakes.

United States: Social media platform X takes legal action against California over Defending Democracy from Deepfake Deception Act of 2024

Type Litigation
Status Ongoing

On 14 November 2024, the social media platform X filed a lawsuit to challenge California's new law regulating political deepfakes on social media under the Defending Democracy from Deepfake Deception Act of 2024 which was passed in September 2024 and is set to be implemented next year. The law mandates that large online platforms, defined as platforms that have at least 1,000,000 Californian users during the previous 12 months, must either prohibit certain election-related deepfake content or label it as "inauthentic" during a specified period around elections. The lawsuit claims that the Act infringes upon First Amendment rights by potentially leading to the censorship of political discourse. X argues that the Act fails to provide clear consequences for misapplication of the regulation and encourages excessive censorship.

United Kingdom: Ofcom publishes open letter on Online Safety Act's (OSA) applicability to Generative AI and chatbots

Type Regulatory (open letter)
Status Published

On 8 November 2024, Ofcom published an open letter addressing online service providers operating in the UK on the application of the Online Safety Act (OSA) to generative AI and chatbots. It highlights recent incidents involving online harm linked to Generative AI, emphasising that platforms which enable user interaction with generative AI, including those allowing users to create their own chatbots, fall under the Act as "user-to-user services". The letter outlines the need for compliance, including conducting risk assessments, implementing effective content moderation, and ensuring robust age verification measures to protect children from harmful content. The letter urges providers to prepare for impending regulatory duties under the OSA, some of which will begin in December 2024, to ensure user safety and mitigate risks associated with generative AI.

United Nations: United Nations (UN) approve draft resolutions on implications of AI in the military

Type Resolution
Status Adopted

On 6 November 2024, the United Nations (UN) First Committee of the General Assembly approved 14 draft resolutions, including implications of AI in the military. The draft emphasises the potential security risks associated with AI in the military,

including risks of an arms race, the possibility of miscalculation, lowering the threshold for conflict, escalation of conflict and proliferation to non-state actors. States are encouraged to take action at all levels to address the opportunities and challenges posed by AI in the military domain from various perspectives, including humanitarian, legal, security, technological, and ethical. The resolution calls for states to cooperate voluntarily, particularly in providing assistance and knowledge-sharing, including exchanging good practices and lessons learned. The Secretary-General is tasked with seeking views from Member States and observer states and submitting a report summarising these views and existing normative proposals for further discussion. The draft resolution regarding AI in the military was approved by a significant majority, with 165 votes in favour, two against (Democratic People's Republic of Korea, Russian Federation), and six abstentions.

Global: G20 Leader's Declaration addresses digital and emerging technologies

Type Voluntary (joint declaration)
Status Published

On 19 November 2024, the G20 leaders published a joint Declaration as a result of the G20 November Summit in Rio de Janeiro. On digital and emerging technologies, the Declaration underscores the need to address digital inequalities and strengthen connectivity and digital inclusion with the aim to empower societies, including women, girls and vulnerable individuals. It stresses that effective digital governance is essential to improving lives while at the same time protecting privacy, personal data, human rights and fundamental freedoms. To tackle the impact of mis- and disinformation, hate speech and other online harms, the Declaration highlights the need for transparency and responsibility for online platforms in line with relevant policies and legal frameworks. To do so effectively, the Declaration commits to work with platforms and relevant stakeholders and facilitate cross-border data flows, adhering to domestic and international legal frameworks.

Section 2 Topic-specific snapshot: "Risk Assessments and Audit Reports"

This section summarises selected analyses and responses published by government agencies, civil society organisations and academia on risk assessment and audit reports.

Assessing Systemic Risk Under the Digital Services Act,
Gabby Miller and Justin Hendrix, 2024

Tech Policy Press hosted a discussion between three experts involved in understanding systemic risk under Europe's Digital Services Act (DSA): Jason Pielemeier, Executive Director of the Global Network Initiative, David Sullivan, Executive Director of the Digital Trust & Safety Partnership, and Chantal Joris, Senior Legal Officer at Article 19. Under the requirements of the DSA, Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) must conduct systemic risk assessments which assess the spread of illegal content on the respective platforms or content that could negatively impact fundamental rights, civil discourse, etc. The participants described the current difficulties from systemic risk assessments under the DSA:

- Lack of transparency on what is happening between the companies and the European Commission. From the civil society side, this has made it extremely difficult to ensure that the risk assessments are meaningful and build upon important existing technology frameworks, namely the UN Guiding Principles on Business and Human Rights.
- Companies have received very little feedback from the European Commission with communication between the two primarily being requests for information and enforcement actions.
- There has been minimal public guidance from the European Commission on what the risk assessments should contain and how they should be conducted.

The participants also detailed the different approaches to risk assessments between the DSA and the UK's Online Safety Act (OSA), both of which are pioneering digital policy legislation. After the DSA came into force, VLOPs and VLOSEs had to comply almost immediately with its provisions so there is minimal guidance on what should be included in systemic risk assessments. On the other hand, Ofcom has held multiple consultations on the OSA and provided thousands of pages of guidance detailing what service providers must include in their risk assessments before any company undergoes a risk assessment of illegal content.

Quick Guide to Illegal Content Risk Assessments,
Ofcom, 2023

In anticipation of the risk assessments required for certain services under the Online Safety Act (OSA), Ofcom, the UK's telecommunications regulatory agency, released its guidance on how platforms can prepare to conduct risk assessments. The assessments primary focus should be on the accessibility of illegal content to users or how the service can be used to commit a criminal offense and its impact. Ofcom outlined four-steps that online services can use while completing a risk assessment:

- Understand the harms – Companies need to identify the illegal harms and risk factors that are applicable to the service.

- Assess the risk of harm – Services need to assess how characteristics of the platform's design can impact the risks of harm. Based on evidence gathered about the respective service, companies need to assess the likelihood of illegal content on the service and its impact.
- Decide measures, implement, and record – Services must address the identified risks through mitigation efforts and record the results of the risk assessment.
- Report, review, and update risk assessments – Ofcom recommends that services report their risk assessment outcomes to the relevant internal governance body or associate. Additionally, it's recommended that risk assessments are reviewing annually to ensure that they are kept up to date.

Identifying and Assessing Human Rights Risks related to End-Use,

United Nations Human Rights Office of the High Commissioner (OHCHR), 2020

The United Nations Human Rights Office has developed guidance for technology companies as part of its B-Tech project on business and human rights, aimed at addressing human rights risks associated with the products and services of these companies, in alignment with the UN Guiding Principles on Business and Human Rights (UNGPs). It lays out how to identify and assess human rights risks, considering potential impacts across all aspects of a business. This includes the design, development, promotion, sales, and usage of their services. Companies are encouraged to carry out thorough evaluations and continuously engage with stakeholders to ensure a deeper understanding of the human rights landscape related to their operations. They are also advised to prioritise risks that could have the most significant consequences, based on three critical factors:

1. Scale: Refers to the severity of the potential impact;
2. Scope: Encompasses the breadth of the possible effects, including how many individuals or communities could be affected by the risk;
3. Remediability: Refers to the ability to reinstate affected individuals or groups to their condition before the impact.

By focusing on these aspects, companies should be able to effectively manage risks to human rights and mitigate any violations of such. Additionally, the importance of a human rights due diligence process is also highlighted, which begins with the identification and assessment of actual or potential human rights impacts arising from the company's activities, business relationships, or service usage. When prioritising responses to human rights impacts, companies must pay particular attention to groups at greater risk of marginalisation or vulnerability, including children, women, indigenous communities, ethnic minorities, members of the LGBTI community, and human rights defenders. Furthermore, it is highlighted that to achieve comprehensive protection, it is necessary that this is an ongoing process, including regular reassessments of risks and redefining approaches - particularly, as information, user behaviour and technology changes. The guidance is designed to embed human rights considerations into core business processes and foster a culture of accountability within the technology industry.

AI Transparency in practice: What was learnt from third-party audit of recommender systems at LinkedIn and Dailymotion,

Jiahao Chen, Jack Bandy, Dave Buckley and Ruchi Bhatia, Christchurch Call, 2024

This report outlines the findings from the first phase of the Christchurch Call Initiative on Algorithmic Outcomes (CCIAO), during which OpenMined's PySyft software was employed to enable external access to impression data associated

with the production recommender systems at LinkedIn and Dailymotion. It also details the research conducted via this platform by four independent researchers. They were able to perform quantitative analyses of the recommender systems on both platforms, investigating how these algorithms influence the content recommended to users, all while safeguarding the security and privacy of personal or commercially sensitive information through the application of PySyft and the OpenDP differential privacy library. Through their research, the researchers derived these key insights:

- Third-party auditors can successfully run analyses against the recommender systems and determine insights into recommendation behaviour across different demographics and how content recommendation is impacted by various algorithms
- There is significant difficulty in establishing a baseline for comparison, but baselines are necessary for understanding the algorithm's impact
- External audits of a specific system are only effective when the auditor and system owner maintain transparency and work closely with the auditors

---

Towards meaningful fundamental rights impact assessment under the DSA,
Access Now and the European Center for Not-for-Profit Law, 2023

This report highlights the importance for companies to recognise, assess, and mitigate both actual and potential risks to fundamental rights linked to their services by conducting Fundamental Rights Impact Assessments (FRIAs) under the Digital Services Act (DSA). It seeks to aid Very Large Online Platforms' (VLOPs) and Very Large Online Search Engines' (VLOSEs) in executing FRIAs effectively, emphasising the need for comprehensive engagement rather than just fulfilling compliance requirements.

The authors advocate for a thorough evaluation that identifies systemic risks associated with automated content moderation, with a particular focus on safeguarding freedom of expression and access to information. They propose that FRIAs should be aligned with the EU Charter of Fundamental Rights, promoting transparency and ensuring engagement with various stakeholders throughout the process. To enhance the effectiveness of these assessments, the report recommends the establishment of clear guidelines for identifying negative impacts on fundamental rights. This involves broadening the scope of assessments to include all relevant service components and considering the implications for both direct users and affected communities. The authors highlight the necessity for regular consultations with external stakeholders, particularly civil society organisations, to ensure that the assessments reflect diverse perspectives. Ultimately, the report calls for a harmonised approach across the EU to ensure consistent, high-quality impact assessments that can effectively address risks to fundamental rights in the digital environment.

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.