Digital Policy Lab

Policy Digest #11

26 June 2023

ISD Powering solutions to extremism, hate and disinformation

Sponsored by

Federal Republic of Germany
Foreign Office

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: Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023

Type Regulatory
Status Exposure draft

On 24 June 2023, Australia's Department of Infrastructure, Transport, Regional Development, Communications and the Art opened a call for feedback on an exposure draft of the Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023. According to the Fact sheet, the Australian Communications and Media Authority (ACMA) would receive powers "to gather information from digital platform providers, or require them to keep certain records about matters regarding misinformation and disinformation; to request industry develop a code of practice covering measures to combat misinformation and disinformation on digital platforms, which the ACMA could register and enforce; to create and enforce an industry standard (a stronger form of regulation)". The ACMA would not have the power to request specific content be removed from services. The powers will apply to "digital platform services that are accessible in Australia" such as "social media, search engines, instant messaging services (although the content of private messages will be out of scope), news aggregators and podcasting services". Submissions close on 6 August.

Australia: Reporting notice issued to Twitter on online hate

Type Regulatory
Status Legal notice issued

On 21 June 2023, the eSafety Commissioner issued a non-periodic reporting notice to Twitter under section 56(2) of the Online Safety Act 2021. In their reasoning, eSafety states that it has received more complaints about online hate on Twitter in the past 12 months than any other platform, and has received an increasing number of reports of serious online abuse since Elon Musk's takeover of the company. This notice requires Twitter to explain what it is doing to minimise online hate, including how it is enforcing its terms of use and hateful conduct policy. eSafety's regulatory powers cover serious adult online abuse as well as the cyber bullying of children and image-based abuse. In some cases, hate speech may meet the statutory thresholds of adult cyber abuse. In particular, the regulation allows eSafety to require online service providers to report on how they are meeting any or all of these Basic Online Safety Expectations. The obligation to respond to a reporting requirement is enforceable and backed by civil penalties. If Twitter fails to respond to the most recent notice within 28 days, the company could face maximum financial penalties of nearly AUD 700,000 a day for continuing breaches.

[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.

Australia: Industry codes registered under the Online Safety Act 2021

Type Regulatory
Status Implementation

On 16 June 2023, the eSafety Commissioner registered mandatory industry codes that cover five sections of the online industry and operate under the Online Safety Act 2021. The codes require industry to take adequate steps to reduce the availability of seriously harmful online content, such as child sexual abuse and pro-terror material. The registered online safety codes cover Hosting Services, Equipment, Internet Carriage Services, App Distribution Services and Social Media Services, and will come into effect six months following registration. The Commissioner found that two of the codes (Relevant Electronic Services and Designated Internet Services) did not provide appropriate community safeguards, while reserving a decision on the code covering Internet Search Engine Services, granting industry an additional four weeks to resubmit.

EU: European Media Freedom Act (EMFA)

Type Regulatory
Status Awaiting Parliament's position (ordinary legislative procedure)

On 21 June 2023, the Council of the EU adopted its general approach on the proposed regulation for 'Establishing a common framework for media services in the internal market' (European Media Freedom Act, EMFA). The EMFA would introduce safeguards against the "unjustified removal" of content produced by media services providers (MSPs) that meets the "editorial standards" in the Member State in which the provider is established. The EMFA would require very large online platforms (VLOPs) - defined in the Digital Service Act (DSA) - to notify MSPs of the violations of their Terms of Service and provide the reasons for content removal before deleting it from their platform. VLOPs would be required to process the complaints received from MSPs and issue an annual report outlining the number of restrictions imposed on MSPs.

The Council's position "clarifies the responsibility of the Member States to guarantee the plurality, independence and proper functioning of public MSPs operating within their borders; ensures that Member States are able to adopt stricter or more detailed rules than those set out in the EMFA; broadens the scope of the requirements on transparency, both for transparency of ownership and for the transparency of state advertising; and provides clearer rules on the relationship between VLOPs and MSPs that adhere to regulatory or self-regulatory regimes of editorial control and journalistic standards in Member States, with the aim to ensure that such content is treated with extra care". The presidency of the Council now has a mandate to begin negotiations with the European Parliament once the latter has established its position on the regulation.

EU: Artificial Intelligence (AI) Act

Type Regulatory
Status Tripartite meetings (ordinary legislative procedure)

On 14 June 2023, the European Parliament adopted its negotiating position on the proposed AI Act. The rules would ensure that AI developed and used is fully in line with EU rights and values including human oversight, safety, privacy, transparency, non-discrimination and social and environmental wellbeing. AI systems with an "unacceptable level of risk" to people's safety would therefore be prohibited, such as those used for social scoring. AI systems used to influence the outcome of elections and in recommender systems used by very large online platforms were added to the "high-risk" list. Generative AI systems such as ChatGPT would have to comply with transparency requirements and ensure safeguards against generating illegal content. Tripartite meetings will now begin with the Council and the Commission to negotiate the final text of the law. While these 'trilogues' can be a lengthy process, it is possible that the Act will be adopted by the end of 2023.

New Zealand: New regulatory framework for safer experiences on online services and media platforms

Type Regulatory
Status Public consultation

On 6 June 2023, New Zealand's Department of Internal Affairs opened a public consultation on a new "regulatory framework for safer experiences on online services and media platforms". It published a discussion document which outlines the government's proposals. The objective of this framework would be to "enhance protection for New Zealanders by reducing their exposure to harmful content, regardless of delivery method". Under the proposals, platforms "would be brought into one cohesive framework with consistent safety standards". The government proposes to create "codes of practice that set out specific safety obligations for larger or riskier platforms". These codes would be enforceable and approved by a newly created independent regulator.

Regulatory efforts would "focus on the areas of highest risk, such as harm to children or content that promotes terrorism". Regulated platforms would cover platforms "where their primary purpose is to make content available". The platform or service is likely to have "an expected audience of 100,000 or more annually; or 25,000 account holders annually in New Zealand". Alternatively, the regulator may designate platforms "if it is unclear whether the threshold has been met, or the risk of harm from that platform is significant". The discussion document also reiterates that the new framework "would retain powers of censorship for the most extreme types of content (called 'objectionable' material)", which is already illegal. The regulator would have powers to "require illegal material to be removed quickly from public availability in New Zealand" in cases of 'objectionable material' as well as "material that is illegal for other reasons, such as harassment or threats to kill". The closing date for feedback is 31 July.

US: Supreme Court ruling in Twitter v. Taamneh

Type Court ruling
Status Issued

On 18 May 2023, the Supreme Court in Twitter v. Taamneh ruled against the family of a 2017 ISIS attack victim who sought to hold social media companies liable for allowing ISIS to use their platforms. The lawsuit relied on the Anti-Terrorism Act, which allows U.S. nationals to sue anyone who "aids and abets" international terrorism. The plaintiffs argued that Twitter contributed to the terrorist organisation's growth by allowing ISIS to use the platform for recruitment and propaganda, alleging that Twitter recommendation algorithms do not constitute "passive aid" but rather qualify as "substantial assistance". The Court differed and noted that Twitter recommendation algorithms are part of the infrastructure the platform provides to users. The fact that the algorithms matched ISIS content with users does not qualify as "active abetting" or "substantial assistance".

The Court thus sidestepped a ruling in a separate case (Gonzalez v. Google) on the scope of Section 230 of the Communications Decency Act, which generally shields platforms from liability for content published by their users. The case concerns a lawsuit filed by the family of Nohemi Gonzalez, who was killed in the 2015 ISIS attack in Paris. The family argues that Google aided ISIS's recruitment by allowing ISIS to post videos on YouTube that incited violence. The U.S. Court of Appeals for the Ninth Circuit ruled that Section 230 protects algorithmic recommendations, at least if the provider's algorithm treated content on its website similarly. In an opinion issued on the same day (18 May), the Supreme Court noted, "much (if not all) of" the family's "complaint seems to fail under either our decision in Twitter or the Ninth Circuit's unchallenged holdings". Therefore, the Court reasoned, there was no need for it to weigh in on the scope of Section 230 now.

Section 2 Topic-specific snapshot: "Foreign Information Manipulation and Interference (FIMI)"

This section presents summaries of selected analyses and commentary published by governments, civil society and academia on the topic of foreign information manipulation and interference (FIMI).

Foreign digital interference – France's detection of an information manipulation campaign, *VIGINUM*, 13 June 2023

On 13 June 2023, the French public authority monitoring digital foreign interferences VIGINUM published a report on the Russian digital information manipulation campaign RRN, named after the website "Reliable Recent News", which publishes pro-Kremlin content in English, German, French, Italian, Chinese and Arabic. According to the investigation conducted by VIGINUM, the campaign relies on several modus operandi: creating websites which share audio-visual content criticising Ukrainian leaders; impersonating media outlets and government websites; creating French-speaking news websites; and creating networks of inauthentic accounts mainly on Facebook and Twitter. The content of the campaign focuses on four main themes:

1. The alleged ineffectiveness of the sanctions against Russia;
2. The alleged Russophobia of Western states;
3. Barbaric acts allegedly committed by Ukrainian armed forces and the neo-Nazi ideology that would predominate among Ukrainian leaders;
4. The negative effects on Europe allegedly created by Ukrainian refugees.

VIGINUM revealed information pointing to the involvement of Russian or Russian-speaking individuals and several Russian companies in the design and conduct of this campaign. The report also notes that several government bodies or bodies affiliated with the Russian state participated in spreading certain content produced under this campaign.

VIGINUM cites previous investigations into the pro-Kremlin impersonation of news websites conducted by the EU Disinfo Lab, the Institute for Strategic Dialogue (ISD) and the Atlantic Council's Digital Forensic Research Lab (DFRLab). French Foreign Minister Catherine Colonna condemned the campaign, stating, it is "not worthy of a permanent member of the UN Security Council. No manipulation attempt will dissuade France from supporting Ukraine in the face of Russia's war of aggression". A summary of the VIGINUM report in English can be found here.

1st EEAS Report on Foreign Information Manipulation and Interference Threats, *European External Action Service (EEAS)*, 7 February 2023

The first edition of the EEAS report on Foreign Information Manipulation and Interference (FIMI) applies a novel framework to a sample of 100 FIMI incidents detected and analysed between October and December 2022. The report defines FIMI as, "a mostly non-illegal pattern of behaviour that threatens or has the potential to negatively impact values, procedures, and political processes. Such activity is manipulative in character, conducted in an intentional and coordinated manner. Actors of such activity can be state or non-state actors, including their proxies inside and outside of their own territory." The definition "overlaps with the notion of disinformation, but is at the same time narrower and broader". For one, "it only refers to information manipulation by actors foreign to the EU and its member states, thus not applying to domestic sources". It is broader insofar as it "does not require the information spread by threat actors to be verifiably false or misleading. The deciding factor for whether something can be considered FIMI is "deceptive or manipulative behaviour" – these patterns of behaviour are described as "Tactics, Techniques, and Procedures (TTPs)". The analytical framework for FIMI threat analysis applies James Pamment's ABCDE framework to differentiate FIMI incidents in terms of actors, behaviours, content, degree, and effect, as well as the DISARM framework for operationalising the concept of 'behaviour' in the ABCDE framework.

The main findings of threat analysis include:

- Russia's full-scale invasion of Ukraine dominates observed FIMI activity. Ukraine and its representatives have been the direct target of 33 incidents;
- Diplomatic channels are an integral part of FIMI incidents;
- Impersonations of international and trusted organisations and individuals are used by Russian actors particularly to target Ukraine. Print and TV media are most often impersonated;
- FIMI actor collusion exists but is limited;
- Incidents do not occur in just one language, content is translated and amplified in multiple languages;
- FIMI is mostly intended to distract and distort. In the case of incidents carried out by Russia, 42% were intended to distract, mostly in the context of the Russian invasion of Ukraine, for example, to turn attention to a different actor/narrative. Another 35% aimed to distort, twist and frame narratives around the invasion. In the case of China, the majority (56%) of incidents intended to distract, for example, to promote China as a reliable partner;
- FIMI remains mostly image and video-based. The cheap and easy production and distribution of such material online makes these formats the most commonly used, for example, fabricated image and video-based contents were used to degrade the adversaries' image or ability to act and to discredit credible sources.

The report contributes to the implementation of the Strategic Compass' call for a "FIMI Data Space". The 2023 annual progress report on the Strategic Compass states that FIMI is "increasingly used as part of broader hybrid campaigns". The progress report highlights that EEAS is working with international partners, including the G7 and NATO, as well as stakeholders from civil society and private sector on "establishing a new central FIMI data space for gathering information on threats stemming from disinformation and foreign information manipulation". Furthermore, the EEAS "stepped up efforts to equip CSDP missions and operations, in particular in Sub-Saharan Africa, with the capabilities and resources to help counter FIMI campaigns".

DISinformation Analysis and Risk Management (DISARM) framework as a common methodology, *EU-US Trade and Technology Council (TTC),* 29 May 2023

The Trade and Technology Council (TTC) Fourth Ministerial published a statement on "Foreign information manipulation and interference in third countries", noting their mutual concern about FIMI and disinformation. As part of its response, the TTC seeks a common methodology for identifying, analysing and countering FIMI to increase transatlantic cooperation. It thereby approved the DISARM framework as part of a "common standard for exchanging structured threat information on FIMI". The TTC agreed that "information will be shared more efficiently, effectively and with a greater level of detail when it comes to understanding the manipulative tactics, techniques and procedures".

DISARM is a "open-source, master framework for fighting disinformation through sharing data and analysis and coordinating effective action". It provides a taxonomy of techniques and tactics initially developed as the AMITT (Adversarial Misinformation and Influence Tactics and Techniques) framework. DISARM involves two main frameworks: DISARM Red, for describing incident creator behaviours, and DISARM Blue, to describe potential response behaviours. The frameworks contain object types, including tactic stages (steps in an incident) and techniques (activities at each tactic stage). Alliance4Europe supported the establishment of the DISARM Foundation as an independent entity dedicated to maintaining the intellectual property of the framework to protect its openness to the stakeholder community, and ensure the fair, open and transparent governance necessary for its enhancement, promotion and support by and for the counter disinformation community.

FIMI: Towards a European Redefinition of Foreign Interference, *EU DisinfoLab*, 7 April 2023

In this report, *EU DisinfoLab* unpacks the development, terminology, limitations and potential evolutions of the FIMI concept applied by the European External Action Service (EEAS). The author notes that "the choice to maintain a very clinical approach to FIMI, focused on operating modes and detached as much as possible from the content or the actors, also responds to strong political constraints and consensus requirements demanded by the different perspectives of the Member States and international partnerships". In terms of the limitations of the threat analysis, the author highlights that a restriction to Russia and China as threat actors leaves out large grey areas – referring for example to the malign actors exposed by the Forbidden Stories journalistic collective which exposed profiles much more varied than those linked to Russia and China alone.

The report concludes that, "while the change of focus from content (and the obsession with the ill-named "fake news") to behaviour is refreshening, it should not be exclusionary, nor should it sacrifice the analysis of narratives, the verification of facts, or the understanding of the political motivations of actors".

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.