Digital Policy Lab

Policy Digest #18

6 March 2025

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: Report of the Online Safety Act Review released

Type Statutory review of regulation
Status Published

On 12 February 2025, the Minister for Communications, Hon Michelle Rowland MP, tabled the Report of the Statutory Review of the Online Safety Act 2021 in Parliament. The independent review assessed the Act's effectiveness and examined additional protections against online harms, including risks posed by emerging technologies. The report sets out 67 recommendations to strengthen Australia's online safety framework, including enhanced complaints mechanisms, stricter penalties for non-compliance, increased transparency requirements and governance reforms for the Office of the eSafety Commissioner.

Australia: Publication of Online Safety Industry Standards under the Online Safety Act

Type Regulatory (standard)
Status In force

On 22 December 2024, the Online Safety Industry Standard under the Online Safety Act came into effect, established by the eSafety Commissioner. This standard mandates electronic services, including dating websites, gaming platforms and messaging apps, to eliminate child sexual abuse material (CSAM) and pro-terror content. In addition, the Designated Internet Services (DIS) and Relevant Electronic Services (RES) standards will require major technology companies, including file and photo storage services along with messaging platforms, to implement robust measures to prevent the misuse of their services for such harmful material. Generative AI "nudify apps" and online marketplaces offering these models must also comply. These standards come after multiple revisions due to disagreement on appropriate community safeguards regarding content causing the highest harm. They also complement six existing industry Codes for social media, search engines, app stores, internet service providers, hosting services and device manufacturers to enhance online safety.

European Union: Code of Practice on Disinformation to become part of the Digital Services Act (DSA)

Type Regulatory
Status Adopted

On 13 February 2025, the Commission and the European Board for Digital Services approved the integration of the voluntary Code of Practice on Disinformation into the DSA framework. The Code aims to combat disinformation risks while upholding the freedom of speech and enhancing transparency under the DSA. In January 2025, signatories, including Very Large

[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.

Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) such as Facebook, Instagram, and Google Search, submitted documentation to transition the initially voluntary agreement to the binding Code. The integration of the Code will take effect on 1 July 2025, making its commitments auditable from that date forward.

European Union: The European Commission withdraws work on AI Liability Directive

Type Regulatory (Directive)
Status Withdrawn

On 11 February 2025, the European Commission published its work programme for the year, excluding the AI Liability Directive along with other proposed legislations. This decision comes amidst the technology sector's advocacy for simpler regulations concerning the management of harms caused by AI. Initially proposed in 2022, the EU AI Liability Directive aimed to establish uniform rules for non-contractual civil liability involving AI systems, complementing the EU AI Act. This decision reflects a shift in focus towards enhancing AI development, with plans to raise €200 billion of investment for such work.

European Union: Commission publishes the Guidelines on prohibited AI practices under the AI Act

Type Regulatory (Guidelines)
Status Published

On 4 February 2025, the Commission published Guidelines regarding prohibited AI practices as outlined by the AI Act due to their potential risks to European values and fundamental rights. The Guidelines specifically highlight practices such as harmful manipulation, social scoring and real-time remote biometric identification. While providing insights into the Commission's views on these prohibitions, the Guidelines are non-binding, with authoritative interpretations reserved for the Court of Justice of the European Union (CJEU). However, they are intended to promote consistent and effective enforcement of the AI Act across the European Union. These Guidelines include legal clarifications and practical examples to assist stakeholders in understanding and complying with the requirements of the AI Act. The Guidelines are now pending formal adoption.

European Union: Revised Code of Conduct on Countering Illegal Hate Speech Online integrated into the DSA framework

Type Regulatory
Status Published

On 20 January 2025, the revised Code of Conduct on Countering Illegal Hate Speech Online ("Code of Conduct+") was formally integrated into the DSA framework following a positive assessment by the European Commission and the European Board for Digital Services. The updated Code strengthens enforcement measures for online platforms in addressing illegal hate speech under EU and national law. For VLOPs and VLOSEs, adherence may serve as a risk mitigation measure under the DSA. The Code of Conduct+ was signed and submitted for integration under the DSA by Dailymotion, Facebook, Instagram, Jeuxvideo.com, LinkedIn, Microsoft-hosted consumer services, Snapchat, Rakuten Viber, TikTok, Twitch, X and YouTube. The Commission and the Board will monitor compliance and recommend further measures, including enhanced reporting obligations and transparency on content moderation outcomes.

France: Framework for establishing minimum technical standards for age verification systems in accessing online pornographic content

Type Regulatory (standards)
Status Implemented

On 9 January 2025, the Regulatory Authority for Audiovisual and Digital Communication (ARCOM) implemented a technical standard detailing the minimum requirements for age verification systems used by online services that distribute pornographic content. The aim of this standard is to safeguard minors from exposure to harmful material. It applies to online public communication services and video-sharing platforms, requiring that age verification systems effectively differentiate between minors and adults, incorporate strong anti-fraud measures, and adhere to data protection regulation. Furthermore, the standard specifies that these systems must employ a double anonymity mechanism to safeguard user identities and offer at least two methods for generating proof of age. It also highlights that audits of these systems will be necessary, with results published to inform users about privacy protections. ARCOM has the authority to issue compliance notices for failures to meet these standards, and ongoing non-compliance may result in financial penalties.

Germany: Introduction of law against digital violence

Type Legislative
Status Draft

On 9 December 2024, the German Federal Ministry of Justice released a draft for a proposed law aimed at addressing digital violence (Gesetz gegen digitale Gewalt). The legislation seeks to enhance individuals' capacity to take legal action against breaches of their personal rights online. It would streamline the processes for victims of online abuse to identify offenders by mandating that platforms reveal specific information about users who engage in unlawful behaviour online. The draft law further establishes a legal entitlement for victims to request the suspension of a perpetrator's account on online platforms in cases of particularly severe violations. Additionally, it strengthens the responsibilities of social media platforms by requiring them to appoint a designated representative in Germany to manage content moderation and compliance matters. The draft law is a first step towards implementing the EU Directive on combating violence against women and domestic violence, which EU Member States must implement by 14 June 2027.

Republic of Korea: Development of Artificial Intelligence Bill and Establishment of Trust-Based Foundations signed into law

Type Legislative
Status Passed

On 21 January 2025, the Bill on Development of Artificial Intelligence and Establishment of Trust-Based Foundations was signed into law after it passed in the National Assembly on 26 December. The "AI Basic Act" guides the ethical development and use of AI and regulates the use of "high impact" AI. This is similar to the "high-risk" classification under the EU AI Act, covering sensitive areas such as access to essential services, health care and employment practices. Companies must inform users when offering products or services that utilise generative AI or "high-impact" AI. The Act also empowers the government to establish a principles-focused three-year strategy. Additionally, the Act allows organisations to undergo voluntary verification and certification processes to ensure the reliability and safety of AI. The legislation is set to take effect on 22 January 2026.

United Kingdom: Proposed Data (Use and Access) Bill moves onto next stage

Type Legislative
Status Committee stage (House of Commons)

On 13 February 2025, the proposed Data (Use and Access) Bill had its second reading in the House of Commons, marking progress in adopting the Bill. The Bill aims to enhance access to and regulation of customer and business data, including on online platforms. It seeks to establish standards for processing personal information, including provisions for making available data for online safety research while ensuring privacy and improving public service delivery in the digital space. Additionally, it addresses the retention of biometric data and the issuance of electronic signatures to contribute to online safety and trust in digital transactions. The dates for the next sessions are to be announced.

United Kingdom: New laws to combat AI-Generated child sexual abuse material announced

Type Legislative
Status Announced

On 2 February 2025, the UK announced new proposed laws to tackle AI-generated CSAM. Key measures include criminalising the creation, possession, or distribution of AI tools for generating CSAM, banning AI "paedophile manuals", targeting websites facilitating CSAM sharing as well as granting Border Force powers to inspect digital devices of suspected offenders. The laws are part of the upcoming Crime and Policing Bill, aimed at strengthening protections for children amid the growing misuse of AI in online abuse.

United Kingdom: Statement: "Tackling Intimate Image Abuse and Sexually Explicit Deepfakes"

Type Legislative
Status Announced

On 7 January 2025, Alex Davies-Jones, Parliamentary Under-Secretary of State for Justice, outlined new measures to tackle intimate image abuse and creation of sexually explicit deepfakes. The Crime and Policing Bill, to be proposed later this year, will introduce new criminal offences for taking intimate images without consent, and the installation of equipment with the intent to facilitate such actions.

United Kingdom: Ofcom launches "year of action" by issuing multiple Codes and Guidelines

Type Regulatory (enforcement)
Status Published

Since mid-December 2024, Ofcom has published multiple Codes and Guidelines for implementing the Online Safety Act. Publishing the Illegal Harms Codes of Practice and the Illegal Content Risk Assessment Guidance on 16 December marked the beginning of services to comply with their duties, with full implementation expected by March 2025. On 16 January 2025, Ofcom issued its Age Assurance Guidance for services which publish pornographic content, as well as children's access assessment Guidance. On 25 February, Ofcom published its draft Guidance on protecting women and girls for consultation in pursuance of phase two of the implementation progress, focusing on child safety, pornography and the protection of women and girls. Ofcom's implementation timeline for 2025 outlines further Codes and Guidance expected to be published by Ofcom in 2025.

United States: TAKE IT DOWN Act passes the US Senate

Type Legislative
Status Awaiting action in the US House of Representatives

On 13 February 2025, the TAKE IT DOWN Act unanimously passed the Senate. The legislation criminalises the publication of non-consensual intimate imagery (NCII), including AI-generated content, and requires social media platforms to remove NCII within 48 hours of notice from an affected individual. Companion legislation was introduced in the House where it awaits action. The bill previously passed the Senate unanimously in the last legislative session but failed to pass the House.

United States: Kids Off Social Media Bill (SB 278) introduced in the Senate

Type Legislative
Status Introduced

On 28 January 2025, a group of bipartisan Senators introduced the Kids Off Social Media Act (KOSMA). The bill would set a minimum age of 13 for use of social media and bans social media companies from using algorithms to recommend content to users under the age of 17. Enforcement authority will be given to the Federal Trade Commission and state attorneys general. The bill was introduced in the previous legislative session but failed to pass. Notably, KOSMA has faced opposition from civil liberties groups which believe age verification would infringe on users' First Amendment rights and increase surveillance of children.

United States: President Trump Signs an Executive Order Prohibiting the Federal Government from Moderating Content on Online Platforms

Type Executive
Status Implemented

On 25 January 2025, President Donald Trump signed an executive order titled "Restoring Freedom of Speech and Ending Federal Censorship". The executive order bans federal officials and taxpayer resources from engaging in or facilitating "any conduct that would unconstitutionally abridge the free speech of any American citizen". The EO also directs the Department of Justice to investigate the federal government's actions over the past four years and "identify and take appropriate action to correct past misconduct by the Federal Government related to censorship of protected speech".

United States: President Trump grants TikTok a 75-day extension to sell the app or face a ban

Type Executive
Status Implemented

On 20 January 2025, President Donald Trump delayed the enforcement of a TikTok ban with an executive order titled "Application of Protecting Americans from Foreign Adversary Controlled Applications Act to TikTok". The executive order grants TikTok a 75-day extension to sell the app before it faces a ban as required by the Protecting Americans from Foreign Adversary Controlled Applications Act (PAFACAA). Trump directed the Attorney General to not enforce PAFACAA, signalling to app stores that the US government will not impose fines if they continue to host the app.

Global: AI agreement signed at Paris AI Action Summit

Type Voluntary (joint statement)
Status Published

On 10 and 11 February 2025, representatives from governments, international organisations, civil society, academia and the private sector gathered in Paris for the AI Action Summit. The final output of the Summit, the multilateral "Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet" was signed by 63 countries and unions, including the African Union Commission and the European Union and its 27 Member States. However, the UK and US did not sign the statement over alleged disagreements on security, over-regulation, and the framing of inclusivity and sustainability. The next Summit will take place in India.

Section 2 Topic-specific snapshot: "Youth radicalisation"

This section summarises selected analyses and responses published by government agencies, civil society organisations and academia on youth radicalisation.

Young guns: Understanding a new generation of extremist radicalization in the US,
Isabel Jones, Jakob Guhl, David Leenstra, Jacob Davey and Moustafa Ayad, 2023

This report examines the evolving dynamics of youth radicalisation in the United States, highlighting how online ecosystems facilitate the recruitment and mobilisation of young individuals into extremist movements. The report indicates that the distinctions between extremist groups, such as Salafi-jihadists and white supremacists, are becoming increasingly blurred. It underscores the role of fringe social media platforms, gaming communities and encrypted messaging apps in shaping extremist subcultures, where ideological narratives are adapted to appeal to younger audiences. It points out the strategic use of digital spaces to normalise extremist ideologies, often through irony, memes and gamified engagement tactics that lower barriers to entry.

A core finding of the report is that the extremist threat landscape in the US has become more fragmented and decentralised, with a diverse array of extremist groups, movements and ideologies operating across a range of digital platforms. The study also assesses the limitations of existing counter-radicalisation efforts, pointing to the need for more adaptive, platform-specific interventions that account for the rapidly shifting nature of online youth culture. To mitigate these risks, the report recommends a multi-stakeholder approach, emphasising platform accountability, digital literacy programs, and targeted prevention strategies that address both online and offline vulnerabilities contributing to youth radicalisation.

764: The Intersection of Terrorism, Violent Extremism, and Child Sexual Exploitation,
Marc-André Argentino, Barrett G and M.B. Tyler, 2024

This article explores the intersection of terrorism, violent extremism, and child sexual exploitation through case studies of the extremist 764 network, which overlaps with the Order of Nine Angles (O9A)[2] ideology. The 764 network is an online child extortion group established by a Discord user named Bradley Cadenhead in 2021. It encourages acts of sextortion, animal abuse, incest, rape, self-harm, and bestiality. 764 recruits new members through online platforms, particularly in gaming communities and specifically targets vulnerable individuals, including LGBTQ+ youth, racial minorities, and individuals with mental illness. The report frames the recruitment process as cross-platform abuse, where perpetrators initiate contact on one platform before moving to less moderated spaces like Discord to manipulate, groom and extort victims.

The authors argue that social media platforms must recognise the risks inherent on their platforms and strengthen content moderation efforts to address these threats. Additionally, they caution that well-meaning initiatives aimed at

[2] O9A is a secretive, neo-Nazi, and satanic extremist movement which promotes occultism, violence, terrorism, and accelerationist ideology, seeking to undermine society through chaos and destruction. O9A combines esoteric fascism, Satanism, and Social Darwinism.

raising awareness may inadvertently backfire by exposing individuals to these networks, increasing the likelihood of them becoming victims.

Childhood Innocence?: Mapping Trends in Teenage Terrorism Offenders,
Hannah Rose and Gina Vale, 2023

This report examines children's involvement in terrorism in England and Wales since 2016, with the goal of deepening the understanding of the evolving terrorist threat landscape. The authors utilise both descriptive statistics and qualitative analysis of 43 cases involving children convicted of terrorist offenses in 2016 to explore the characteristics of this "new generation" of terrorists. The findings highlight the prevalence of young males, with terrorist acts predominantly tied to Islamist and extreme-right ideologies. The report identifies that children are most frequently convicted for the preparation of terrorist acts, followed by encouragement of terrorism, noting that all cases of the latter were linked to social media posts. The authors suggest that encouraging terrorism serves as an intermediary step between spreading propaganda and planning an actual attack.

The study reveals the significant role of online influence and interconnectivity across ideologies. The authors find that radicalisation dynamics can unfold in multiple directions. Specifically, in online environments where the age of individuals is often hidden, young people can have a considerable impact. By influencing not only peers but also older and younger individuals, they contribute to the broader process of radicalisation.

While acknowledging that children can be vulnerable to recruitment, the authors challenge the stereotype of helpless victims. They present a nuanced picture of children in terrorism, demonstrating how young individuals can pose a security threat themselves.

Thirty-fifth report of the Analytical Support and Sanctions Monitoring Team submitted pursuant to resolution 2734 (2024) concerning ISIL (Da'esh), Al-Qaida and associated individuals and entities,
United Nations Security Council, 2025

This monitoring report focuses on the Islamic State in Iraq and the Levant (ISIL) and Al-Qaida, highlighting recent developments in the global terrorist threat landscape. Regarding youth radicalisation, the report lays out that the average age of individuals being radicalised is decreasing in several Member States. It provides information on emerging tools such as 3D printing and artificial intelligence (AI), enabling AI-driven propaganda and targeted recruitment tactics that pose significant risks to young people.

When examining the Arabian Peninsula, it was observed that Al-Qaida in the Arabian Peninsula's (AQAP) media arm leveraged technologies such as cryptocurrencies for activities like campaigning. Additionally, they have increasingly utilised video games as a platform to recruit and influence youth.

In Europe, radicalised individuals are increasingly of a younger age or even minors, often recruited through encrypted messaging platforms. The report refers to recent arrests in Austria and France to underscore the direct links between young radicals and ISIL (Da'esh) networks. Furthermore, the report highlights the Gaza-Israel conflict as a catalyst for the radicalisation of highly vulnerable young individuals, emphasising that they are driven more by violence than by ideological beliefs. It also underscores the significant role of online platforms, where social media algorithms amplifying extremist content contribute to radicalisation processes. In South-East Asia, while the overall terrorist threat remains relatively low, online self-radicalisation among youth is on the rise, with the period from radicalisation to activation becoming shorter. Due to the evolving threat of youth radicalisation, the report recommends strengthening preventive efforts that tackle the root causes of radicalisation and violent extremism. Enhanced monitoring and regulation of alternative internet platforms and encrypted chat applications are necessary to curb online recruitment. Furthermore, collaboration with the UN Office on Drugs and Crime is welcomed by the Monitoring Team.

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.