Digital Policy Lab

Policy Digest #10

28 April 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

EU: European Centre for Algorithmic Transparency (ECAT)

Type Regulatory
Status Enforcement

On 18 April 2023, the European Centre for Algorithmic Transparency (ECAT) was officially inaugurated by the European Commission's Joint Research Centre in Seville, Spain. The ECAT was created to support enforcement of the EU's Digital Services Act (DSA), which imposes risk management obligations for very large online platforms (VLOPs) and very large online search engines (VLOSEs). ECAT will provide the Commission with in-house technical and scientific expertise to ensure that algorithmic systems used by the VLOPs and VLOSEs comply with the risk management, mitigation and transparency requirements. This includes the performance of technical analyses and evaluations of algorithms. According to ECAT, an interdisciplinary team of data scientists, AI experts, social scientists and legal experts will assess the functioning of algorithms and propose best practices to mitigate their negative impact.

EU: Proposed European Media Freedom Act (EMFA)

Type Regulatory
Status Awaiting committee decision (Ordinary legislative procedure)

On 31 March 2023, the Committee on Culture and Education (CULT) in the European Parliament published its draft report on the proposed regulation for 'Establishing a common framework for media services in the internal market (European Media Freedom Act, EMFA). On 13 April 2023, the Committee on the Internal Market and Consumer Protection (IMCO), the opinion giving committee, published its draft opinion and amendments (see here and here). In their amendments, MEPs emphasised the independence of the European Board for Media Services, which will replace the European Regulators Group for Audiovisual Media Services (ERGA). Amendments also considered platforms' power in media service providers' self-declarations, for example, stipulating that "media service providers' self-declarations must be easy to verify", including proposed measures of setting up an "independent, rapid and effective complaint and redress mechanism" or transferring the responsibility to Member States. This provision was criticised by civil society organisations for leaving the decision of whether a media outlet qualifies as a 'media service provider' and is editorially independent entirely to the discretion of platforms. Amendments further addressed the notification of 'media service providers' prior to any removal of content or suspension of accounts taking effect, with some MEPs proposing to do so "within no longer than 48 hours from a complaint being lodged".

In March 2023, German media associations (representing Axel Springer SE, Bauer Media Group and Frankfurter Allgemeine Zeitung, amongst others) published an open letter in support of splitting the EMFA, which would remove and revise certain provisions into a directive "to give leeway to the Member States while implementing them in a way that fits best for their relevant media markets". For example, revising "Article 3 – Rights of recipients of media services", the letter argues, "Media pluralism and diversity of opinion to the benefit of the public discourse don't have any link to internal market regulation."

[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.

The proposal is faced with significant opposition from within the European Parliament. In the Council of the EU, a similar proposal was informally raised by Germany with the support of Poland but ultimately lacked necessary support from other Member States.

The Swedish EU Council presidency is planning a ministerial discussion at the Education, Youth, Culture and Sport Council on 16 May 2023 on how the EMFA will strengthen the media sector. Ministers are expected to endorse Council amendments introduced to the text and provide political guidance on crucial aspects of the regulation.

Germany: Law against digital violence

Type Regulatory
Status Key points (Proposal)

On 12 April 2023, the Federal Ministry of Justice (Bundesministerium der Justiz, BMJ) presented proposed key points for a 'Law against digital violence' (Gesetz gegen digitale Gewalt). According to the points laid out in the proposal, the law intends to make it easier for victims of digital violence to take action in the digital space. Specifically, the proposed law will allow for account blocking if an agitator repeatedly insults a target. Blocking an account should also be possible in cases where only the account's name is available, and the holder's real name is unidentifiable. To implement account blocking in flagrant cases, district courts should be able to issue a temporary injunction. Additionally, those affected by digital violence should also be entitled to receive more information about the agitator. For this reason, platform providers will have to disclose the IP address. Subsequently, the IP address can be used to ask the telecom provider which person was using this IP address at a particular time. The identified individuals can then be sued for damages, or criminal charges could be filed. Civil society organisations have until 26 May to comment on the key points. Based on the feedback, a draft bill will be prepared and is expected to be presented in the second half of 2023.

Germany: Federal Office of Justice (BfJ) investigation into Twitter regarding content moderation

Type Regulatory
Status Compliance

On 4 April 2023, the German Federal Office of Justice (Bundesamt für Justiz, BfJ) opened fine proceedings against Twitter concerning its content moderation under the Network Enforcement Act (NetzDG). According to the BfJ, there are sufficient indications of systemic failure to comply with content moderation requirements (to delete unlawful content within seven days of reporting or 24 hours in case its unlawfulness is manifest) and the obligation to provide an effective and transparent procedure for dealing with complaints from users about illegal content. The case is built on a series of tweets published over a period of four months and containing similar defamatory statements directed to the same person, which users had reported to Twitter. Accordingly, the failure of Twitter to remove reported content indicates a systemic failure to comply with content moderation requirements.

The BfJ has given Twitter time to provide evidence of its compliance with the content moderation rules. Following an assessment of the evidence provided, the BfJ can, in case the evidence is not sufficient, file a lawsuit with the responsible District Court for a preliminary ruling. In case the Court finds the content to be illegal, the BfJ will be able to impose fines under the NetzDG.

Switzerland: Draft bill for consultation on the regulation of communication platforms

Type Regulatory
Status Call to prepare a draft bill for consultation

On 5 April 2023, the Federal Council instructed the Federal Department of the Environment, Transport, Energy and Communications (DETEC) to prepare a draft bill for consultation on the regulation of communication platforms. The announcement notes that platforms lack transparency about their content moderation systems, stating that "a platform may block a user's account or delete content" with "little or no opportunity for recourse". This draft bill would thereby strengthen the rights of users in Switzerland and demand more transparency from the platforms "without limiting their positive influence on freedom of expression". Where appropriate, the new regulation would be based on the EU's Digital Services Act.

Specifically, the Bill would require large platforms to have a point of contact and a legal representative in Switzerland; provide users with review mechanisms for deleted or blocked content, including the set-up of an independent Swiss arbitration board funded by the platforms; require platforms to indicate all advertising as such and, in the case of target group specific advertising, publish the main parameters; and require platforms to establish reporting mechanisms for certain illegal speech. The Federal Council has instructed DETEC, with the involvement of the Federal Office of Justice (FOJ), to prepare a bill for consultation on this issue by the end of March 2024.

UK: Online Safety Bill (OSB)

Type Regulatory
Status Committee Stage

As the Online Safety Bill (OSB) entered its Committee Stage in the House of Lord on 19 April 2023, amendments were published in the 'marshalled list' of amendments. Proposed changes include amendments to create a duty for 'providers of user-to-user services' to manage harmful suicide or self-harm content; to set out a clear definition of harmful suicide or self-harm content; and to create a duty on Ofcom to report about suicide and harm. The amendments also include new clauses (following clause 117), outlining "responsibilities" for providers when they are issued a notice by a senior coroner in an investigation or inquest into the death of a child, as well as "duties of OFCOM in certain cases where a child has died".

Regarding clause 39, several amendments would remove the Secretary of State's (SoS) ability to direct Ofcom on a draft code of practice. The SoS may instead "write to Ofcom with non-binding observations to which Ofcom must have regard". Carnegie UK published a note on SoS powers and Ofcom's independence, noting that need to amend SoS extensive powers has been echoed by the Delegated Powers and Regulatory Reform Committee and the Select Committee on the Constitution.

Furthermore, Lord Bethell and Lord Clement-Jones proposed an amendment to accelerate data access for accredited researchers and civil society organisations. The amendment includes a process for Ofcom to issue a code of practice for researchers. The amendment aims to enable data access while ensuring user privacy and protecting trade secrets.

Previously, the new Department for Science, Innovation and Technology (DSIT) Ministerial team answered questions in the Commons, during which Secretary of State Michelle Donelan confirmed that the Government was "committed to ensuring that [the OSB] is passed before the end of the current session". As Committee progresses, daily amendment lists and updated, numbered marshalled lists of amendments are published here.

US: Annual Threat Assessment of the U.S. Intelligence Community

Type Non-regulatory
Status Published

On 6 February 2023, the Office of the Director of National Intelligence published the Annual Threat Assessment of the U.S. Intelligence Community. The assessment includes a section on "trends in digital authoritarianism and malign influence" that highlights threats posed by "foreign states' malicious use of digital information and communication technologies". Furthermore, the assessment notes "a backdrop of broader digital influence operations that many autocrats are conducting globally to try to shape how foreign public's view their regimes, create social and political upheaval in some democracies, shift policies, and sway voters' perspectives and preferences".

Additionally, the assessment considers the threat of transnational Racially or Ethnically Motivated Violent Extremists (RMVEs) – identified as "a decentralised movement of adherents to an ideology that espouses the use of violence to advance white supremacy, neo-Nazism, and other exclusionary cultural-nationalist beliefs" – which poses "the most lethal threat to U.S. persons and interests" through "attacks and propaganda that espouses violence". The assessment highlights, "The transnational and loose structure of RMVE organizations challenges local security services and creates a resilience against disruptions." It finds that RMVEs "capitalise on societal and political hyperpolarisation to try to legitimatise their aims and mainstream their narratives and conspiracy theories into the public discourse". The assessment also asserts that Terrorgram, "a loosely connected network of channels on the messaging application Telegram", would seek to to circumvent content moderation efforts "to share propaganda, exchange operational guidance, and valorise the perpetrators of previous terrorist attacks". Finally, in the context of the immediate aftermath of Russia's invasion of Ukraine, the assessment observes that "Ukraine featured heavily in online discussions among foreign RMVEs".

Section 2 Topic-specific snapshot: "Post-organisational extremism"

This section presents summaries of selected proposals, analyses and commentary published by governments, civil society and academia on the topic of 'post-organisational' extremism.

Radical reinforcement: The January 6 attack and the methodology of hybridized extremism, *Institute for Strategic Dialogue (ISD),* 17 February 2023

In this analysis, ISD showcases the "fundamental shifts in the organising principles of violent extremism". It notes that while this threat has traditionally been driven by highly organised and coordinated groups, recent years have seen the growth of post-organisational manifestations of extremism, driven by individuals and loose communities with little connection to formal organisations. Extremist actors are thereby more likely to be mobilised by "amorphous ideological convictions, political grievances and shared hatred and hostility toward specific groups and institutions", suggesting that extremism is becoming "hybridised" with other threats. ISD finds that "fringe, conspiracy-driven subcultures have grown closer to the political mainstream in the US", adding to a conceptual blurriness.

Previous ISD analysis on the events leading up to January 6 shows that "mobilisation to violence began well before the election itself" and was "nurtured" in both mainstream and fringe online spaces, as extremist actors capitalised on COVID-19 disinformation and conspiracy theories to expand their audiences and influence. These movements capitalise on "latent prejudices and hateful opinions", including racism, misogyny, xenophobia and anti-LGBTQ+ hate, which are then swept up in "broader conspiratorial worldviews". ISD finds that channels on fringe and mainstream platforms regularly pivot focus from unfounded claims of election fraud to disinformation about vaccines, conspiracy theories about global elites, and hateful and exclusionary sentiments, typically following the drumbeat of right-wing media, fringe online influencers and politicians. With such "flexible points of focus", extremist groups could "maintain a broad base that is angry, engaged and primed to accept the next talking point that fits within their conspiratorial frame".

The Buffalo Attack: The Cumulative Momentum of Far-Right Terror, *Combating Terrorism Center at West Point,* July 2022.

This article examines the 2022 Buffalo racist mass shooting, its perpetrator, his pathway to violence as well as the techniques, tactics, and practices that underpinned his attack. The article argues that "the Buffalo massacre was not an isolated phenomenon. Indeed, one can only fully comprehend it when considered within a continuum of self-referential extreme right terrorism inspired by the March 2019 terrorist attack on two mosques in Christchurch, New Zealand," that had "a catalytic effect upon extreme-right actors, sparking a chain reaction of mass shootings". In each instance, the article asserts, extreme-right terrorists "sought to exceed its death toll, incite further violence, and honor the attacks with their own violence".

The Buffalo terrorist's ideological affinity with the Christchurch attacker extended to copying his modus operandi and attack aesthetics, underscoring the need for researchers, analysts and policymakers to consider such manifestos "collectively rather than singularly" since "they constitute part of the same body of work". Still, the article notes ideological variations, which reflect "the diversity and intersectionality of extreme-right prejudices" as well as "significant structural difference between the Buffalo and Christchurch manifestos". Moreover, the article finds that while acting alone, each of the attackers was immersed in a shared online ecosystem, which – in many cases – "replaced the physical group or party as the principal point of contact for engaging with extreme-right ideologies".

Transparency Report for the Terrorist Content Analytics Platform, *Tech Against Terrorism*, 17 March 2022

The Terrorist Content Analytics Platform (TCAP) is a tool to identify, verify, and alert terrorist content to tech platforms for removal. The platform has been funded by Public Safety Canada. The first Transparency Report provides a detailed breakdown of the core metrics for the report period between 1 December 2020 and 30 November 2021, and of key TCAP policies and processes. In sum, the TCAP sent 11,074 alerts to 65 tech companies, 94% of which is now offline. It alerted 10,959 pieces official verified content from designated Islamist terrorist groups, 94% of which was removed following alerts. It also alerted 115 pieces of official verified far-right terrorist content, 50% of which was removed following alerts. Researchers explain that the discrepancy in numbers is due to "different online propaganda dissemination techniques deployed by far-right and Islamist terrorist groups", and due to "there being fewer far-right terrorist groups designated by democratic nation states". Researchers emphasised the need to include content spanning multiple ideologies, with a particular focus on the global violent far-right.

Towards a Truly Post-Organisational UK Far Right? The Usefulness of a Newly Emergent Concept, *Global Network on Extremism & Technology (GNET)*, 16 December 2021

In this analysis, Dr. William Allchorn argues that the notion of "a decentralised, leaderless group or movement – at least on the far right – is not a recent innovation", referring to Louis Beam who coined the concept of "leaderless resistance" to describe the need for white supremacists to "take action in small cells of one to six men". The analysis acknowledges, "The ability to quantify movement cohesiveness, unity, and impact is becoming more and more difficult for researchers and practitioners alike." The concept of post-organisation is defined as "decentralised, non-hierarchical movement that requires little or no organisational unity in order to work broadly in the same tactical direction and towards similar ideological goals".

Using the CARR FRGB Dataset Research Report, the analysis applies "organisational disunity", "strategic diversity" and "ideological unity" as proxies to measure the level of post-organisation in the UK far-right protest scene between 2009 and 2020. First, the analysis observes an "implosion of key electoral actors" as well as a "broader fragmentation", noting "a more closed opportunity structure for [others] to develop into more formal political actors". This could allow for the possibility of "a predominant actor" to appear on the UK scene. Secondly, the analysis looks at the trend towards ideological fragmentation, noting "clear upticks in the range of [mobilisation] rationales – especially with the onset and emergence of white supremacist organisations". Still, the analysis notes that the level of rationales could be linked to "an organic outgrowth of an uptick in protest activity". Finally, the analysis emphasises a convergence of "strategies of demonstrations, counter-demonstrations, and disruption/vigilante type events", recalling a closed opportunity structure when it comes to engaging in formal party politics. Ultimately, the author calls for an analysis of online versus offline trends to enable a fuller picture.

A Taxonomy for the Classification of Post-Organisational Violent Extremist & Terrorist Content, *Institute for Strategic Dialogue (ISD)*, 9 December 2021

In recognition of 'post-organisational' violent extremism and terrorism, this paper outlines a prototype taxonomy for classifying terrorist and violent extremist content. Accordingly, it is designed to be group-agnostic and shaped around analysis of content, which is influential to violent extremism and terrorism beyond that produced by proscribed terrorist organisations. The paper aims to inform content moderation decisions made by platforms, including adjustments to the hash sharing database of the Global Internet Forum to Counter Terrorism (GIFCT), which provides unique digital "fingerprints" of known terrorist content which has been removed from social media platforms.

Terrorist Definitions and Designations Lists What Technology Companies Need to Know, *Global Research Network on Terrorism and Technology*, 19 July 2019

This publication is part of a series of papers released by the Global Research Network on Terrorism and Technology. The research conducted by this network seeks to better understand radicalisation, recruitment and the myriad of ways terrorist entities use the digital space. The research notes that platforms usually refer to third-party terrorist definitions and designation lists when moderating potential terrorist content and accounts. Yet, those definitions and lists are often produced for specific legal, political or academic purposes and may not be suitable for general use.

The research suggest that platforms should define terrorist entities in a way that distinguishes them from non-violent dissidents, state actors, conventional rebel groups, and criminals or criminal syndicates, while using government designation lists with caution, since such lists compiled are more likely to include some terrorist groups but not others. Moreover, the research suggests that governments, platforms and civil society representatives should work together to develop a global, unbiased and real-time database of possible terrorist entities.

A post-organisational far right?, *Hope Not Hate*, 2018

In this blog post, Dr. Joe Mulhall argues that we should "no longer measure the strength or likely influence of the [far-right] movement solely by how cohesive it is", noting that the "link between unity and impact is no longer as clear as it once was". The analysis links this development partially to the rise of social media platforms, noting how "far-right social media personalities who, despite not being part of traditional activist organisations or parties, have the ability to reach unprecedented numbers of people". This would allow people to actively engage in far-right politics, outside formalised organisational structures. While political activism for most of the post-1945 period had required "finding a party, joining, canvassing, knocking on doors, dishing out leaflets and attending meetings", social media today lowers the (social) costs of activism and enables a more international outlook".

Mulhall observes how a decentralised collective of far-right activists can act "completely anonymously without the danger and risk of being ostracised for doing so". The analysis concludes that the danger of a, even if appearing fractured and splintered, collective of anonymous online activists is the slow reshaping and radicalising of political agendas.

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.