Digital Policy Lab

Policy Digest #14

12 February 2024

ISD Powering new generations against extremism

ALFRED LANDECKER FOUNDATION

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 Commissioner releases Online Safety Act industry standards drafts

Type Regulatory (consultation)
Status Published

On 20 November 2023, the Australian eSafety Commissioner released drafts of two new industry standards for consultation. The standards are intended to support the 2021 Online Safety Act and introduce mandatory compliance measures to address the most harmful 'Class 1' content. The first standard covers relevant electronic services (RES), a broad category of communication services that includes email, instant messaging, SMS, multimedia message services, dating services, and services that enable people to play games together. The second standard covers designated internet services (DIS), which includes online file storage websites and a variety of apps and websites not covered under the RES draft or other existing standards.

The standards will operate alongside industry-drafted codes covering other services. Previous codes on RES and DIS were initially drafted by industry but "did not meet the statutory requirements for registration because they did not contain appropriate community safeguards for users in Australia," requiring eSafety to decline registration of the codes in May 2023. This process resulted in eSafety drafting the current proposed codes. The consultation closed on 21 December 2023. In January and February 2024, the consultation responses will be reviewed and considered in possible amendments to the standards. A Lodgement of Standards and explanatory statements will then be brought to the Office of Parliamentary Counsel for review in March 2024.

Australia: Government publishes consultation on amended Online Safety (Basic Online Safety Expectations) Determination

Type Regulatory (consultation)
Status Published

On 22 November, the Australian federal government launched a consultation on proposed amendments to the Online Safety (Basic Online Safety Expectations) Determination 2022 (BOSE Determination). The BOSE Determination is part of the 2021 Online Safety Act and sets the government's minimum safety expectations regarding online service providers. Proposed amendments to the Determination are intended to reflect the rapidly evolving online landscape, and cover generative AI, recommender systems and user controls; children's wellbeing and restricting access to age-inappropriate content; safety impacts of business and resourcing decisions; hate speech online; transparency; and enforcement of terms of use.

These proposed amendments, coupled with the recently published draft industry standards, point towards an increased specificity in the Australian online safety framework. The BOSE Determination consultation will remain open until 16 February 2024.

[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 Commission launches infringement proceedings against X

Type Regulatory (enforcement)
Status Investigation proceedings commenced

On 18 December 2023, the Commission opened formal proceedings against X, previously known as Twitter, under the Digital Services Act (DSA). The proceedings will focus on X's compliance with the DSA regarding the dissemination of illegal content in the EU; the efficacy of measures to combat information manipulation and risks to civic discourse and electoral processes; measures to increase transparency and researcher data access; and suspected deceptive design issues. These are the first formal proceedings launched under the DSA and empower the Commission to take further enforcement steps, such as interim measures and noncompliance decisions. There is no deadline for the conclusion of formal proceedings.

European Union: Political agreement reached on the European Media Freedom Act

Type Regulatory (agreement)
Status Finalised, subject to formal approval

On 15 December 2023, the European Commission, Parliament and Council came to a political agreement on the European Media Freedom Act (EMFA). The EMFA will facilitate a free and independent media landscape on- and off-line and includes the creation of a new and independent European Board for Media Services. It also includes safeguards against the unwarranted removal of media content by Very Large Online Platforms (VLOPs), and protects editorial independence, independence of public media, and the transparency of media ownership and audience measurement methods. The agreement now must receive formal approval from the European Parliament and Council. Once adopted and published in the Official Journal of the European Union, the EMFA will be binding and applicable in all Member States after 15 months. It is intended to build on the DSA and the Audiovisual Media Services Directive (AVMSD).

European Union: European Commission sends formal requests for information to Very Large Online Platforms (VLOPS) and Very Large Online Search Engines (VLOSEs)

Type Regulatory (investigation)
Status Processing information received

On 14 December 2023, the Commission sent formal requests for information to Apple and Google under the DSA. Both companies have been requested to provide more information on systemic risks they have identified regarding the App Store and Google Play. The Commission also is seeking more information on Apple and Google's compliance with rules on online marketplaces and recommender systems, and online advertising transparency.

On 18 January 2024, 17 VLOPs and VLOSEs also received formal requests for information from the Commission. The selected VLOPs and VLOSEs must provide further information on the measures they have taken to comply with obligations under the DSA to provide eligible researchers with access to publicly accessible data via their online interfaces. Researcher access to publicly available data is an important contribution to the DSA's goals, with the Commission citing access as being particularly important in the lead-up to national and bloc-wide elections within the EU this year.

European Union: European Commission designates new VLOPs under the DSA

Type Regulatory (designation)
Status Complete

On 20 December 2023, the Commission designated a second set of VLOPs under the DSA. The new designated services are Pornhub, Stripchat and XVideos; all three services fulfil the VLOP designation threshold of 45 million average monthly users in the EU. Their designation will subject them to higher scrutiny and accountability from the Commission and national regulators.

European Union: Artificial Intelligence Act unanimously voted in by Committee of Permanent Representatives

Type Regulatory (agreement)
Status Awaiting formal adoption

On 2 February 2024, the European Union Committee of Permanent Representatives, composed of representatives of all member states, voted unanimously in favour of the Artificial Intelligence Act. The vote marks a significant step in the progress of the Act, which faced recent opposition from some member states over concerns about its effects on industry. Once formally adopted by the European Parliament's Internal Market and Civil Liberties Committees, the Act will become the world's first piece of comprehensive AI legislation.

The final agreement sets out tiered categories of risk for AI systems, which will be subject to increasingly stringent requirements design to protect the health, fundamental rights, and safety of people in the EU. General Purpose AI (GPAI) models will be subject to varied levels of stringency, depending on whether they pose systemic risks. The Act also compromises on placing biometric categorisation and identification within its risk framework, leaving many exceptions to the prohibited category while maintaining standards for preventing abuses of the technology.

European Union: Commission launches public consultation on draft Digital Services Act electoral integrity guidelines

Type Regulatory (consultation)
Status Published

On 8 February 2024, the European Commission opened their draft guidelines on electoral integrity under the DSA to public consultation. The guidelines are the first under Article 35 of the DSA, which covers codes of conduct to address specific risks. The guidelines are intended to provide VLOPs and VLOSEs with guidance on best practices and measures to mitigate systemic risks to electoral integrity online. This includes mitigation measures regarding general election-related risks, generative AI content, and the upcoming European Parliament elections from 6-9 June 2024.

The draft guidelines build on the dialogues between the Commission and several of the first 19 designated services under the DSA. The consultation is open until 7 March 2024.

Germany: Government introduces the DSA implementing law "Digitale-Dienste-Gesetz"

Type Regulatory (Implementing law)
Status In Parliament

On 15 January 2024, the German government introduced the law implementing the EU's DSA, the "Digitale-Dienste-Gesetz (DDG)". The DSA establishes a horizontal legal framework for digital intermediary services, standardises rules for consumer protection, and a robust supervisory structure. The Federal Network Agency ("Bundesnetzagentur") is designated as the Digital Services Coordinator (DSC) in Germany. Its responsibility will be to monitor compliance with the DSA. Providers are obligated to take measures against illegal content, and users will be able to report any non-compliance with these obligations to the Federal Network Agency.

Ireland: Coimisiún na Meán launches consultation on the draft Online Safety Code and designates Video-Sharing Platform Services

Type Regulatory (consultation)
Status Published

On 18 December 2023, the Irish media regulator (CNaM, "Coimisiún na Meán") published their draft Online Safety Code for video-sharing platform services. The draft Code sets out measures to improve user safety online, with a focus on children's safety; it covers illegal content as well as cyberbullying, promotion of eating disorders, self-harm, and suicide. It forms part of the larger Irish online safety framework, which includes its domestic Online Safety and Media Regulation Act, as well as the EU DSA and Terrorist Content Online Regulation. Once enacted, the Code will be legally binding for all designated Irish-based video-sharing platforms. The consultation closed on 19 January 2024.

On 9 January 2024, the CNaM published details about which video-sharing platforms will fall under the remit of the Code. The designated services are Facebook, Instagram, Udemy, YouTube, TikTok, LinkedIn, X, Pinterest, Tumblr, and Reddit. Breaches of the Code will leave platforms liable for fines of up to €20 million. Reddit responded to their designation by launching legal action on 16 January 2024 at the Irish High Court, although no date for a hearing has been set yet.

Netherlands: Authority for Consumers and Markets publishes DSA implementation consultation and signs support agreement with the European Commission

Type Regulatory (consultation)
Status Published

On 18 January 2024, the Dutch Authority for Consumers and Markets (ACM, "Autoriteit Consument & Markt") published its draft guidelines for market participants that fall under their remit in enforcement of the DSA. The ACM is the regulator-designate for the Netherlands. Their guidance is formulated to address the compliance requirements of different types of providers under the DSA. Topics covered include dispute settlement, illegal content, and protecting minors. The guidance has been issued in advance to allow businesses to prepare for the incorporation of the DSA into Dutch law, which is expected to come later this year. Once the law has been nationally implemented, the ACM will be able to enforce the DSA in the Netherlands.

On 19 December 2023, the ACM also signed an administrative arrangement with the European Commission, in preparation for the Commission's supervision of VLOPs and VLOSEs under the DSA. The arrangement will support the Commission in identifying and assessing systemic risks, as well as its investigations of services. It is also expected to improve flows of information, data, best practice, methods, and tools between the Commission and the ACM. Similar arrangements have already been signed between the Commission and the French, Irish, and Italian regulators.

United Kingdom: Data Protection and Digital Information Bill passes second reading

Type Regulatory (draft)
Status Awaiting committee reading

On 19 December, the Data Protection and Digital Information Bill passed its second reading in the House of Lords. The Bill would establish an updated and simplified data protection framework in the UK through a range of measures. This includes a framework for digital identity verification services; greater clarity on data protection rules, especially for researchers; reforms to the Information Commissioner; and provisions on the flow of personal data for law enforcement and national security purposes. Various amendments have been proposed by members of the House of Lords to the Bill, including one on access to data for vetted researchers that is closely modelled on Article 40 of the EU DSA and aims to ensure "UK-based academic researchers are not put at a disadvantage when it comes to researching matters of public interest regarding whether the largest online services - including services most used by children - are safe, private and comply with UK law." The next step for the Bill is to go through the Committee stage in the House of Lords; a date for this has not yet been announced.

United States: States consider and enact laws on AI in elections

Type Regulatory (state-level)
Status Variable state-by-state

In January and February 2024, at least 40 US states introduced or took action on bills to regulate the use of AI in elections and political campaigning. Many of these efforts have bipartisan backing. Five states (California, Michigan, Minnesota, Texas, and Washington) have already passed laws, with bipartisan backing, either prohibiting or placing requirements on certain relevant use cases. When not banned, disclosure that content is synthetically generated is often required instead. Washington State's legislation, for example, requires that any synthetic media (image, video, or audio) used in an electoral context must be clearly labelled via audio or text. As federal legislation on the use of AI in elections stalls, the role of state-level legislation has the potential to play an important role in the 2024 elections.

United States: Senate Judiciary Committee conducts hearing on online child sexual exploitation

Type Hearing
Status Complete

On 31 January 2024, the Senate Judiciary conducted a hearing on online child sexual exploitation, featuring the CEOs of Meta, X, TikTok, Snap, and Discord. Several Judiciary members specifically mentioned that there is currently no legal remedy for victims and their families who have been harmed by content or actions that have happened on the platforms, proposing that individuals be able to sue the platforms directly for these harms. Many Judiciary members also referenced either amending or repealing the Section 230 protections for the platforms within the Communications Decency Act of 1996, which shield them from taking liability for content posted on their platforms. Senator Lindsay Graham said he would be introducing a bill to repeal Section 230 completely over the coming weeks. The CEOs were asked about their willingness to endorse several relevant bills, with X CEO Linda Yaccarino endorsing both the STOP CSAM Act and Kids Online Safety Act (KOSA) during the hearing, after endorsing the SHIELD Act the day before.

United States: Biden-Harris Administration announce creation of the U.S. AI Safety Institute Consortium

Type Multistakeholder forum
Status Announced

On 8 February 2024, the US Secretary of Commerce Gina Raimondo announced the creation of the US AI Safety Institute Consortium (AISIC). AISIC will bring together an initial group of over 200 leading stakeholders in the AI field, including industry developers and researchers, academics, government, and civil society organisations. Stakeholders will contribute towards key directives from President Biden's October 2023 Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. These include developing red-teaming guidelines, capability evaluations, risk management, safety and security, and watermarking synthetic content.

Section 2 Topic-specific snapshot: "Electoral Safeguarding: Assessing and Mitigating Evolving Risks Posed to Electoral Integrity Online"

This section provides an overview of how elections are covered by two key pieces of online safety legislation: the EU Digital Services Act and the UK Online Safety Act. It also summarises private sector announcements regarding their plans for elections in 2024, as well as selected analyses and recommendations published by government agencies, civil society organisations and academia on the topic of electoral safeguarding.

Regulatory Refresher: EU and UK Online Safety Legislation and Safeguarding Electoral Integrity

The EU Digital Services Act includes various provisions with regards to electoral integrity through its regulation of systemic risks and targeted advertising via online services. All online services are required to:

- Not profile users with online advertising based on protected data. This ban includes profiling based on categories such as ethnicity, political views, or religion.
- Prevent the dissemination of illegal content, including terrorist content or illegal hate speech.

Online services of all sizes are also encouraged by the European Commission to sign on to the 2022 Code of Practice on Disinformation, a set of self-regulatory standards designed to fight disinformation. The Code of Practice is set up to function as a mitigation measure and recognised Code of Conduct under the DSA, and forms part of the larger framework of EU digital regulation. Signatories include Meta, Microsoft, and TikTok.

However, the highest level of obligations fall on designated Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) – the largest companies within the bloc in terms of user numbers – who are subject to several additional obligations. Collectively these obligations aim to mitigate risks stemming from legal but harmful online content or activity that can influence electoral processes, such as dis- and misinformation. They require VLOPs and VLOSEs to:

- Assess systemic risks posed by their services. Negative impacts on democratic processes, civil discourse and electoral processes are considered systemic risks under Article 34. All VLOPs and VLOSEs must identify, analyse, and effectively mitigate such risks, as well as risks from the potential misuse of services by recipients of the service. Systemic risk assessments from services must include the systems and elements that contribute to the risk, including algorithmic systems such as those used for recommendation and advertising.
- Have a crisis response mechanism, which should include specific measures to take when their service is used for the rapid spread of disinformation.
- Maintain a repository of public advertisements, which allow researchers to better study emerging risks to political participation and civil discourse, as well as from disinformation campaigns.

The UK Online Safety Act (OSA), passed in October 2023, primarily focuses on illegal online content and activity and harms to children. While it does not include as extensive disinformation or elections-related provisions as the EU's DSA, it does include new foreign interference criminal offences created by the new National Security Act (2023) as a 'priority offence' under its illegal content duties. These new offences include illegitimate activities that misrepresent a person's identity or participation

in political processes, including elections and proceedings of registered political parties. As a result, the OSA requires online service providers to assess the risk of foreign interference activities occurring on their service. Providers are obliged to put in place proportionate measures to effectively mitigate and manage any risks of foreign interference, including those relating to elections and other democratic processes. The UK communications regulator, Ofcom, expects to begin enforcement of illegal harms codes by early 2025, and all codes by early 2026.

Platform Approaches to 2024 Elections

Several technology companies have already published their approaches to 2024 global and national elections. The following list, current at the time of writing, summarises these approaches:

- On 28 November 2023, Meta announced its approach to upcoming elections. While elements of its approach remain "consistent for some time," other policies are new: advertisers are now required to disclose the use of AI or other digital techniques to edit ads containing realistic photo, video, or audio content. Meta is also expanding its threat detection capacity and policies to address threats against election officials and poll workers.
- On 19 December 2023, Google outlined its policies regarding the November 2024 US elections. Its new tools and policies include watermarking, disclosure, and labelling requirements regarding generative AI; efforts to help users access high-quality information via Maps, YouTube, News, and Search; and partnerships with security organisations on campaign and information security risks.
- On 15 January 2024, Open AI indicated its approach to 2024 elections worldwide. It outlines its usage policies to prevent abuse of tools such as ChatGPT, including restrictions on chatbots pretending to be real people, or misrepresentation of democratic processes and/or discouragement of political participation. It is also working to implement the Coalition for Content Provenance and Authenticity's digital credentials into images generated by DALL-E 3, which is non-removable and will identify images as AI-generated. In the United States, ChatGPT will also direct users to the authoritative and nonpartisan CanIVote.org for voting information.
- On 18 January 2024, TikTok outlined its efforts to protect election integrity worldwide. Information quality efforts include media literacy partnerships, and, in the US, an Elections Centre for users to access reliable voter information, as well as requirements for political figures to have verified accounts. It also intends to continue moderating content and accounts by disrupting and removing influence operations, restricting misleading AI-generated content, and countering misinformation both proactively and reactively.

Further Reading: Election-Related Analyses and Recommendations

Foreign Information Manipulation and Interference Threats: A Framework for Networked Defence,
European Union External Action Service (EEAS), January 2024

This report describes the current FIMI threat landscape and provides a comprehensive response framework based on 750 FIMI incidents investigated by the EEAS. It also provides detailed insights into FIMI operations during electoral processes, including cross-case patterns and expected threat progression in the lead-up to an election. The report also offers a brief but concise framework to respond to election-specific FIMI threats. The framework identifies key actions for government, political parties, platforms, and civil society to take throughout the elections process, starting months before the election and extending post-election.

Democracy by Design: A Content-Agnostic Election Integrity Framework for Online Platforms,
Accountable Tech and partners, September 2023

This framework provides a set of content-agnostic election integrity recommendations to online platforms, designed to maximise impact. All recommendations are designed to be easily actionable, and are developed based on platforms' existing product design, policy toolkits, and – where possible – research. The recommendations are intended to resonate across a variety of national cultures and legislative systems, as well as distinct types of platforms.

There are three categories of recommendations:

- Bolstering Resilience: 'soft interventions' to introduce targeted friction and context to platform design to mitigate harms to civic discourse and democracy. The framework particularly recommends using interstitials for misleading information, 'circuit breakers' for mitigating harms from viral content, restricting resharing, and implementing strike systems for repeat offenders.
- Countering Election Manipulation: preventative measures against evolving threats posed by hostile actors and automation. Recommendations focus on safeguards such as content provenance standards, transparency on and user adjustability of the parameters of recommendation systems, and prohibitions on the use of generative AI to share electoral misinformation or create micro-targeted ads.
- Paper Trails: key transparency measures required to adequately assess systemic threats, intervention efficacy, and trust building. The framework recommends that platforms make all election-related policies available in one clear and accessible place, release regular transparency reports across prominent languages, and provide researcher data access to support impact and threat analysis.

Terrorism, Extremism, Disinformation and Artificial Intelligence: A Primer for Policy Practitioners,
Institute for Strategic Dialogue, 22 January 2024

This policy briefing outlines the ways in which artificial intelligence technologies may interact with democracy, with the focus on extremism, harmful online content, misinformation, and disinformation. It provides a brief conceptual primer on AI, including high level explanations of relevant technical processes and the current landscape of AI opportunities and risks.

The briefing analyses the risks posed by text, image, video, and audio-based content generation, including politically motivated 'deepfakes' and the scaled production of persuasive and deceptive content online. It also covers the risks posed by AI-augmented political marketing and bot nets, as well as issues posed by recommender systems. This risk landscape overview is accompanied by corresponding public policy solutions and technical mitigation measures, including suggestions for further reading.

For practitioners looking specifically for an analysis of risks and public policy solutions regarding information ecosystems and democratic processes, we recommend reading the beginning of Section 2 (pp. 15-22) and Section 3 (pp. 27-31).

Diaspora Communities and Computational Propaganda on Messaging Apps,
Centre for International Governance Innovation, January 2024

This policy brief covers research on the use of computational propaganda – including misinformation, disinformation and foreign information manipulation and interference (FIMI) – to politically influence and target diaspora minorities in the United States. It focuses on messaging apps, which are particularly important amongst diaspora communities and often lack effective content moderation. Research findings indicate that minorities are more frequently targeted via messaging apps, and that targeting occurs based on different narratives, and often in different languages, than material directed at the American majority population. These differences mean that propaganda targeting minorities and members of diaspora communities is often missed by policymakers and analysts.

Researchers identify four types of false and misleading information commonly shared on messaging apps by diaspora communities: "sowing of confusion via translational ambiguities; leveraging falsehoods to redraw ideological fault lines; use of religion to sow doubt about candidates' views; and oversimplification of complex perspectives, policies and procedures to alter voting decisions." These techniques also leverage the political histories in countries of origin, drawing on factors such as pre-existing distrust of government to support mis- and disinformation. The brief also provides policy recommendations, focusing on the role of inclusive media literacy programming and minority and diaspora voices in the regulation of messaging apps.

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.