

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 publishes summary of transparency notices under the Online Safety Act 2021

Type Regulatory

Status Non-periodic notices summary published

In October 2023, Australia's eSafety Commissioner [published key findings](#) of the responses to the non-periodic notices issued to Google, Twitter (X), TikTok, Twitch and Discord on 22 February 2023 regarding their steps to detect and address child sexual exploitation and abuse (CSEA) on their services, including sexual extortion and the safety of recommender systems. The summary includes key findings regarding the detection of livestreamed CSEA; the detection of grooming; blocking of URLs linking to known CSEA material; community moderation; language coverage; user reporting; and the scale of abuse. A full report of the industry responses [can be found here](#).

EU: Delegated Regulation on independent audits under the Digital Services Act

Type Regulatory

Status Delegated Regulation published

On 20 October 2023, the European Commission [published a Delegated Regulation](#) on independent audits under the Digital Services Act (DSA). This delegated act provides a framework to guide providers of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) as well as auditing organisations in the preparation and issuance of audit reports and audit implementation reports. The publication of audit reports and audit implementation reports should increase the transparency and accountability of providers and offer a comparative basis for public scrutiny. The audit reports should also be a source of information for the European Commission, the national Digital Services Coordinators (DSCs), as well as other competent authorities.

EU: Administrative arrangements with French, Irish and Italian media regulators to support enforcement of the Digital Services Act

Type Regulatory

Status Enforcement

On 23 October 2023, the European Commission services [signed administrative arrangements](#) with the media regulators of France (Autorité de régulation de la communication audiovisuelle et numérique, Arcom) and Ireland (Coimisiún na

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.

Meán) to support its supervisory and enforcement powers under the DSA. On 30 October 2023, the Commission signed an arrangement with the Italian media regulator (Autorità per le garanzie nelle comunicazioni, AGCOM). These bilateral arrangements are aimed at developing expertise and capabilities for coordinating their response to the spread and amplification of illegal content on VLOPs and VLOSEs. They allow the Commission services and relevant national authorities to exchange information, data, good practices, methodologies, technical systems and tools. The arrangements will be of particular importance until the establishment of the Board of Digital Services Coordinators, which is due for February 2024. Once the Board is operational, the arrangements will continue to provide added value to organise the practical relationships between Commission services and national authorities under the DSA.

EU: Regulation on the transparency and targeting of political advertising

Type Regulatory

Status Provisional agreement

On 7 November 2023, the Council of the EU and the European Parliament reached a provisional agreement on the main political elements of a new regulation on the transparency and targeting of political advertising. Under the provisional agreement, political advertising is defined as the “preparation, placement, promotion, publication, delivery or dissemination of messages: by, for, or on behalf of political actors, unless they are of a purely private or a purely commercial nature; or which are liable and designed to influence voting behaviour or the outcome of an election, referendum, or a legislative or regulatory process at EU, national, regional or local level.” The provisionally agreed rules also place strict limits on the use of targeting and ad delivery techniques. In addition, they contain a blanket ban on profiling using special categories of personal data, such as data revealing racial or ethnic origin or political opinions. To prevent foreign interference, co-legislators also agreed to prohibit the provision of advertising services to third country sponsors three months before an election or a referendum. Furthermore, the European Commission should set up a European public repository for online political advertisements to make publicly available information on all online advertisements and their transparency notices.

The Council and the Parliament will continue to work on a technical level to finalise the text of the regulation. The full agreement will need to be adopted by both institutions in their entirety and undergo legal-linguistic revision before formal adoption.

France: Christchurch Call Leaders’ Summit convenes in Paris

Type Non-regulatory

Status Issued

On 10 November 2023, New Zealand and France co-hosted the 2023 Christchurch Call Leaders’ Summit in Paris, resulting in the publication of a joint statement. The Summit brought together government, technology companies, partner organisations, and civil society to cover a range of issues at the intersection of technology and extremism. It took place in the context of the conflict between Israel and Hamas, and rising volumes of terrorist and violent extremist content, and content that glorifies or promotes violence, antisemitism, and Islamophobia. On crisis response capability, Leaders welcomed progress to bring more online service providers into the Call and endorsed actions to update the Call’s Shared Crisis Response Protocol. A key challenge identified by the Call was the rapid advances in technology, including foundation AI models and how these create vulnerabilities that can be exploited by terrorists and violent extremists. Other issues

covered included the challenge of radicalisation to violence and the impacts of recommendation systems, including the development of new governance structures and ethics frameworks, and multistakeholder approaches to delivering on the Christchurch Call commitments. Leaders welcomed the report on [Misogynistic Pathways to Radicalisation](#) prepared by ISD within the context of the DPL, in association with the Christchurch Call and the Global Partnership for Action on Gender-Based Online Harassment and Abuse. The Summit also welcomed a number of new signatories with highly relevant expertise on the above topics, including technology companies like OpenAI, Anthropic, and Discord, as well as partner organisations such as CASM Technologies and The Extremism and Gaming Research Network (EGRN).

South Africa: Competition Commission launches Media and Digital Platforms Market Inquiry

Type Regulatory

Status Inquiry established

On 17 October 2023, South Africa's Competition Commission [launched the Media and Digital Platforms Market Inquiry \(MDPMI\)](#), designed to scrutinise the distribution of media content on South African digital platforms and the advertising technology (adtech) markets. Commissioner Doris Tshepe noted that the inquiry would delve into various issues facing media sustainability, stating that it "comes at a critical moment for the media industry as news consumption rapidly shifts online and traditional sources of funding to print and broadcasting advertising decline." The MDPMI will also assess how digital platforms and adtech market dynamics affect the visibility and competitiveness of smaller news organisations, including community and African language news media, and the diversity of news available to consumers. It will concentrate on digital platforms, including search engines, social media sites, video-sharing platforms, and news aggregation platforms, as well as adtech market participants on the supply and demand sides, and ad exchanges. During the initial phase, the MDPMI plans to conduct two rounds of information gathering, with public hearings scheduled for March 2024. The Terms of Reference, Statement of Issues, Administrative Timetable, and Guidelines for Participation are all accessible on the [MDPMI website](#).

UK: Online Safety Act receives Royal Assent and Ofcom publishes first consultation on illegal harms

Type Regulatory

Status Entering into effect

On 26 October 2023, the Online Safety Act [received](#) Royal Assent. The Act gives Ofcom the authority to enforce the new requirements, including regulatory supervision of internet services that enable the sharing of user-generated content ('user-to-user services') and search services. Ofcom will be able to issue provisions to implement the regulation and will be required to issue codes of practice on the existing duties. An [Implementation Roadmap](#) for the Act includes three phases:

- Phase one (Illegal content): On 9 November 2023, Ofcom [published its first consultation on illegal harms](#), including child sexual abuse material, terrorist content and fraud. It covers the causes and impacts of illegal harms; how services should assess and mitigate the risks of illegal harms; how services can identify illegal content; and the approach to enforcement. Ofcom also published an '[at a glance](#)' summary of the proposals and who they apply to; a [summary of each chapter](#) of the consultation; as well as first drafts of risk assessment, codes of practice and associated guidance, that together underpin the illegal harms part of the regulatory regime.

- Phase two (Child safety, pornography, and protecting women and girls): A consultation, due in December 2023, will set out draft guidance for services that host pornographic content. Further consultations relating to child safety duties will follow in Spring 2024, while Ofcom will publish draft guidance on protecting women and girls by Spring 2025.
- Phase three (Additional duties for categorised services): These duties, including the publication of transparency reports and development of user empowerment measures, apply to services which meet certain criteria related to their number of users or high-risk features of their service. Ofcom will publish advice to the Secretary of State regarding categorisation, and draft guidance on its approach to transparency reporting, in Spring 2024.

UK: Artificial Intelligence Safety Summit leads to signing of Bletchley Declaration, announcement of UK AI Safety Institute, and 'State of the Science' Report

Type Non-regulatory

Status Completed

On 1-2 November 2023, the UK government convened government leaders and ministers, as well as select members of industry, civil society, and academia, for the world's first AI Safety Summit at Bletchley Park, UK. 27 countries and the European Union agreed to the [Bletchley Declaration](#), an agreement to support a network of research on AI safety, as well as commitments to securing human rights and the United Nations Sustainable Development Goals. While non-binding, the Declaration is significant for being the first global AI governance tool that brings together democratic and non-democratic countries, from every continent, in common agreement on the benefits and risks of AI.

The UK government announced that Yoshua Bengio, a leading academic in the field, has been commissioned to lead a global team of experts on drafting a '[State of the Science](#)' Report on AI. The Report will summarise existing research on AI risks and identify key research priorities moving forward. It is to be published before the [next AI Safety Summit](#), to be held by France in November 2024. The Republic of Korea will co-host an additional miniature virtual summit in six months' time. The [United States](#) and [United Kingdom](#) also announced the launch of respective national AI Safety Institutes, which will evaluate advanced AI systems, conduct AI safety research, and facilitate information exchange with national and international stakeholders.

US: Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

Type Presidential action

Status Issued

On 30 October 2023, US President Biden issued a sweeping [Executive Order](#) covering the impacts and governance of AI across the areas of safety, privacy, civil rights, education, healthcare, government, workplaces, innovation and competition, and global cooperation. A particularly notable element of the Executive Order is that developers of particularly advanced AI systems will be required to notify the federal government when training the model, as well as share their safety test results and other critical information with the US government. The National Institute of Standards and Technology (NIST) has also been directed to develop AI safety standards, which will be adopted and applied by the Department of Homeland Security to critical infrastructure sectors. The Executive Order also urges action in other branches of government. It calls on Congress to pass bipartisan legislation on data privacy, with special attention to child protection. It additionally urges the Federal Trade Commission (FTC) to assess its regulatory scope and exercise existing authority where appropriate.

While the Executive Order is not a comprehensive legal framework – and therefore lacks the detail of legislation – it nevertheless has significant implications for standard-setting and outlines further priorities of action by the US government. By setting standards for the AI systems that the federal government will buy, the US may also be able to influence what the market considers as acceptable safety standards.

G7: Hiroshima Artificial Intelligence Process publishes its Comprehensive Policy Framework

Type Non-regulatory
Status Published

On 30 October 2023, as part of the ongoing Hiroshima Artificial Intelligence (AI) Process, the [G7](#) launched two living documents to guide the responsible development of advanced AI systems. The [Hiroshima Process International Guiding Principles for Advanced AI Systems](#) outlines 11 key principles to the safe and trustworthy development of advanced AI systems. These principles build on [existing OECD AI Principles](#), and are intended to assist organisations with the development, deployment, and use of advanced AI systems. The [International Code of Conduct for Organizations Developing Advanced AI Systems](#) provides more specific guidance to developers on how to implement the principles in practice. The G7 called organisations developing advanced AI systems to commit to the application of the Code of Conduct.

Section 2 Topic-specific snapshot: “TVEC, Hate Speech and Foreign Interference in the Context of the Israel-Hamas Conflict: Risks, and Regulatory and Platform Responses”

This section summarises selected analyses and responses published by regulators, academia, and civil society organisations on the topic of TVEC, hate speech and disinformation in the context of the Israel-Hamas conflict.

Rise in antisemitism on both mainstream and fringe social media platforms following Hamas’ terrorist attack,

Institute for Strategic Dialogue (ISD), 31 October 2023

Using a bespoke hate speech classifier, ISD analysis found an over 50-fold increase in the absolute volume of antisemitic comments on YouTube videos about the Israel-Hamas conflict, following Hamas’ attacks. The data demonstrated an increase in the overall proportion of antisemitic messages during the same period by a factor of 2.4. A total of 15,720 antisemitic YouTube comments were identified after the attack. This increase coincided with a peak in overall volume of collected videos, as well as Israel’s response, including the mobilisation of troops and the targeting of sites in Gaza. Among the significant volume of identified antisemitic comments, several broad themes aligned with the International Holocaust Remembrance Alliance’s (IHRA) working definition of antisemitism. This included dehumanising language, often used with reference to Jewish people, including slurs or comparing Israelis or Zionists to Nazis, and their actions to those of Hitler. Cross-platform monitoring showed increased online threats against Jewish communities, as well as a three-fold increase in the use of antisemitic language on alternative platforms, including 4chan, Gab, Telegram and BitChute.

Use of words, phrases and hashtags associated with anti-Muslim mobilisation surges amid Israel-Gaza conflict,

Institute for Strategic Dialogue (ISD), 2 November 2023

This Dispatch explores the prevalence of keywords associated with anti-Muslim hate speech on X; hashtags associated with campaigns promoting negative anti-Muslim stereotypes; and the prevalence of anti-Muslim hate on alt-tech platforms. On X, the number of posts containing anti-Muslim keywords spiked during the weekend of Hamas’ terrorist attacks on 7-8 October, resulting in a 422% increase from the previous two days. The entire week of Hamas’ terrorist attacks (7-13 October) saw an increase of 250% in anti-Muslim hate compared to the previous week. Beyond mis- and disinformation, the use of anti-Muslim keywords served to stereotype and essentialise Muslim communities as violent, attributing Hamas’ actions to Muslims in Western countries. An initial examination of anti-Muslim hashtags used on X similarly shows a significant increase in online anti-Muslim hate following Hamas’ terrorist attacks. In the 10 days immediately following the attack, use of hashtags #fuckislam, #stopislam and #banislam rose 132% on the previous period. Utilising the same dehumanising anti-Muslim keywords, ISD researchers investigated content on alternative platforms, including Patriots.win, Bitchute, Gab and Odysee. Across these platforms where extremist movements are known to be active, a clear spike in anti-Muslim hatred can be evidenced, followed by a sustained increase across many platforms in the following period.

Recommendation on coordinating responses to incidents in particular arising from the dissemination of illegal content, ahead of the full entry into application of the DSA,*European Commission, 18 October 2023*

The European Commission published a Recommendation (which allows EU institutions to make their views known and to suggest a line of action without imposing any legal obligation on those to whom it is addressed) in the context of “an unprecedented increase in illegal and harmful content being disseminated online” within the EU, as the document refers to both Russia’s war in Ukraine and the terrorist attack by Hamas. The document recommends the following measures:

- Prior to 17 February 2024, Member States are encouraged, through an “informal network”, to coordinate their actions in relation to the dissemination of illegal content on designated VLOPs and VLOSEs, including via appointing their DSCs and sharing their contact details.
- The informal network is recommended to cooperate with the Commission to respond via specific meetings to incidents, in particular arising from the dissemination of illegal content, “posing a clear risk of intimidating groups of population and destabilising political and social structures in the Union or parts thereof, including those which risk leading to a serious threat to public security or public health in the Union or in significant parts of it.”
- Regular Incident Response meetings to: (a) Exchange of information, good practices, methodologies, technical systems and tools with the purpose of supporting supervisory efforts regarding VLOPs and VLOSEs in the context of crisis; and (b) Exchange of information collected at national level from competent national authorities as regards the identification of online illegal content related to the situation of crisis and its amplification by very large online platforms and very large online search engines, including, where available, information regarding its effect in the local public opinion.
- The informal network could, where relevant, provide “meaningful information regarding the functioning and design of relevant VLOPs and VLOSEs.”
- Support to the Commission in monitoring and enforcing the DSA via conducting of interviews and carrying out inspections within their territory.
- Member States are encouraged to participate in “existing voluntary cooperation frameworks to address the dissemination of illegal content online.” Such voluntary cooperation frameworks include the EU Crisis Protocol, which provides for a voluntary mechanism to respond to a suspected crisis in the online space, stemming from a terrorist or a violent extremist act. Member States are also encouraged to coordinate via international fora such as the Christchurch Call and the industry-led Global Internet Forum to Counter Terrorism.

European Commission sends request for information to Meta under the DSA,*European Commission, 19 October 2023*

The Commission requested Meta to provide more information on the measures it has taken to comply with obligations related to risk assessments and mitigation measures to protect the integrity of elections and following the terrorist attacks across Israel by Hamas, in particular with regard to the dissemination and amplification of illegal content and disinformation.

European Commission sends request for information to TikTok under the DSA,*European Commission, 19 October 2023*

The Commission requested TikTok to provide more information on the measures it has taken to comply with obligations related to the risk assessments and mitigation measures against the spreading of illegal content, in particular the spreading of terrorist and violent content and hate speech, as well as the alleged spread of disinformation. The request also addresses TikTok's compliance with other elements of the DSA, in particular with regards to its provisions related to the protection of minors online.

European Commission sends request for information to X under the DSA,*European Commission, 12 October 2023*

The Commission is investigating X's compliance with the DSA, including with regard to its policies and actions regarding notices on illegal content, complaint handling, risk assessment and measures to mitigate the risks identified. The request followed indications received by the Commission services of the alleged spreading of illegal content and disinformation, in particular the spreading of terrorist and violent content and hate speech.

In addition to these formal requests, European Commissioner for the Internal Market Thierry Breton sent letters to [Meta CEO Mark Zuckerberg](#), [X owner Elon Musk](#) and [TikTok CEO Shou Zi Chew](#). Based on the assessment of the company responses to the requests, the Commission will assess next steps. This could entail the formal opening of proceedings pursuant to Article 66 of the DSA. Pursuant to Article 74 (2) of the DSA, the Commission can impose fines for incorrect, incomplete or misleading information in response to a request for information.

Ofcom sends letter to UK-based video-sharing platforms regarding the crisis in Israel and Gaza,*Ofcom, 11 October 2023*

Under Ofcom's current regulation, video-sharing platforms that fall within Ofcom's jurisdiction are required to put in place measures that are appropriate to protect the general public from terrorist videos and video material that incites hatred or violence, and to protect children from video material that may cause them harm. The letter calls on platforms to ensure their systems and processes are effective in anticipating and responding to the potential spread of harmful video material stemming from the developing crisis.

Israel-Hamas war: how social media companies are handling the response,*The Verge, 17 November 2023*

The Verge tracked platform responses following the Hamas terrorist attack on 7 October. The following list summarises and expands on these updates:

- On 13 October 2023, Meta [published](#) a blogpost on its "ongoing efforts regarding the Israel-Hamas war," including how it enforces its policies around [Dangerous Organizations and Individuals](#), [Violent and Graphic Content](#), [Hate Speech](#), [Violence and Incitement](#), [Bullying and Harassment](#), and [Coordinating Harm](#).

- On 15 October 2023, TikTok [published](#) a blogpost on its measures to protect the community, which include for example launching a command centre; evolving its proactive automated detection systems in real-time; adding more moderators who speak Arabic and Hebrew; enforcing its [policies](#) against [violence](#), [hate](#), and [harmful misinformation](#); and adding opt-in screens over content that could be shocking or graphic.
- On 25 October 2023, Telegram [blocked](#) access to the official channel of Hamas, [hamas_com](#) and al-Qassam brigades, the military wing of Hamas, but only on Android phones due to violations of Google's app store guidelines.
- On 13 November 2023, TikTok [published](#) a blogpost entitled "The truth about TikTok hashtags and content during the Israel-Hamas war". The aim of the blogpost was to debunk "misinformation and mischaracterization about how the TikTok platform actually operates," referring among other things to its recommendation algorithms and polling data outlining "skewed attitudes among young people toward Palestine."
- On 14 November 2023, X [published](#) a blogpost outlining its actions with regard to Policy Enforcement, Community Notes, and Brand Safety. The blogpost referred to X's violent speech and [hateful conduct](#) policy, [Violent and Hateful Entities](#) policy, and [Synthetic and manipulated media](#) policy. It also noted that X's [Sensitive media](#) policy allows Graphic Content "so long as it is placed behind a sensitive media warning interstitial," while generally removing content that meets the definition of Gratuitous Gore. However, in times of "certain newsworthy events," X would allow "a range of media" due to the "public's interest to understand what's happening in real-time."

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.