Research Paper

Digital Society Initiative

January 2024

Towards a global approach to digital platform regulation

Preserving openness amid the push for internet sovereignty

Yasmin Afina, Marjorie Buchser, Alex Krasodomski, Jacqueline Rowe, Nikki Sun and Rowan Wilkinson

CHATHAM HOUSE

Chatham House, the Royal Institute of International Affairs, is a world-leading policy institute based in London. Our mission is to help governments and societies build a sustainably secure, prosperous and just world.

Contents

Summary

- After decades of reluctance, governments around the world are moving to regulate, and more actively direct, digital platforms in an effort to tackle perceived harms and to strengthen state oversight and control. Digital sovereignty is emerging as a critical goal, but the agenda is complicated by national security considerations, the influence of tech companies and domestic politics.
- There is significant diversity among countries in their approaches to platform regulation at present, with no clearly established norms or best practice. As yet, no one model is winning out. Neither legislation from Brussels, nor decisions made in Beijing, London or Washington are guaranteed to set the global agenda going forward. As such, there is a real risk of fragmentation becoming entrenched. Multilateral organizations are not currently providing sufficient leadership on the shape and execution of regulation at the international level.
- Whether the trend towards global divergence continues or moves towards convergence is a critical policy question. It is probable that while some convergence among like-minded nations will occur over time, further fragmentation is likely without the promotion of new approaches to global governance. A jurisdictional, fragmented internet could emerge as a result, with the world becoming a vast 'Venn diagram' of partially porous internets built around national languages, cultures and platforms, but accessible to all and controlled crudely.
- US technology provision remains the dominant force in shaping global norms, while EU regulation is its most influential check. Insofar as values-based lawmaking around digital platforms remains the primary way in which global regulatory efforts are made, the EU will continue to lead. But translating policy priorities and laws into technical standards presents its own unique challenge, and China outcompetes the EU in offering a 'full stack' of digital technologies, complete with standards and infrastructure, to developing countries seeking to digitize at pace.
- Despite being overlooked by some in the tech industry, human rights provide a well-established and compelling framework that could contribute to a global regulatory approach. The core principles and standards of human rights are universal and have long been institutionalized by the international community, meaning that regulatory regimes aligned to human rights norms would have force from the outset.

- Human rights are not a complete solution to the problem. Governments must also adopt other measures to help preserve an open, global internet, including tackling national divergence on platform regulation, increasing investment in international digital cooperation and securing the continued relevance and strength of existing institutions responsible for maintaining openness.
- To understand the divergence in global approaches to platform regulation, the authors reviewed 55 laws and proposals for legislation from around the world placing requirements on how platforms should moderate content as of October 2022. This research paper explores the results of that analysis, defining and detailing a set of global regulatory trends, including five notable approaches: strict custody; independent regulation; user rights and capacities; extensive platform monitoring; and data localization as part of content moderation regulation.
- The paper also outlines possible pathways for the future of platform regulation, including those of the major digital centres in the EU, China, the UK and the US. It discusses the issues around establishing global frameworks and the potential role for human rights. The paper concludes with recommendations on how policymakers can make progress towards alignment on platform regulation and preserve the open, global internet.

01 Introduction

A small number of platforms set global news agendas, culture and norms. But governance of those platforms is fragmented. Coordination and interoperability would strengthen states' ability to deal with corporate power, while also reducing compliance burdens on industry.

Over the past decade, large platforms like Amazon, Facebook, Google, Instagram, TikTok and WeChat have become a ubiquitous presence in daily life. As a result, these largely Chinese and US digital platforms are renegotiating the relationship between people and the world around them.

Governments have previously been slow on the uptake. The governance of digital platforms and services is now a central priority. An array of government bodies, technology firms and civil society actors have contributed to a patchwork of principles, laws and best practices that attempt to reflect the new primacy of digital technologies in shaping our lives. Those groups' attention has most frequently been focused around the twin poles of data protection and online harm.

National governments are motivated by a diverse set of ambitions in relation to platforms. For some, the hegemony of those platforms over their citizens' experience of the internet has challenged the social contract, opening a rift between citizen expectations and government capacity. For others, the spread of platforms has proved an unwelcome challenge to central power. Accordingly, national governance frameworks vary significantly and reflect the diversity of societal concerns, challenges and cultural and political approaches.

The US, for example, follows its free market and free speech traditions. Its hesitancy to regulate platforms has been a defining feature of their growth, with US-based platforms operating largely free of intermediary liability, able to set their own rules and taking minimal legal liability for what their users do or say. This has created tension as those platforms spread beyond the US, particularly

with European authorities that are pursuing a more vigilant, co-regulatory model with greater focus on balancing liberties, in contrast to the US focus on freedom of expression.[1]

In countries where platforms might be perceived to challenge state hegemony over information, regulatory and legislative responses have tended to be stricter. Governments like those of Nigeria and Singapore are increasingly enacting laws to exert greater state control over online space. Under such regimes, platforms can be required to proactively monitor and filter broadly defined categories of online content, make user data available to authorities indiscriminately and reduce user-level protections.

Fitting approaches to platform regulation into neat categories is an imperfect process. Paradoxically, countries with a poor track record on human rights have sometimes mandated platforms to carry out human rights audits, as seen in China. Elsewhere, comparatively liberal platform regulation may include employee liability or proactive content moderation requirements, as in New Zealand or India respectively. India in particular highlights the added difficulty of marrying a regulatory approach with its domestic use and application, and the strength of its oversight and democratic protections.

Striking a balance between the substantial benefits of an open internet and the push by countries to exercise their power online is the policy challenge for future platform regulation.

Whether one approach can or will win out over others, or whether diverse approaches can co-exist, remains to be seen. Efforts to find commonality across regulatory regimes – either from groups of countries or from international bodies like the OECD – are in their infancy. Outside of highly technical spaces such as standards-setting bodies, there is no single major international institution through which platform regulation is currently negotiated. The idea of harmonizing global regulatory approaches to the internet is controversial: the one-size-fits-all approaches that have defined the design of digital platforms to date have regularly failed to account for diverse local contexts, sometimes with catastrophic results. For instance, digital and social media platforms have been accused of high-profile failures in stewardship in Myanmar, Somalia and, most recently, during the Israel–Hamas conflict.

As well as having an integral role in underpinning global business, communication and community, for many people around the world, the web represents the most powerful tool for maintaining values such as freedom of expression and access to information, and for coordination on global challenges like climate change and sustainable development. But new digital jurisdictions have mapped poorly

1 This tension is particularly evident at present, with the upcoming entry into force of the EU Digital Services Act. The act marks a shift away from the model of the EU E-Commerce Act, which exempted platforms for intermediary liability.

onto existing political and legal institutions, creating significant new challenges for sovereign nations seeking to protect their citizens, enforce their laws and set the fundamental norms of the societies they govern.

While the internet and its benefits should not be equated with or reduced to a handful of large digital platforms, such platforms constitute the main – and, sometimes, only – entry point to the digital space for many users across the world. Existing and upcoming national regulations on platforms will therefore have a direct impact on citizens' experience and access. By extension, they will also partially define how open the global internet will be in the future. Striking a balance between the substantial benefits of an open internet on the one hand and the push by countries to exercise their power online on the other is the policy challenge for future platform regulation. Addressing technology governance, devising the appropriate policy and regulatory responses will require global cooperation. The internet could still form the basis of such cooperation. But a jurisdictional, fragmented internet threatens to undermine this promise at the time when it may be needed most.

This research paper – produced in partnership by Chatham House and Global Partners Digital – examines the divergent approaches to platform regulation to date. These approaches range from limited and independent regulation such as in Canada, to much firmer regimes aimed at preserving social order, like that in Belarus. Some approaches threaten companies with fines and others place liability with their directors. Some focus on the protection and promotion of civil liberties. Meanwhile, others look to empower users through technical tools, or gloss over the empowerment or protection of users entirely and lay out lists of illegal content for platforms to tackle.

The paper takes stock of where we are today, lays out where we might aim to get to tomorrow, and considers how we might measure the distance between the two.

02 Global regulatory trends

There is some international consensus around the idea that action is required to tackle the power of large digital platforms. But the diversity of approaches adopted so far threatens the future of the open, global internet.

There is no average regulatory regime. Countries cannot be easily grouped together according to other characteristics. For example, regional and linguistic groupings may obscure as much as they reveal, with significant differences in approach between groups or countries that share borders and languages. Despite this, there are some overarching trends and commonalities. Given digital platforms' global reach over multiple jurisdictions with competing or conflicting requirements, this chapter attempts to identify those trends and common features across a global dataset (compiled by Chatham House researchers) and to place them within a set of defined approaches. It then explores how these approaches diverge and interact.

Common requirements

Across the dataset, the most common features in regulations were enforcement through imposing fines on platforms (71 per cent) or threatening them with blocking of their services or blocking platforms altogether (51 per cent). In general, platforms were held accountable primarily for content classed as illegal (75 per cent) once they had been notified of its existence (76 per cent). There were also provisions for content that is not illegal but is seen, in some way, as harmful – such content ranges from disinformation to abuse online (51 per cent).

Table 1. Categorized survey questions gauging legislative approaches to platform regulation, per cent

	% yes	% no
Scope and governance		
Does the regulation require a multi-stakeholder approach to platform governance?	**33**	67
Does the regulation differentiate between types of digital platforms? (For example, between video streaming services and social network platforms?)	**29**	71
Is the regulation enforced by an independent authority?	**29**	71
Does the regulation differentiate between sizes of digital platforms? (For example, as measured by annual revenue or number of users or employees?)	**27**	73
Penalties and sanctions		
Does the regulation impose fines?	**71**	29
Does the regulation threaten platforms with restrictions or blocking for non-compliance?	**51**	49
Does the regulation threaten prison sentences for platform employees for non-compliance with content moderation requirements?	**22**	78
Content-based duties		
Does the regulation require platforms to remove prohibited content whenever it is notified of such content?	**76**	24
Does the regulation tackle content which is already designated as illegal under other legislation?	**75**	25
Does the regulation require platforms to remove or deal with content that is not illegal?	**51**	49
Does the regulation require platforms to remove prohibited content within a specific timeframe?	**46**	54
Does the regulation require platforms to proactively monitor for prohibited content?	**27**	73
Does the regulation require platforms to remove prohibited content when ordered to do so by a court?	**22**	78
Does the regulation designate new types of content as illegal?	**20**	80
Business-based duties		
Does the regulation require platforms to establish a local office or local contact?	**51**	49
Does the regulation require platforms to report regularly on the performance of their content moderation systems?	**40**	60
Does the regulation require platforms to register its services with authorities?	**26**	74

	% yes	% no
Does the regulation require platforms to carry out human rights risk assessments?	20	80
Does the regulation require platforms to submit to independent audit?	16	84
Does the regulation require platforms to store data locally?	7	93
Does the regulation require platforms to report regularly on advertising revenue?	4	96
Considerations for freedom of expression		
Does the regulation explicitly mention freedom of expression?	38	62
Are there limitations on the powers of regulators in line with freedom of expression safeguards?	31	69
Does the regulation reference platforms' responsibility to consider freedom of expression in their operations?	27	73
Are there regulatory exemptions for journalistic, scientific or public interest content?	20	80
Considerations for user capacities		
Does the regulation require platforms to implement complaints mechanisms?	49	51
Does the regulation require platforms to publish terms of service?	40	60
Does the regulation require platforms to notify users of ongoing complaints or appeals?	33	67
Does the regulation require platforms to implement appeals mechanisms?	27	73

Global approaches show greater coherence around identifying and dealing with sanctioned content than around tackling systems that might prevent or mitigate the spread of such content. On questions of transparency, risk assessments and audit, a few pieces of legislation tackle business practices and processes in the context of content moderation, though more holistic approaches such as the EU Digital Services Act (DSA) are buttressed by regulatory approaches to business practices, data privacy and antitrust that go beyond questions of content.[2] Nevertheless, sophisticated approaches to platform content regulation, like transparency reporting on advertising revenue – which is one of the critical drivers of the design and functioning of online spaces – have rarely been called for. Similarly, the absence of multi-stakeholder participation in regulatory consultation in two-thirds of the regulatory regimes in the dataset is a concern, and core questions remain about who to include and when, where to fold inclusion into policy processes, and how to organize these efforts with a balance of flexibility and fairness.[3]

2 See the EU Digital Markets and European AI Acts.
3 Chatham House Director's Office and International Law Programme (2021), *Reflections on building more inclusive global governance: Ten insights into emerging practice*, Synthesis Paper, London: Royal Institute of International Affairs, https://www.chathamhouse.org/2021/04/reflections-building-more-inclusive-global-governance/03-ten-insights-reflections-building.

Regulatory approaches

Grouping regulatory regimes simply by language or region fails to capture similarities between regulations in different parts of the world. For instance, national approaches to platform regulations in Europe vary widely from Belarus to France, or in South Asia from Bangladesh to Pakistan. Characterizing regulation by similarities in approach provides an alternative window.

By clustering regulations into five approaches based on similarities and differences across 29 analytical metrics (see Appendix 1), researchers were able to identify and explore core features or characteristics that define certain different regulatory approaches taken to platform regulations around the world. While these categories are informed by legal and data analysis, they are designed to be narrative and descriptive rather than exhaustive; and to be starting points for discussion about convergence and divergence.

Some regulatory landscapes can be seen as representing more than one of the five broad approaches identified, particularly when geographies have passed or proposed multiple regulatory regimes for online platforms. For example, Australia's Online Safety Act (and Basic Online Safety Expectations Determination) 2021 focuses on business and content duties, while its Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 centres on criminal charges for illegal content. Similarly, the EU's wide-ranging DSA and the UK's Online Safety Act 2023 meet the threshold for representation in both the *independent regulation* and *user rights and capacities* groups.

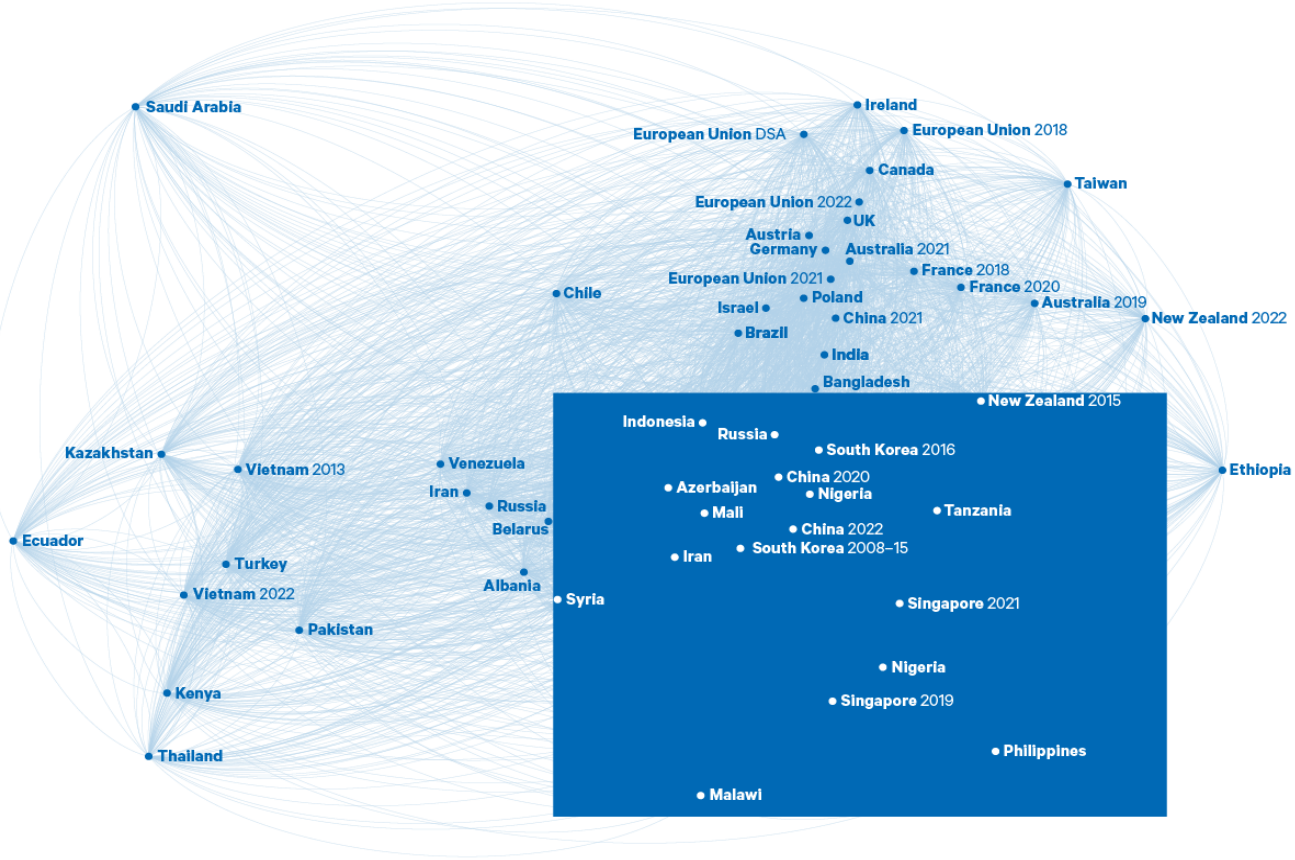
A summary of the five approaches and their main features is shown in Table 2.

Table 2. Approaches to platform regulation and main characteristics

Approach		Features
1	Strict custody	Prison sentences for non-compliance; legal but harmful content in scope; little proportional regulation
2	Independent regulation	Independent oversight; stronger emphasis on freedom of expression considerations
3	User rights and capacities	Transparency, redress and appeals processes mandated; few proactive monitoring requirements
4	Extensive platform monitoring	Extensive proactive monitoring requirements; legal but harmful content in scope
5	Data localization as part of content moderation regulation	Data localization requirements; powers to block access to platforms; no independent oversight

Approach 1: Strict custody

Figure 1. Cluster depicting countries following a *strict custody* approach



Of the 55 global regulations examined, 10 regimes exemplified the *strict custody* approach:

- Bangladesh's Draft Regulation for Digital, Social Media and OTT Platforms (2021);
- Malawi's Electronic Transactions and Cyber Security Act (2017);
- Mali's Law No. 2019-056 on the Suppression of Cybercrime (2019);
- New Zealand's Harmful Digital Communications Act (2015);
- Nigeria's Protection from Internet Falsehood and Manipulation Bill (2019) and Draft Code of Practice for Internet Intermediaries (2022);
- The Philippines' Anti-False Content Bill (2019);
- Singapore's Foreign Interference (Counter-measures) Act (2021);
- South Korea's Act No. 14080 on Promotion of Information and Communications Network Utilization and Information Protection (2016);
- Syria's Law on Cybercrime (2022); and
- Tanzania's Electronic and Postal Communications (Online Content) Regulations (2020).

Regulations in this group were categorized by:

- Prison sentences for platform employees for non-compliance with content moderation requirements or orders (10/10 regimes);
- Platforms being required to remove or deal with content that is not illegal (8/10 regimes);
- No distinction in the regulation between types of online platform (10/10 regimes); and
- No requirements on platforms to report on advertising revenue, submit to independent audit, localize data or implement appeals mechanisms (10/10 regimes).

Regulations in this group diverge from global trends by imposing potential custodial sentences for platform employees as a result of content moderation failings. Malian regulations, for instance, threaten local employees of non-compliant service providers with up to two years in prison, while the Singaporean Foreign Interference (Counter-Measures) Act threatens prison sentences of up to four years.[4,5] Incarceration of individual employees for platform shortcomings threatens in-country platform operations, and regulations of this kind are unpopular with companies.

Eight of the 10 regulations in this group require platforms to remove or otherwise address content that is not necessarily illegal, but is defined as, in some way, harmful. This broad category is difficult to define and varies across jurisdictions depending on a government's policy goals. Content deemed to be harmful can range from pornography or violence to misinformation or anti-government messages. The Nigerian Draft Code of Practice for Internet Intermediaries, for instance, aims to prevent the transmission of 'false statements/declaration of facts'. South Korean regulations prohibit 'information that infringes on the rights of others', while Tanzanian law prohibits the 'ridicule, abuse or harming the reputation' of the country or its flag.[6,7,8]

Prohibitions of 'legal but harmful' content are controversial. Without clear legal guidelines or definitions, they present challenges to users and industry in deciding whether content or behaviour falls foul of the regulation. By incentivizing and/or requiring platforms to remove legal content, states risk implementing restrictions on freedom of expression that do not satisfy the 'tripartite test' of legality, legitimacy, and necessity and proportionality (see chapter 4) under international human rights law (IHRL) and guidance.

4 Law No 2019-056 on the Suppression of Cybercrime, 24–27.
5 Foreign Interference (Counter-measures) Act (FICA), 45 (3a, 4a).
6 The Protection from Internet Falsehood and Manipulation Bill 2019, 1, 16(a).
7 Act on Promotion of Information and Communications Network Utilisation and Information Protection (Act No. 14080, Mar. 22, 2016), 44(1).
8 Tanzanian Electronic and Postal Communications (Online Content) Regulations, 2020, and their Amendment Regulations 2022, Third schedule (under reg 16).

Platform regulations in the dataset that follow a *strict custody* approach do not differentiate between types of online platforms and services. This lack of differentiation hampers proportional regulation, and likely favours larger platforms with the resources to comply. Absence of independent auditing and appeals processes suggests that these regimes prioritize state control over online spaces.

It is notable that this regulatory approach did not correlate with requirements mandating data localization, pointing to states looking to boost their legal leverage over online platforms without resorting to expensive and skill-intensive technical leverage built on accessing local data.

Overall, business conditions for platforms under *strict custody* regimes are restrictive. For example, the risk of harsh punishments for individual employees tends to be viewed by businesses as regulatory overreach, and threats to individuals working at technology companies have even been described as 'hostage-taking laws', following controversy over their use in Brazil, India and Russia over the past decade.[9]

Industry groups have also raised the risks of directorial criminal liability for failure to comply with content moderation requirements. Civil society organizations, meanwhile, warn that these types of sanctions could lead to overzealous removal of content by risk-averse decision-makers.[10] This risk is further exacerbated by requirements to remove content that is legal where that content might be deemed harmful in the local context.

Strict custody regimes are geared towards strengthening state or judicial power over online platforms.

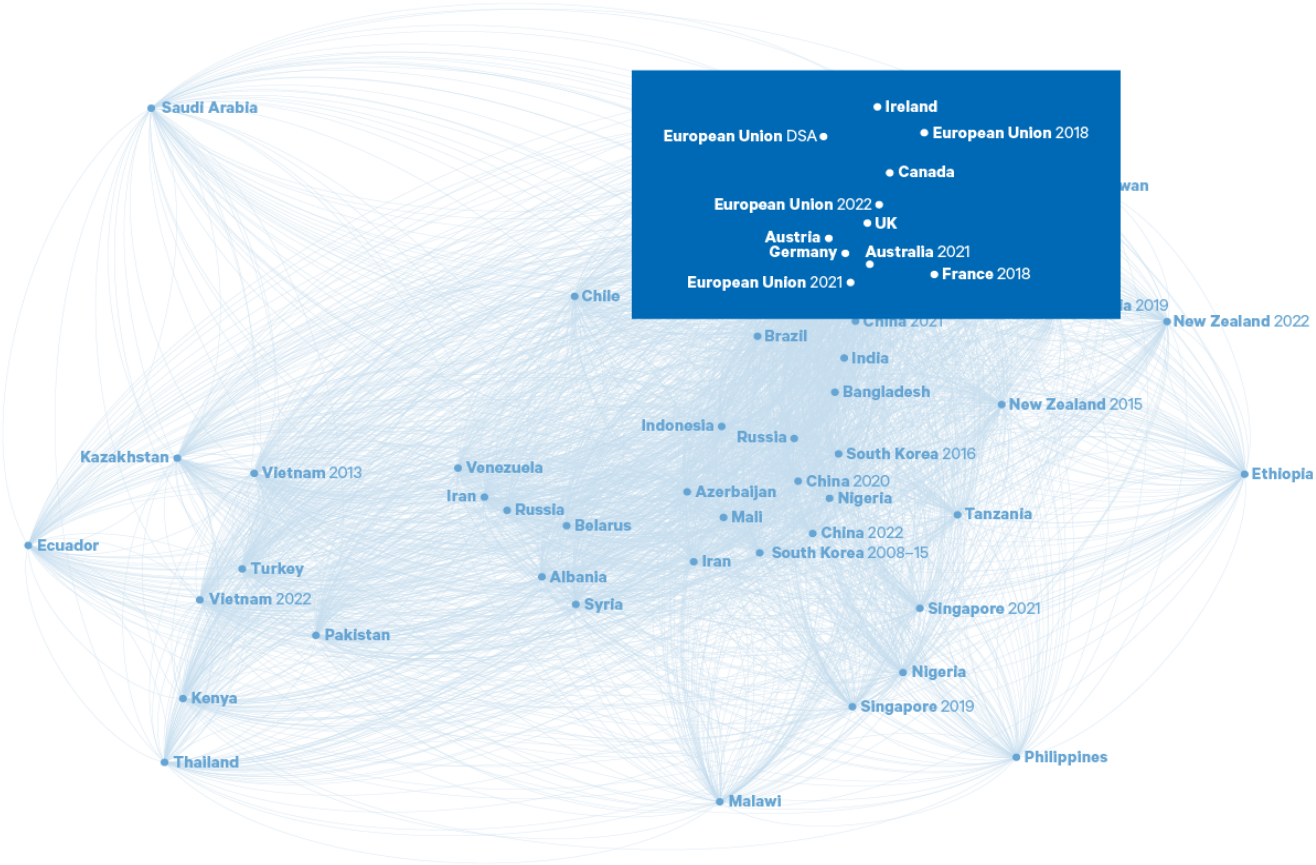
Regulations assigned to the *strict custody* regimes group tend to put less emphasis on systemic change at a platform level, and tend not to build in significant user protections or routes towards additional platform oversight. For the most part, the regulations implement neither independent audits of platforms nor appeals or transparency mechanisms for users. There is some variation: regulations in Bangladesh, Nigeria and South Korea do call for transparency in terms of service and around platform decisions, while regulations in The Philippines, Singapore and Syria make no mention of protections for users or systemic platform change.

9 Elliot, V. (2021), 'New laws requiring social media platforms to hire local staff could endanger employees', Rest of World, 14 May 2021, https://restofworld.org/2021/social-media-laws-twitter-facebook.
10 Burns, H. (2021), 'Online abuse: Why management liability isn't the answer', Open Rights Group, 5 May 2021, https://www.openrightsgroup.org/blog/online-abuse-why-management-liability-isnt-the-answer; Keller, D. (2021), 'Empirical Evidence of Over-Removal By Internet Companies Under Intermediary Liability Laws: An Updated List', Stanford Law School Center for Internet and Society, 8 February 2021, https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws.

Strict custody regimes are geared towards strengthening state or judicial power over online platforms. Civil society and individual experiences of platforms under *strict custody* regimes should be expected to reflect this emphasis, with reduced freedom to challenge state-aligned narratives, and reduced access to material deemed by the national government to threaten social order.

Approach 2: Independent regulation

Figure 2. Cluster depicting countries following an *independent regulation* approach



Of the 55 global regulations examined, nine regimes exemplified the *independent regulation* approach:

- Australia's Online Safety Act (2021); Online Safety (Basic Online Safety Expectations) Determination (2022); and Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act (2019);
- Canada's Proposed Approach to Online Safety (Discussion Guide and Technical Paper) (2021);

- The EU's Proposal for a regulation of the European parliament and of the council for laying down rules to prevent and combat child sexual abuse (2022); Digital Services Act (2022); and Directive 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU (Audiovisual Media Services Directive);
- Ireland's Online Safety and Media Regulation Bill (2022);
- New Zealand's Harmful Digital Communications Act (2015); and
- The UK's Online Safety Bill (now the Online Safety Act; 2023).

Regulations in this approach were categorized by:

- Regulation enforced by an independent regulatory authority (9/9 regimes);[11]
- Explicit mention of freedom of expression (8/9 regimes) and limitations on regulators' enforcement powers in line with freedom of expression safeguards (9/9 regimes);
- Proportional or differing regulation between types of online platform (7/9 regimes) backed by fines (8/9 regimes);
- Multi-stakeholder input into the regulatory process (6/9 regimes); and
- Legal content that could be damaging to social order being in scope (6/9 regimes), but including neither proactive monitoring requirements (2/9 regimes) nor data localization requirements (0/9 regimes)

Regulatory regimes following this approach were marked by distance placed between the government and private interests and the regulator, and by limits to the power of the regulator on the grounds of freedom of expression. Provisions for independent regulators are mainly concentrated among liberal democracies – for example, Austria, France and the UK all propose an independent regulator – although the approach of separating regulatory supervision of platforms from government has also been followed by countries taking a different approach to regulation like Kenya, Malawi and Tanzania.

Provisions that limit regulatory enforcement in line with freedom of expression safeguards are often written into these types of regulation. For example, the Australian eSafety Commissioner's powers are limited if their enforcement 'would infringe any constitutional doctrine of implied freedom of political communication'.[12] Ireland's Online Safety and Media Regulation Bill includes provisions for platforms to notify the regulator if, in their view, a regulation excessively infringes on their users' freedom of expression.[13]

The commitment to multi-stakeholder input further strengthens the levels of societal, industrial and third-sector input into regulatory decision-making. For instance, the inclusion of multi-stakeholder consultations in the policy development process likely makes the DSA function on a measured and

11 OECD (2019), 'Independent sector regulators and competition', 2 December 2019, https://www.oecd.org/daf/competition/independent-sector-regulators.htm.
12 16/233 and 474.38, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019.
13 Online Safety and Media Regulation Bill, Part 12 (46: *139ZO(3, 4), 139ZU).

compromise-based approach, boosting regulatory innovation and reducing ecosystem disruption.[14] At a national level, French law calls for collaboration between platforms and news agencies, publishers and journalists to tackle disinformation.[15]

Under an *independent regulation* approach, powers and regulations tend to differentiate between different types of platform – i.e. a social media platform will carry different responsibilities to those of a search engine.[16]

Regulations assigned to the *independent regulation* approach tend to be those proposed by liberal democracies. Such regulatory regimes try to achieve a balance between increased state authority over platforms and a reduction in harmful content on the one hand, and commitments to preserving liberties on the other.[17]

Regulations assigned to the *independent regulation* approach try to achieve a balance between increased state authority over platforms and a reduction in harmful content on the one hand, and commitments to preserving liberties on the other.

Regulatory regimes under this approach are not absolute: they are limited in autonomy, power and scope. Although the fines proposed by these regimes are significant – the DSA alone can fine a platform up to 6 per cent of global revenue – *independent regulation* regimes do not go as far as demanding costly compliance requirements, such as the proactive monitoring of speech or user data localization.[18]

Nevertheless, they are given a significant remit. Most go beyond sanctioning only illegal content and demand platforms tackle a diverse range of harmful material or behaviour, with greater platform latitude around how precisely that content or behaviour will be tackled.

For businesses, this may prove a compliance challenge. For the most part, industry actors will feel listened to under *independent regulation* regimes and protected from heavily prescriptive regulation on the grounds of freedom of expression. Co-regulatory models recognize the evolving nature of digital ecosystems and the complex responsibilities of the companies concerned.[19]

However, the issue of 'legal but harmful' content or behaviour has dogged the debate on digital regulation, despite repeated international guidance – including interventions from the UN Special Rapporteur on the Protection and Promotion

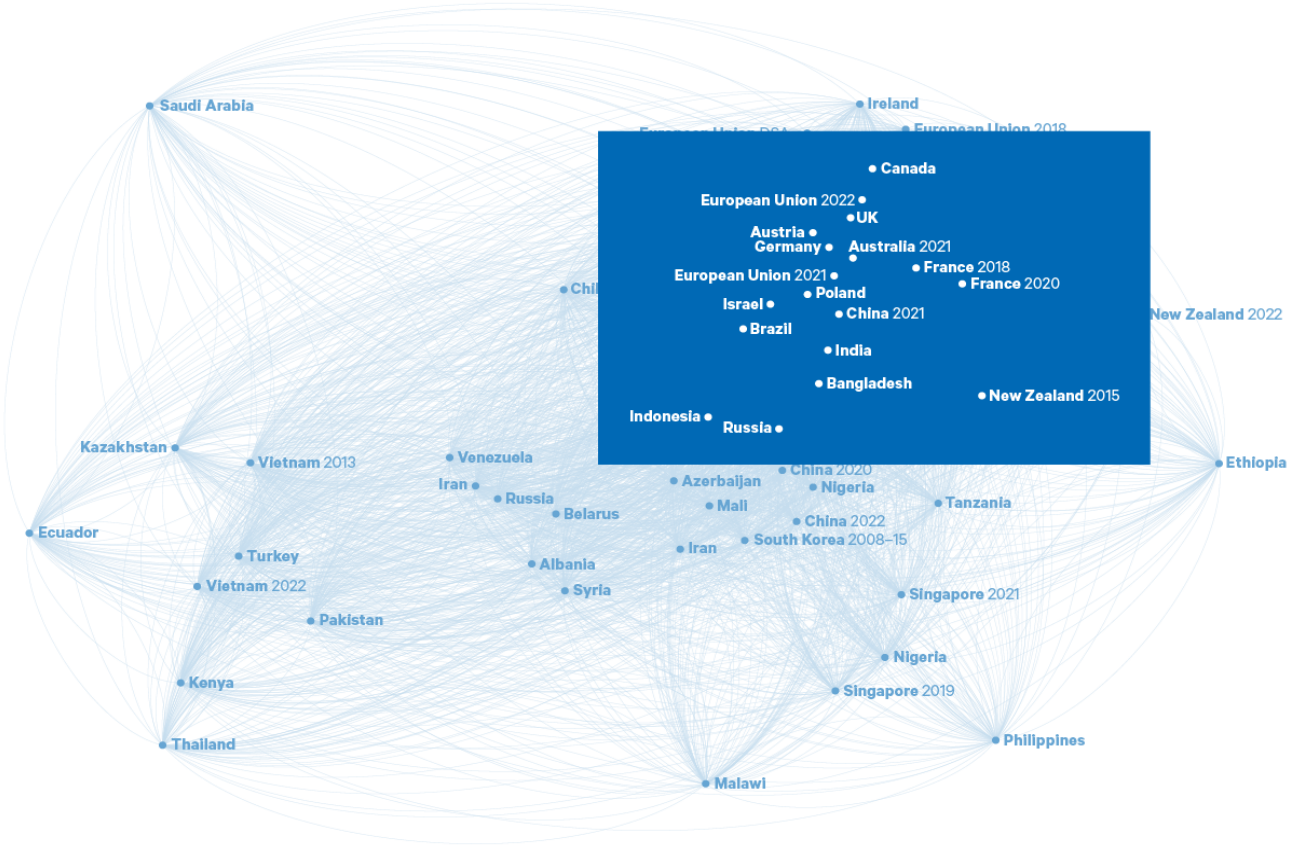
---

14 Morar, D. and Santos, B. (2022), 'Is the DSA a New Dawn of Legislating Platform Governance Globally?', Lawfare, 30 November 2022, https://www.lawfareblog.com/dsa-new-dawn-legislating-platform-governance-globally.
15 Law n° 2018-1202 of 22 December 2018 relating to the fight against the manipulation of information, §14.
16 Online Safety Bill, 183-186, Sched. 1, 11.
17 Reisman, R. (2023), 'From Freedom of Speech and Reach to Freedom of Expression and Impression', Tech Policy Press, 14 February 2023, https://techpolicy.press/from-freedom-of-speech-and-reach-to-freedom-of-expression-and-impression.
18 EU Digital Services Act, articles 42, 59 and 60.
19 Buchser, M. and Moynihan, H. (2021), 'Can global technology governance anticipate the future?', Chatham House Expert Comment, 27 April 2021, https://www.chathamhouse.org/2021/04/can-global-technology-governance-anticipate-future.

of Freedom of Expression and the UN Human Rights Committee – that states should not force companies to remove speech that is not explicitly illegal.[20,21] Some regulators have now dropped such requirements, but those persisting will require clarity on a regulator's expectations to avoid a significant burden to platforms in scope.[22]

Approach 3: User rights and capacities

Figure 3. Cluster depicting countries following a *user rights and capacities* approach



The most common approach, *user rights and capacities*, contained 17 of the 55 regulatory regimes:

- Austria's Communication Platforms Act (2020);
- Bangladesh's Draft Regulation for Digital, Social Media and OTT Platforms (2021);

20 UN Office of the High Commissioner for Human Rights (2018), *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, 6 April 2018, https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/38/35.
21 UN International Covenant on Civil and Political Rights Human Rights Committee (2011), *General comment No. 34: Article 19: Freedoms of opinion and expression*, 12 September 2011, https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.
22 MacCarthy, M. (2022), 'U.K. government purges "legal but harmful" provisions from its revised Online Safety Bill', blog, Brookings Institution, https://www.brookings.edu/blog/techtank/2022/12/21/u-k-government-purges-legal-but-harmful-provisions-from-its-revised-online-safety-bill.

- Brazil's Draft Bill 2630 on Freedom, Responsibility and Transparency on the Internet ('The Fake News Law') (2020);
- Canada's Proposed Approach to Online Safety (Discussion Guide and Technical Paper) (2021);
- The EU's Digital Services Act (2022); Proposal for a regulation of the European parliament and of the council for laying down rules to prevent and combat child sexual abuse (2022); and Regulation 2021/784 on addressing the dissemination of terrorist content online (2021);
- France's Law n° 2021-1109 consolidating respect for the principles of the Republic (2019);
- Germany's Network Enforcement Act (NetzDG) (2017);
- India's Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules ('IT Rules') (2021);
- Indonesia's Minister of Communication and Informatics Regulation No. 5 of 2020 on Private Electronic System Operators (2020);
- Israel's Social Networks Bill (2021);
- Nigeria's Draft Code of Practice for Internet Intermediaries (2022);
- Poland's Law on the Protection of Freedom of Speech on Social Networking Sites (2021);
- Russia's Federal Law No. 149-FZ on Information, Information Technologies and Protection of Information (2006);
- Taiwan's Draft Digital Communication Law (2022); and
- The UK's Online Safety Bill (now Online Safety Act; 2023).

Regulations in this group were broadly categorized by:

- A focus on empowering platform users, with mandates on transparency, redress and appeals (at least three processes mandated; 17/17 regimes);
- Provisions for illegal content to be removed on notification (15/17 regimes), rather than requirements for proactive monitoring (5/17 regimes);
- Sanctions limited to fines for platforms (16/17 regimes) over prison sentences (1/17 regimes); and
- Requirements for local contacts (13/17 regimes), rather than data localization (1/17).

This category covers a diverse set of regulatory approaches brought together by their mandating processes and mechanisms aimed at strengthening the tools available to users and civil society when interacting with platforms.

The study checked legislation for four commonly used tools: appeals mechanisms, transparency reporting, complaints procedures and terms of service. A requirement for platforms to implement a complaints mechanism was most common among the regimes analysed, with one-half of the regimes mandating this. Germany's NetzDG,

for instance, requires platforms to implement a means for users to report illegal content. French regulation mandates the same, noting the importance of users being able to report content promoted on behalf of a third party.

Other requirements included publishing terms of service (42 per cent of regimes), and providing appeals mechanisms to allow users to challenge platform removals (27 per cent) and be kept up to date on those challenges and other complaints (33 per cent).

Content monitoring requirements under *user rights and capacities* regimes are usually lighter than under the other approaches identified. The focus is on illegal content and its removal by platforms after notification, usually referred to as 'notice and takedown' regimes. Only a minority of such regimes demand proactive monitoring by platforms. Infractions are punishable by fines, with limited liability for individual employees.

A *user rights and capacities* approach puts the onus on the platform to improve, while being comparatively less prescriptive about how those improvements are made.

User rights and capacities regimes further impose a number of business duties. Transparency reporting, for instance, is a requirement in 88 per cent of such regimes (compared with 40 per cent across all the regimes reviewed for this study). A local office or point of contact also tends to be mandated.

In many respects, *user rights and capacities* regulation is rooted in a tradition of encouraging self-regulation, through setting targets but leaving execution to platforms outside of a handful of serious offences. Over time, legislatures have increasingly viewed platform improvements as necessary but not sufficient. This has contributed to the rise of independent regulation and stricter regulation of sanctioned content, but alongside these laws, regimes continue to mandate changes to platform design and function. In short, a *user rights and capacities* approach puts the onus on the platform to improve, while being comparatively less prescriptive about how those improvements are made.

Mandating the creation and maintenance of tools that improve platform users' experience of a platform contributes to industry innovation in meeting regulatory requirements. This kind of principle-based approach – provided it can be shown to be effective in meeting regulatory objectives – reduces burdens on both the regulator and the regulated platforms. These tools – often referred to as 'middleware' – could sit independently of major platforms.[23]

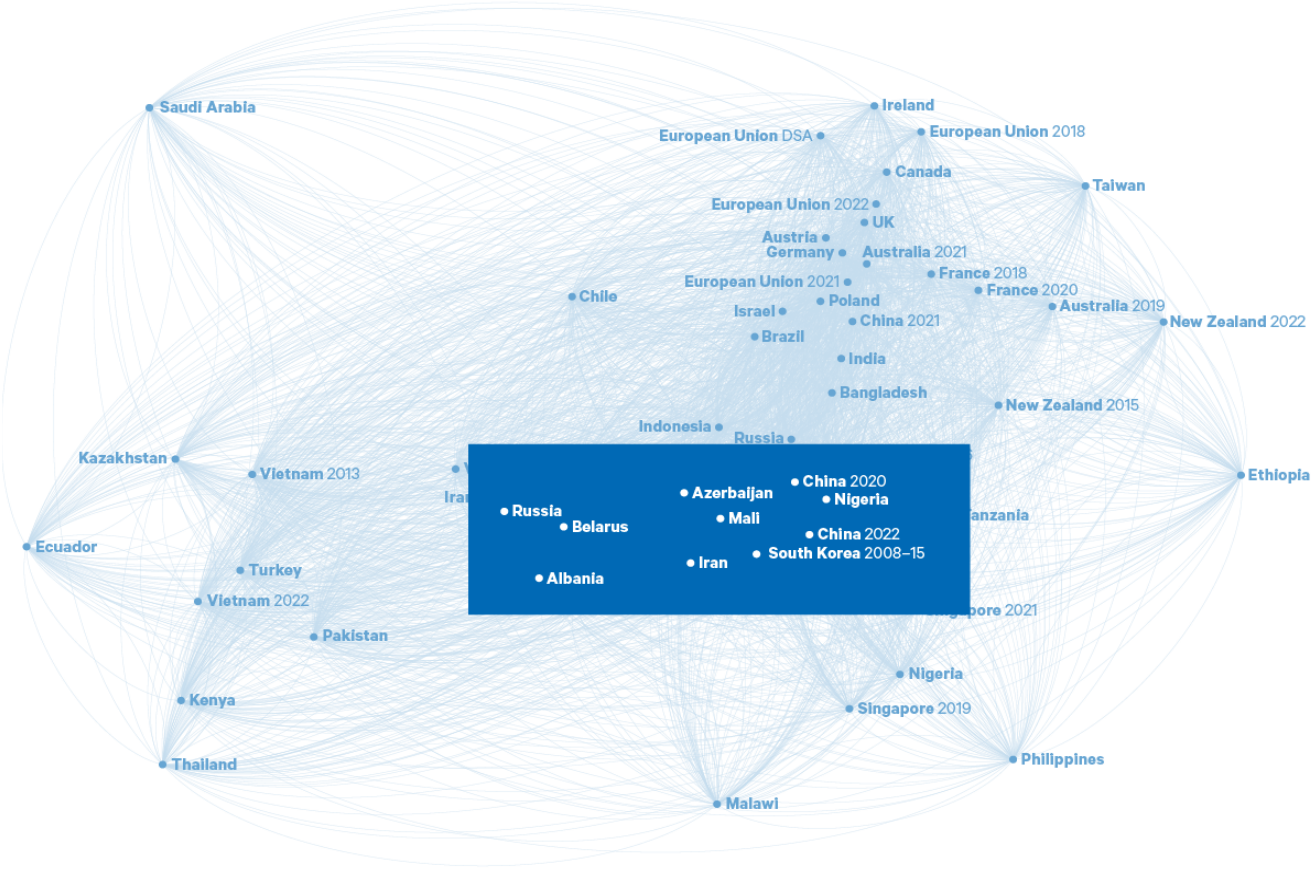
23 Fukuyama, F. et al. (undated), *Middleware for Dominant Digital Platforms, A Technological Solution to a Threat to Democracy*, Stanford, CA: Stanford University Cyber Policy Center, https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/cpc-middleware_ff_v2.pdf (accessed 19 May 2023).

A risk with non-prescriptive regulation comes in delegating definitions around the design of critical compliance tools to the regulated platforms. An advertising transparency database, or a platform application programming interface (API) – to give two examples – should be designed to standards agreed on by a wider range of stakeholders than just the platforms themselves. Provision of routes to input by a multi-stakeholder audience is critical, and platforms and regulators must be open to working with the third sector and academics on assessing and establishing best practice. There is an added risk that best practice comes to be defined by the biggest platforms, whose resources to develop and deploy solutions cannot be matched by smaller competitors. In the absence of an open solution and increased platform collaboration, smaller platforms may find themselves forced out of the market. Tackling this power imbalance through government software cooperation may be a viable path forward.[24]

Although rarely explicit in the legislation, a focus on processes and tools that allow users, communities and platforms to wield power and balance rights against one another is the approach truest to a human-rights based framework. Empowering individual users through dedicated processes and tooling is therefore likely the closest analogue to a human-rights based approach to platform regulation (see chapter 4).

Approach 4: Extensive platform monitoring

Figure 4. Cluster depicting countries following an *extensive platform monitoring* approach



24 Riley, C. and Ness, S. (2022), 'Modularity for International Internet Governance', Lawfare, 19 July 2022, https://www.lawfareblog.com/modularity-international-internet-governance.

Five regimes from four countries exemplified an approach to content moderation emphasizing enhanced monitoring by platforms:

- Belarus' Law of the Republic of Belarus No. 128-Z on amendments and additions to some laws of the Republic of Belarus (2018);
- China's Provisions on the Governance of the Online Information Content Ecosystem (2019) and Draft Regulations on the Protection of Minors on the Internet (2022);
- Nigeria's Draft Code of Practice for Internet Intermediaries (2022); and
- Russia's Federal Law No. 149-FZ on Information, Information Technologies and Protection of Information (2006).

Regulations in this group were categorized by:

- Proactive content moderation requirements (5/5 regimes);
- Mandating regular reports on content moderation systems (4/5 regimes);
- The inclusion in scope of legal content that could be damaging to social order (5/5 regimes); and
- Sanctions including blocking and restricting access to content or platforms (4/5 regimes), but not extending to prison sentences (1/5 regimes).

Regulations grouped under *platform monitoring* were defined by looser definitions of sanctioned content (for instance, legal content that could be damaging to social order). Belarusian law, for instance, imposes sanctions against the sharing of 'materials containing obscene words and expressions' in addition to illegal content.[25] Nigerian laws concerning the use of automated accounts, meanwhile, sanction content that may 'be prejudicial to public health, public safety, public tranquillity or public finances' or 'diminish public confidence in the performance of any duty or function of, or in the exercise of power by the Government'.[26] Such definitions reduce clarity to end users and platforms as to the limits of acceptable content.

Nevertheless, these looser definitions are backed by stricter monitoring requirements and punishments for offending organizations. Most notably, all five regimes mandate proactive content moderation, rather than liability only after notification of behaviour or content that is in breach of the regulations.[27] Proactive content moderation places expectations on platforms to remove sanctioned content as soon as, or before, it is posted. Most regulatory regimes have clauses for proactive content moderation, but these clauses are largely restricted to illegal content such as that related to child sexual abuse material (CSAM), terrorism and copyrighted materials. Such requirements are inconsistent with IHRL, even when only applied to CSAM or terrorism-related content.[28]

---

25 Law of 17 July 2018 No. 128-Z, 30-1(2.2).
26 The Protection from Internet Falsehood and Manipulation Bill 2019, C770, ii, vi.
27 Multiple Chinese laws require proactive content moderation. These include Provisions on the Governance of the Online Information Content Ecosystem, 10; and Draft regulations on the Protection of Minors on the Internet, 20(5), 26, 30.
28 UN Office of the High Commissioner for Human Rights (2018), *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*.

This approach is further characterized by comparatively far-reaching punishments for platforms failing to comply. For instance, Russian law provides for administrative, civil or criminal liability.[29] In four of the five regimes surveyed, regulators are further empowered to block offending platforms and providers.

The combination of penalties for legal content deemed harmful to individuals or to social order and platform liability beginning with when content is posted, rather than when a platform is notified of it, makes *platform monitoring* regimes troubling from both a regulatory compliance standpoint and a civil liberties perspective. Vague categories of prohibited content contradict international standards on freedom of expression, under which any restrictions on speech must be clearly provided for in law, in pursuit of a legitimate aim, and proportionate and necessary to the achievement of the the stated aim. Such categories make it difficult for technology companies to arbitrate the types of speech allowed and prohibited, and translate poorly to automated content identification systems. Moreover, broad criminal provisions are routinely cited as tools used to suppress freedom of expression.[30]

With sanctions extending to the blocking of access to platforms, user experience under *platform monitoring* regimes is characterized by opacity.

Extending requirements to the more diffuse category of 'legal but harmful' content creates a heavy technological burden on platforms as automated techniques centring on language and image recognition become required. The lines around permissible content are blurred, as these automated technologies can struggle to identify this kind of content accurately. Coupled with risk-averse enforcement, large-scale removal of speech that should be legitimate under IHRL and guidance on freedom of expression becomes more likely. Such regimes incentivize the platform to remove in case of doubt, rather than run the risk of violating regulation, particularly when regulation mandates individual employee liability.

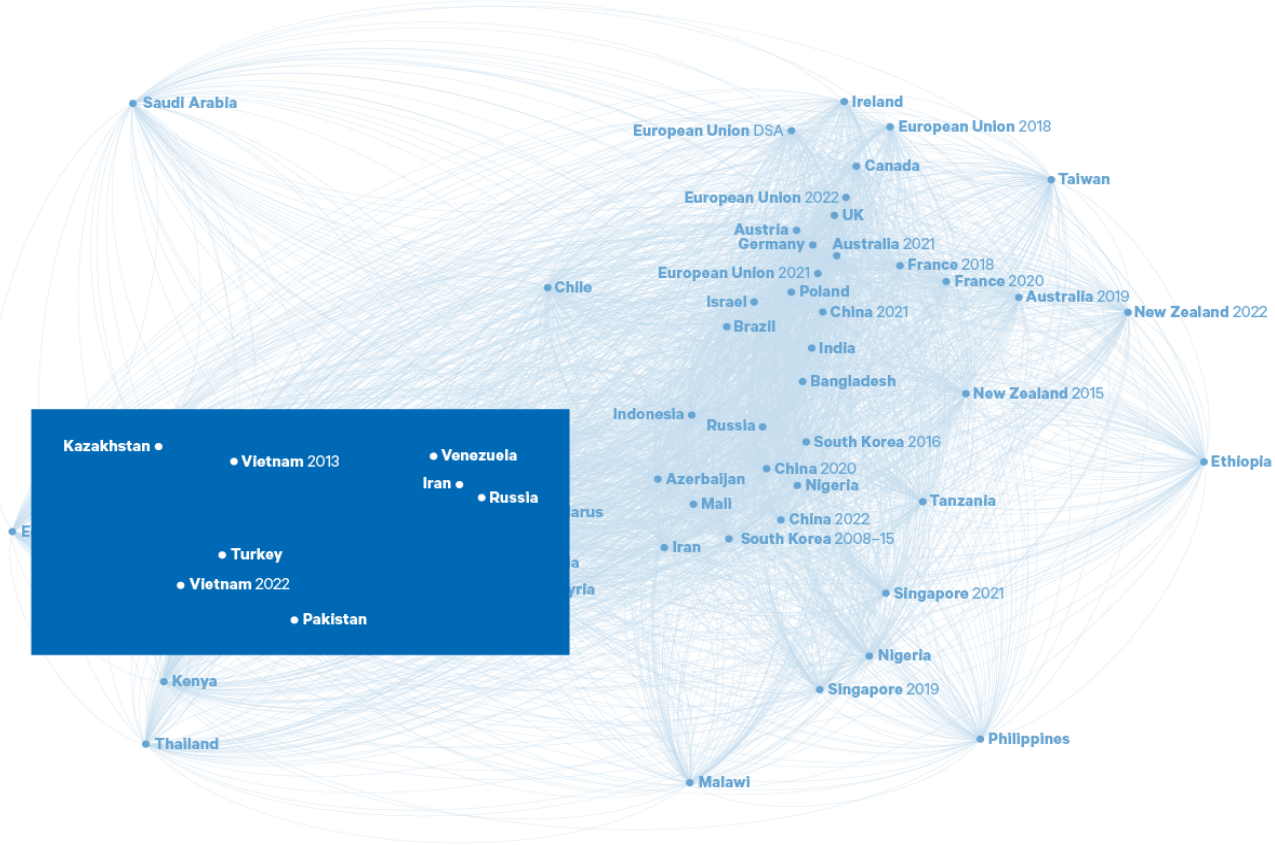
With sanctions extending to the blocking of access to platforms, user experience under *platform monitoring* regimes is characterized by opacity. Decisions made at both state and platform level are unlikely to be clearly explained to a user, particularly those around the removal of content or the suspension of user access. Only one of the regulations in this group – China's Draft Guidelines for Implementing Subject Responsibilities on Internet Platforms – mandates independent audits for 'super-large' platforms.[31]

---

29 Federal Law of 27 July 2006 No. 149-FZ 'On Information, Information Technologies and Protection of Information', 10.4(13), 10.5(17).
30 See, for example, Amnesty International (undated), 'Freedom of Expression', https://www.amnesty.org/en/what-we-do/freedom-of-expression.
31 cqn.com.cn (2021), 'Guidelines for implementing subject responsibilities on internet platforms (Draft for comments)', 29 October 2021, https://www.cqn.com.cn/zj/content/2021-10/29/content_8747098.htm.

Approach 5: Data localization as part of content moderation regulation

Figure 5. Cluster depicting countries following a *data localization as part of content moderation regulation* approach



Five regimes from four countries were distinguished by explicitly referring to data localization as part of their approach to content moderation:

- Pakistan's Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules (2021);
- Russia's Federal Law No. 149-FZ on Information, Information Technologies and Protection of Information (2006);
- Turkey's Law on the Regulation of Publications Made in the Internet Environment and Combating Crimes Committed through these Publications (2007);
- Vietnam's Decree No. 72/2013/ND-CP on the management, provision and use of Internet services and online information (2013); and
- Vietnam's Law on Cybersecurity (2018; including Decree 53/2022 Elaborating a Number of Articles of the Law on Cybersecurity of Vietnam).

Regulations in this group were categorized by:

- Data localization requirements (5/5 regimes);[32]
- Requirements to remove illegal content on notification (4/5 regimes), rather than requirements for proactive monitoring (1/5 regimes);
- Few (5/5 regimes) or no (3/5 regimes) considerations of freedom of expression;
- Sanctions including the blocking and restriction of access to content or platforms (4/5 regimes), but not extending to prison sentences (0/5 regimes); and
- An absence of independent regulator or audit requirements (0/5 regimes).

The relatively small number of countries referring to data localization in their content moderation regulation may be misleading, as many others are beginning to mandate data localization in wider reforms and legislation targeting the digital economy. These countries include Brazil, China, Nigeria and Russia.[33,34] Although it does not mandate data localization, the EU's data protection legislation places certain restrictions and conditions on the transfer of data, while other countries have mandated data localization for certain types of data, including financial or medical.[35]

This approach to content regulation generally requires platforms to store, and likely provide access to, data on territory over which a state has legislative authority.[36] For example, an October 2020 amendment to Turkey's Regulation of Internet Broadcasts and Prevention of Crimes Committed through Such Broadcasts (Law No. 5651) requires domestic or foreign social network providers to store user data in Turkey.[37]

Over the past decade, platforms – most of them based in the US – have had significant discretion in managing government data-access requests. For countries looking to manage speech or behaviour more closely, US-based data storage impedes their attempts to identify users responsible for infringements.

The five regimes demanding data localization in the context of content regulation provide few protections for freedom of expression in their legislation, and have been proposed by countries where protections for speech and other human rights are limited. The legislation often coincides with requirements for companies to hand over user data or identify users to state authorities on request.

---

32 Note that this is not an exhaustive list of laws including data localization requirements, which includes a range of other laws such as cybersecurity and data protection laws that are not included in the database and this paper because they do not also include content moderation requirements. In these five specific cases, the pairing of data localizations requirements with content moderation requirements is particularly of interest, given the opportunities for enforcement of content regulations and forcing platforms to hand over data of non-compliant users.
33 Federal Law No 242-FZ, part 5, article 18.
34 Article 37, Cybersecurity Law of the Peoples' Republic of China, 2017.
35 Pfeifele, S. (2017), 'Is the GDPR a data localization law?', IAPP, 29 September 2017, https://iapp.org/news/a/is-the-gdpr-a-data-localization-law.
36 While there is no universally accepted definition of what data localization means, as well as on its scope and overall legal and administrative force, the EU has attempted to provide a definition. See Article 3(5), EU Regulation 2018/1807: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R1807. For another approach, see also Svantesson, D. (2020), 'Data localisation trends and challenges: Considerations for the review of the Privacy Guidelines', *OECD Digital Economy Papers*, 22 December 2020, https://doi.org/10.1787/20716826.
37 Law on the Regulation of Broadcasts via Internet and Prevention of Crimes Committed through Such Broadcasts, 4 (6).

Taken together, the proposed regulations may represent an attempt by a state to significantly increase control over the internet inside its borders. Notably, none of the regimes grouped under this approach provide for an independent regulator, with the government taking responsibility for enforcement.

Governments may choose to adopt and enforce data localization requirements for several reasons. The main reason is to enable 'easy' access to data by national law enforcement and/or security agencies. When government agencies require access to data that may be hosted in another country outside of their jurisdiction, not only do they require cooperation from the company hosting the data, but also from the government of the territory where that data is hosted. From a law enforcement viewpoint, data localization measures help circumvent such obstacles by ensuring that data (or at least, a copy of the data) is in a certain territory/jurisdiction.[38]

This approach to content regulation generally requires platforms to store, and likely provide access to, data on territory over which a state has legislative authority.

Beyond the issue of access, data stored within a given country's territory would, in principle, be subject to that country's jurisdiction, and thus, laws, regulations and policies. Depending on the processes in place in that country's jurisdiction and the overall legal landscape surrounding data access, the government would be in a better position to apply measures over any data located in its territory.

Yet data localization requirements have implications and give rise to concerns. Mandating the installation or use of hardware inside a country's borders may prove a step too far for all but the largest digital platforms or services in light of the financial and operational implications.[39] When faced with demands for compliance, it is probable that smaller platforms will seek to end their operations and provision of services to users in that territory. If fully enforced, users may see a reduction in service availability and find online services dominated by platforms capable of meeting these regulatory requirements, occasionally punctuated by other platforms and services too small to catch the regulators' attention. These trends combined would likely mean that users are significantly limited in their experience of the open, global internet.[40]

In addition, data localization requirements would mandate companies to store data in physical systems geographically located in the territory of countries putting in place such measures; hence, instead of centralizing all data in a single location, they will have to acquire (or rent) and 'maintain servers in each of these countries

---

38 Cory, N. and Dascoli, L. (2019), 'How Barriers to Cross-Border Data Flows Are Spreading Globally, What They Cost, and How to Address Them', Information Technology & Innovation Foundation, 19 July 2019, https://itif.org/publications/2021/07/19/how-barriers-cross-border-data-flows-are-spreading-globally-what-they-cost.
39 Ibid.
40 Internet Society (2020), 'Internet Way of Networking Use Case: Data Localization: How mandatory data localization impacts the Internet Way of Networking', 30 September 2020, https://www.internetsociety.org/resources/doc/2020/internet-impact-assessment-toolkit/use-case-data-localization.

in which they do business'.[41] Proliferation of data centres would have a considerable environmental impact, due to the compute powers required and subsequent energy consumption, as well as their greenhouse gas emissions and waste.[42]

Data localization carries further risks. Centralizing digital infrastructure inside a country's borders puts that infrastructure at risk should that country come under attack. 'Data embassies' located abroad, such as those employed by Estonia,[43] highlight the advantages of securing a country's data beyond its own physical borders. Localization also presents a risk for users dependent on platforms to act as a buffer against state authority. In dealing with some states around the world, US-based platforms have been reluctant to abide by national laws, up to and including challenging requests for data or information about users in court, frequently on the grounds that data is not held in that particular country. A lack of regulatory safeguards in the *data localization as part of content moderation regulation* group will not reassure platforms, while data localization weakens a company's ability to protect its local users' privacy or freedoms.

Data localization debates can cut both ways. Pressure on ByteDance from the US government to use local US data storage for US users of its app, TikTok, highlights the growing international concern about the security implications of unrestricted data flows.[44]

In a way, it would not be realistic to expect a 'one-size-fits-all', harmonized approach to data localization that would be universally adopted and operationalized. While cooperation agreements are in place to, for example, facilitate cross-border data transfers, data localization measures inherently rest on the concept of data sovereignty and, thus, countries' exercise of prerogatives and control in line with their respective national priorities. Yet stricter approaches to data localization, and the subsequent power authoritarian governments hold over their populations, raise questions regarding implications for human rights. For example, Russia's data localization requirements and strict monitoring and enforcement are seemingly motivated by government concern over the use of social media in anti-government protests. This apparent focus jeopardizes the users' (and, more generally, the population's) right to the freedom of expression, right to protest and enjoyment of broader civil and political rights.[45]

Democratic context

Legislation is inseparable from the context in which it is enforced. A full examination of the regulatory approach to platforms and the democratic integrity of each government is beyond the scope of this paper, but a partial picture can be discerned.

---

41 Komaitis, K. (2017), 'The 'wicked problem' of data localisation', *Journal of Cyber Policy*, 2(3), pp. 355–65, https://doi.org/10.1080/23738871.2017.1402942.
42 Gonzalez Monserrate, S. (2022), 'The Staggering Ecological Impacts of Computation and the Cloud', The MIT Press Reader, 14 February 2022, https://thereader.mitpress.mit.edu/the-staggering-ecological-impacts-of-computation-and-the-cloud.
43 e-Estonia (undated), 'Data Embassy', https://e-estonia.com/solutions/e-governance/data-embassy.
44 Calamug, A. (2022), 'Delivering on our US data governance', Tiktok, 17 June 2022, https://newsroom.tiktok.com/en-us/delivering-on-our-us-data-governance.
45 Newton, M. (2018), 'Russian Data Localization Laws: Enriching "Security" & the Economy', The Henry M. Jackson School of International Studies, 28 February 2018, https://jsis.washington.edu/news/russian-data-localization-enriching-security-economy.

Within the groups identified in this paper, and even globally, there are noticeable outliers. Examples include requirements for human rights due diligence reporting in countries that do not otherwise recognize human rights, and protections for freedom of expression in countries known to habitually suppress anti-government speech.

These outliers, however, tend to be anomalies. Taking the 2021 Freedom House *Global Freedom Index* as an indicator of the strengths of protection for political rights and civil liberties in a given country, it is clear that stricter regulations tended to be concentrated in countries with lower 'freedom scores'. Countries with strong democratic traditions, meanwhile, tend to support multi-stakeholder, independent regulations, with caveats in line with protections for individual liberties and rights.

Similarly, there is a strong correlation between national regimes that task independent regulators with platform regulation and countries scoring highly on the Freedom House index, while requirements around surveillance and data localization are largely found in countries with lower scores.

Figure 6. Regulatory groups and countries, measured against their Freedom House ranking

Note: Some countries – including Canada, New Zealand, Russia and the UK – were included in more than one regulatory group for the purpose of this project.
Source: Freedom House (2021), *Freedom in the World 2021: Democracy Under Siege*, report, Washington, DC: Freedom House, https://freedomhouse.org/report/freedom-world/2021/democracy-under-siege.

03 Regulatory pathways and potential solutions

Decisions made in the significant digital centres of power – Brussels, Beijing, London and Washington, DC – may be influential in shaping global approaches to platform governance.

Around the world, laissez-faire approaches to platform growth are increasingly giving way to government intervention. However, this expansion in national-level scrutiny has not been matched by international cooperation on the substance of platform regulation.

Internet pioneers' hopes of a single, unifying, global digital foundation have been realized in no small part. Never before have countries, economies, citizens and communities been so closely connected. This development has brought substantial benefits, in the spread of information, in access to economic opportunity and in connections forged between individuals and communities around the world. From business to activism, the internet has allowed for global coordination to take place in novel and powerful ways.

But this success story should not obscure the costs. New digital jurisdictions have interacted poorly with existing national political and legal institutions, challenging sovereign nations' capacity to protect their citizens, enforce their laws and set the fundamental norms of the societies they govern. Quite understandably, national platform regulation is now trying to address this capacity gap.

If the global internet has a future, it will be found in compromise and coordination between polities and economies able to find a settlement balancing national sovereignty and international interdependence and interoperability. Techno-libertarian hopes of cyberspace sitting outside the realms of the 'weary giants of flesh and steel' are unrealistic.[46]

Unified language and concepts may provide some like-minded states with a common language. Building on David Kaye's report to the UN as Special Rapporteur for Freedom of Expression, chapter 4 explores one of those concepts, asking whether a human-rights based approach may provide a route towards alignment between nations. As a long-standing framework with significant (if incomplete) global support, human rights may provide a valuable foundation for regulatory coalition-building.

The EU's approach

- The EU's power in setting the agenda for regulation is undisputed. The size of Europe's market and its considerable soft power strengthen the case for global applicability of its approach to and influence on digital platform regulation.
- Europe's collective approach and its core language of human rights make it compelling to other constituencies keen to leverage its legitimacy.
- However, paucity of enforcement and a growing emphasis in the tech industry on technical standards-setting threaten to undermine this advantage.

From data protection standards to standardized chargers for smartphones and other devices, observers point to the existence of a 'Brussels effect' in the area of regulation – i.e. the spread of European norms beyond Europe, as states and businesses elsewhere react to policy decisions made in Brussels. To an extent, this soft power is simply a function of the size of the EU market. But the inclusive, consensus-based and deliberative approach underpinning European policymaking adds further weight to legislative acts internationally.[47]

European regulation is both values-driven – reflecting the EU's democratic values, human rights and the plurality of opinions among EU member states – and strategic.[48] Under the presidency of Ursula von der Leyen, the European Commission has sought to strengthen Europe's independence in many areas of policy under the banner of 'open' strategic autonomy.[49] The European approach

46 Barlow, J. P. (1996), 'A Declaration of the Independence of Cyberspace', The Electronic Frontier Foundation, 8 February 1996, https://www.eff.org/cyberspace-independence.
47 Bendiek, A. and Stuerzer, I. (2023), 'The Brussels Effect, European Regulatory Power and Political Capital: Evidence for Mutually Reinforcing Internal and External Dimensions of the Brussels Effect from the European Digital Policy Debate', *Digital Society*, 2(5), https://doi.org/10.1007/s44206-022-00031-1.
48 European Commission (undated), 'The Digital Services Act package', https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.
49 Renda, A. (2022), 'Beyond the Brussels Effect: Leveraging Digital Regulation for Strategic Autonomy', policy brief, Brussels: Foundation for European Progressive Studies, 1 March 2022, https://feps-europe.eu/publication/853-leveraging-digital-regulation-for-strategic-autonomy.

to platform regulation has accordingly been characterized as a 'third way' – sitting between unfettered platform power and Beijing's regime of close ties between government and large tech companies.[50]

There is little doubt that European regulatory action has shaped digital platforms beyond its borders. Since Germany's NetzDG law was passed in 2017, European national and EU rules around content moderation, data protection and digital advertising have led major digital platforms to choose compliance, often amending their standard global offering to meet the requirements of their large European markets. In a 2018 House Committee on Commerce and Energy hearing in the US Congress, Mark Zuckerberg confirmed that changes to Facebook made in response to the EU's General Data Protection Regulation (GDPR) would be rolled out worldwide.[51] However, the extent to which European regulations have led to genuine change is debatable, as is the extent of the threat of enforcement.[52,53]

Within Europe, an innovative mixture of regulatory packages has emerged, designed to update and rebalance the protections from intermediary liability provided by the EU's e-Commerce Act (2000).

Within Europe, an innovative mixture of regulatory packages has emerged, designed to update and rebalance the protections from intermediary liability provided by the EU's e-Commerce Act (2000). These initiatives include the 2018 voluntary Code of Practice on Disinformation; the 2022 Regulation on Terrorist Content Online; the wide-reaching DSA (which, along with its counterpart Digital Markets Act, begins to apply throughout 2023 and 2024); and, more recently, new proposals for addressing CSAM online. The DSA in particular establishes new obligations for digital platforms to be transparent with regulators and users about their content moderation practices, to have appropriate systems and policies in place to deal with illegal content once notified, and to follow strict rules regarding the use of user data for advertising purposes. For very large online platforms and very large online search engines with over 45 million users in the EU, additional obligations around mandatory risk assessment and mitigation and independent audits apply.

Member states will enforce these rules for regular-sized platforms through national digital service coordinators, whereas the largest platforms will be accountable to the European Commission for compliance, potentially limiting the extent of the 'Brussels effect'. If a regional body is required to supervise the compliance of the largest (and most used) platforms, copycat legislation in individual states would not be enough to recreate the DSA's system of

50 Ibid.
51 Jeong, S. (2018), 'Zuckerberg says Facebook will extend European data protections worldwide — kind of', The Verge, 11 April 2018, https://www.theverge.com/2018/4/11/17224492/zuckerberg-facebook-congress-gdpr-data-protection.
52 Renda (2022), 'Beyond the Brussels Effect'.
53 Constine, J. (2018), 'A flaw-by-flaw guide to Facebook's new GDPR privacy changes', TechCrunch, 18 April 2018, https://techcrunch.com/2018/04/17/facebook-gdpr-changes.

accountability without extensive regional cooperation. However, the DSA undeniably sets a strong precedent for proportionate regulation of digital platforms that seeks to respect individual rights and freedoms. As such, the guidance for platforms and audit and transparency frameworks that the DSA produces are likely to serve as templates that many others will follow.

However, some caution is necessary when forecasting the future strength of the 'Brussels effect'. Governance models for technology are in flux, and the growing importance of international technology standards requires a different set of approaches to the more traditional rule-making that the EU is used to. Continuing negotiations on digital platform regulation – particularly transatlantic ones – are inevitable, as although US platforms depend on European markets for growth, European citizens depend on US technology provision.[54] Insofar as values-based lawmaking around digital platforms remains the primary way in which global regulatory efforts are made, the EU will continue to lead. But translating policy priorities and laws into technical standards is its own unique exercise and the EU is not currently able to compete with China in offering a 'full stack' of digital technologies, complete with standards and infrastructure, to developing countries seeking to digitize at pace.[55]

China's approach

- China's approach to domestic digital platform regulation is primarily driven by the political agenda of the ruling Communist Party of China (CPC), with political stability its main aim.
- Despite significant regulation in recent years mandating improved user capabilities, platform transparency, data protection and changes to business practices, state surveillance and control of online space remain undented and, as such, the Chinese approach is unsurprisingly non-compliant with global human rights frameworks.
- The 'Beijing effect' is an example of how greater state control of a country's domestic internet can be implemented, but not a blueprint for others to follow. Replicating China's approach in countries where US platforms have a strong presence is likely to prove difficult, as most countries lack the resources necessary.

Beijing oversees a significantly greater centralization of control over technology platforms inside its borders than other governments. However, reports of total subjugation are overstated, as evidenced by recent tensions between business practice and popular opinion, and by the inclusion of limited user protections in Chinese platform regulation regimes.

---

54 Bendiek, A. and Stürzer, I. (2022), 'Advancing European internal and external digital sovereignty: the Brussels effect and the EU-US Trade and Technology Council', Berlin: Stiftung Wissenschaft und Politik, https://doi.org/10.18449/2022C20.
55 Shi-Kupfer, K. and Ohlberg, M. (2019), *China's Digital Rise: Challenges for Europe*, report, Berlin: Mercator Institute for China Studies, https://merics.org/en/report/chinas-digital-rise.

On the one hand, the Chinese government relies on the cooperation of platforms to enforce effective control over digital content. On the other, it keeps a close eye on the expanding influence of large platforms, rolling out a series of regulations to keep big tech's power in check.

The Chinese platform ecosystem is dominated by a few large domestic businesses – most notably including Alibaba, Baidu, ByteDance and Tencent – and largely excludes major Western competitors. The government has close ties with the leadership of platform companies; the preservation of 'mainstream' values is a core tenet of Chinese platform oversight. Over the past 10 years, China's regulatory focus has moved from filtering sensitive keywords and punishing individual content uploaders to holding operators of online platforms liable for the content they host.

As such, domestic platform companies are not only required to comply with prescriptive regulatory requirements, but to devise their own rules to systematically ensure their platforms do not risk attracting unwanted government attention. Erring on the side of caution means that content deemed 'politically harmful' is strictly censored in China, and sanctioned categories remain vaguely defined and can cover a wide range of content ranging from insulting national heroes to subverting state power. This caution further leads to the deployment of proactive content moderation technologies, using both artificial intelligence tools and human labour. Chinese platforms often require users to register their real identity and to provide extensive personal information, such as mobile phone number, address and profession, to access services.

Large tech platforms in China cede extensive surveillance and control capabilities to the Chinese state. There remains, however, friction between the state and platform operators. Reporting on privacy abuse and the use of technology in exploiting Chinese workers has caused significant public outcry. The CPC has publicly stressed the need for technology platforms to serve the public and regulated to that end, though Chinese regulations have focused on business rather than on the state's surveillance capacities. The Cybersecurity Law (2017), the Data Security and Personal Information Protection Laws (2021) and, most recently, the Internet Information Service Algorithmic Recommendation Management Provisions (2022) have all led to significant changes in platform design and business practices, as the state looks to curb platform power and emphasize its position as steward of the Chinese people.

No government has had greater success in carving out a national internet than China. Chinese state power over its domestic internet is likely the envy of authoritarian regimes around the world. The Beijing effect may therefore be to provide an ideal for authorities looking to secure or justify greater control over their citizens' experience of the web. But it is less likely to become a model to replicate. This is partly due to the strength of US companies' global presence, and partly to the immense domestic resource required to manage the internet in the way China does. However, a global shift away from the traditional rule-making for digital technologies associated with European

approaches towards standardization as a model for internet governance would likely strengthen the Beijing effect, given China's head-start in engaging with and influencing global telecommunications and digital standards bodies.[56]

The UK's approach

- The UK's approach to domestic digital platform regulation is largely driven by a public conversation about online harms, with decision-makers keen to be seen to tackle high-profile instances of harm to individual users on the major platforms. This emphasis is in part tempered by concerns among some politicians, academics, public figures and citizens about over-regulation of speech.
- Global human rights frameworks are not key forces in shaping the UK's approach to platform regulation. However, a focus on scrutinizing platform systems and on transparency aligns UK regulation methodologically with other global approaches.
- Despite this approach having broad international appeal, London's influence on global regulatory norms may be limited by political barriers to international cooperation.

In March 2022, almost three years since the initial Online Harms white paper emerged and began the debate about digital regulation in the UK, the government's Online Safety Bill was published. In the intervening period, the bill underwent significant revisions, and, even since this analysis was completed in autumn 2022, has been substantially amended in both houses of parliament. (For example, removing some provisions relating to legal but harmful content for adult users and strengthening the requirements for platforms to verify the age of all users.) The bill entered into law in October 2023.

Approaches to British digital platform regulation have largely been driven by a vocal and high-profile public conversation about online harm, and heavily informed by criminal legal norms.

Beginning in earnest around 2014 and prompted in part by the proliferation of content associated with Islamic State, media coverage of online platforms in the UK has for a decade now been relentless in highlighting harms and demanding action from the UK government against the largest and most influential platforms.

Civil society in the UK, however, remains split on the issue. Proponents of far-reaching platform regulation are led by childrens' charities, high-profile whistleblowers and well-known voices in the media calling for issue-specific regulations. For example, the broadcaster and consumer rights campaigner Martin Lewis successfully called for the inclusion of scam advertising in the bill,[57] while the model and television personality Katie Price led a campaign demanding

56 Ibid.
57 UK Parliament (2019), 'Emerging trend in economic crime affecting consumers: Martin Lewis fake adverts scamming vulnerable consumers out of £1000s', written parliamentary evidence submitted to the House of Commons Treasury Select Committee by MoneySavingExpert.com, https://committees.parliament.uk/writtenevidence/90710/html.

ID verification as part of creating a social media account.[58] On the other side of the debate, internet freedom and civil liberty organizations – including, among others, Article19, Demos, Liberty, the Open Rights Group – have raised significant concerns about the compatibility of proposed regulations with legal obligations, democratic norms and protections for freedom of expression, privacy and non-discrimination. Approaches to platform regulation in the UK coalesce around these two poles: a majority wanting to be seen to be tough on platforms, protecting children and tackling harm online; and a minority concerned about implications for existing rights and freedoms in the UK.

Criminal law frameworks have had a significant influence in shaping the UK's approach to platform regulation. More imaginative approaches centred on the establishment of a statutory duty of care for adults have largely been replaced by criminalization of particular types of content or user behaviour: for instance, disinformation is now covered under a new criminal offence of foreign interference, established in the National Security Act of July 2023.[59] Tackling cyberflashing also required a new criminal offence.[60] Legal but harmful content was dropped from the Online Safety Bill before its approval.[61]

Criminal law frameworks have had a significant influence in shaping the UK's approach to platform regulation. More imaginative approaches have largely been replaced by criminalization of particular types of content or user behaviour.

Human rights frameworks have not featured prominently in the UK's approach to regulation. Where rights are mentioned, they mirror the US approach in prioritizing freedom of expression. This emphasis is exacerbated by a political desire to diverge from EU approaches following Brexit.[62]

Although the UK is an important market for major platforms, regional attention will be firmly on the EU and its approach to regulation. Moves by the UK to share its own view of best practice through a network of global digital platform regulators has been welcomed in Australia, Fiji and Ireland, but cooperation with regulators elsewhere is stymied by misalignment on what content to regulate and how. Given the significant resourcing behind the Office of Communications (Ofcom) and Ofcom's commitment to publishing guidance for regulated platforms around the Online Safety Act's passage, the UK may have gained some traction internationally by being a first mover on defining aspects of digital platform regulation.

---

58 UK Parliament (2022), 'Make verified ID a requirement for opening a social media account', https://petition.parliament.uk/petitions/575833.
59 Home Office (2023), 'Foreign interference: National Security Bill factsheet', https://www.gov.uk/government/publications/national-security-bill-factsheets/foreign-interference-national-security-bill-factsheet.
60 Milmo, D. (2022), 'New law banning cyberflashing to be included in online safety bill', *Guardian*, 13 March 2022, https://www.theguardian.com/society/2022/mar/13/new-law-banning-cyberflashing-to-be-included-in-online-safety-bill.
61 MacCarthy (2022), 'U.K. government purges "legal but harmful" provisions from its revised Online Safety Bill'.
62 Schlesinger, P. (2022), 'The neo-regulation of Internet platforms in the United Kingdom', *Policy & Internet*, 14(1), pp. 47–62, https://doi.org/10.1002/poi3.288.

US approaches

- US approaches to domestic digital platform regulation are rooted in the prioritization of market economics and promotion of a business agenda that provides space for tech companies to flourish and flexibility for states to define their own priorities.
- Individual states approach platform regulation in different ways. For example, California and Florida take widely divergent positions on the purpose, extent and deployment of appropriate platform regulation.
- The language of civil rights underpins US conversations surrounding a rights-based approach to platform regulation. The perspective and tone of existing laws and proposals focus on the US Constitution and Bill of Rights, rather than the Universal Declaration of Human Rights (UDHR) and other international legal mechanisms. This includes a heavy emphasis on the First Amendment of the Constitution and the US culture of litigation.

The US is home to dominant social media firms such as Google, LinkedIn, Meta (owner of Facebook, Instagram and WhatsApp), Pinterest, Snapchat and X (formerly Twitter). This capital – cultural, economic and social – provides the US with the capacity, connections and resources to dominate the platform governance landscape. But up to now, US legislation has sought to defend platform autonomy, putting the US at odds with other jurisdictions pushing for greater intervention. A historic reliance on industry standards over regulation has failed to translate to online platforms.

Language used at both ends of the US political spectrum has changed in the past years, with both Democrats and Republicans criticizing the autonomy afforded to platforms in making decisions on content moderation.[63,64] Growing political polarization, however, limits the scope for bipartisan agreement on platform regulation.[65] State positions are further apart still. In September 2023, California successfully passed bill AB 587, which requires social media companies to submit reports to the state by January 2024 on content moderation and policy decisions. Proponents claim this legislation is aimed at tackling 'hate and disinformation'. Meanwhile, officials in Florida are seeking to limit the extent to which platforms can moderate content at all.[66,67] While there is no comprehensive, national consensus on regulation, broad agreement among legislators on the problems caused by a lack of intermediary liability (often known as Section 230, in reference to a section of the 1996 Telecommunications Decency Act) is quickening the development of proposals promoting more bipartisan support

---

63 Kern, R. (2022), 'White House renews call to 'remove' Section 230 liability shield', Politico, 8 September 2022, https://www.politico.com/news/2022/09/08/white-house-renews-call-to-remove-section-230-liability-shield-00055771.
64 Ramseyer Draft Legislative Reforms to Section 230 of the 1996 Communications Decency Act, 2020.
65 DeSilver, D. (2022), 'The polarization in today's Congress has roots that go back decades', Pew Research Center, 10 March 2022, https://pewrsr.ch/3tMrxsF.
66 Office of Governor Gavin Newsom (2022), 'Governor Newsom Signs Nation-Leading Social Media Transparency Measure', press release, 13 September 2022, https://www.gov.ca.gov/2022/09/13/governor-newsom-signs-nation-leading-social-media-transparency-measure.
67 Oremus, W. and Zakrzewski, C. (2022), 'Florida brings battle over social media regulation to Supreme Court', *Washington Post*, 21 September 2022, https://www.washingtonpost.com/technology/2022/09/21/florida-social-media-supreme-court-scotus.

such as the Eliminating Abusive and Rampant Neglect of Interactive Technologies (EARN IT) Act, the Kids Online Safety Act and the Platform Accountability and Transparency Act. Regulatory change that challenges platform businesses is likely to be further slowed by industry lobbying. Technology companies spent a reported $55 million on lobbying the US federal government in 2021.[68]

The US is also unlikely to promote an approach based on IHRL and international standards, as the civil rights movement has historically provided the basis for the defence of minority and constitutional rights, rather than human rights frameworks and language. Recent court cases and calls for legislative change use language specific to domestic US protections for civil rights and freedoms, such as the First Amendment of the constitution, rather than the fundamental and universal rights such as Article 2 of the UDHR. For example, the Anti-Defamation League's report and subsequent policy on preventing anti-Semitic hate and harassment on social media focused solely on US civic rights.[69]

Whether California will capitalize on its internal power and sway the US debate in favour of closer regulation is yet to be determined. However, with a lack of federal-level alignment, a singular US approach to digital platform regulation is extremely unlikely to emerge in the near future. While the EU-US Trade and Technology Council does act as a forum for debate and exchange on digital transformation and cooperation,[70] the EU is therefore likely to remain the leading voice worldwide in calling for greater regulation.

Consensus and cooperation

As the previous sections show, wide gaps remain between the major centres of political power driving digital regulation. The US's constitutional commitments to freedom of expression and its hesitancy to intervene in markets will be the determining forces in shaping the web, as US tech companies continue to dictate the rules and norms for the digital tools used by the global majority. Nevertheless, US dominance has not deterred authorities and jurisdictions with conflicting values. European regulations on data, platforms and digital advertising have put significant pressure on the dominant tech companies, with many of those companies adapting their products globally to meet European standards. Meanwhile, post-Brexit, the UK wants to be seen as providing a 'third way' on technology, balancing the twin aims of enabling growth and ensuring safety. It remains uncertain whether the Online Safety Act passed in October 2023 will add to the UK's credibility on platform governance. China's decision to foster its own digital ecosystem and strictly maintain its barriers is the clearest obstacle to any attempt to establish a global governance framework for online

---

68 Birnbaum, E. (2022), 'Tech spent big on lobbying last year', Politico, 24 January 2022, https://politi.co/33QmIWa.
69 Anti-Defamation League (2021), *Online Hate and Harassment: The American Experience 2021*, report, New York: Anti-Defamation League, https://www.adl.org/resources/report/online-hate-and-harassment-american-experience-2021.
70 Schneider-Petsinger, M. (2022), *Strengthening US–EU cooperation on trade and technology*, Briefing Paper, London: Royal Insititute of International Affairs, https://www.chathamhouse.org/2022/12/strengthening-us-eu-cooperation-trade-and-technology.

platforms. The Chinese vision is not an exceptional one, even if costly and difficult to implement. Many states worldwide would choose to pursue greater digital sovereignty at the expense of global connectivity, given the choice.

Despite this divergence, powerful forces are pulling in the other direction, towards greater alignment. Many would argue that a global internet is good worth pursuing in and of itself – indeed, universal global connectivity by 2030 is one of the UN's Sustainable Development Goals.[71] Demand from citizens and business for digital services hosted or operated by international companies is strong and growing, and participation in the global economy has for decades now been predicated on digital infrastructure provided by online platforms. For instance, in 2014 restrictions on access to the open source software development platform GitHub and a series of other platforms in India were quickly reversed after an outcry from the country's tech industry.[72] Current internet infrastructure is by design better suited to openness and connectivity than to the imposition of national borders.

In the near term, only those countries or geographies with both sufficient will and sufficient resources will be able to pursue a strategy of disconnecting from the US–EU version of the web, described in depth in the *Four Internets* paper by Wendy Hall and Kieran O'Hara.[73] It is probable that only China has both the will and ability to build and maintain the full stack of digital infrastructure required to break away entirely, with the rest of the world becoming in effect a vast 'Venn diagram' of porous internets built around national languages, cultures and platforms but accessible to all, and controlled crudely. This control is more likely to be exercised through blocking access to individual websites or to the internet itself, rather than by implementing new standards or protocols. Even China must allow some internet traffic through the 'Great Firewall' in support of national and international businesses operating in the country. In the medium to long term, though, Chinese leadership – as demonstrated through trade agreements and influence in international standards bodies – and the export of Chinese digital standards and infrastructure could bring other countries into the Chinese internet.

Strong reasons for maintaining the status quo remain. The internet familiar to most users is shaped by an uneasy digital hegemony negotiated between the US and EU. Access to digital services, markets and platforms is enormously significant to businesses and citizens around the world. However, the process of agreeing joint roadmaps, principles and regulation for digital goods and services between the EU, the US and their partners is fiercely contested. Recent regulatory initiatives like the EU's DSA and the UK's Online Safety Bill have prompted significant criticism from prominent voices in the US tech sector, such as from Signal's Meredith Whittaker on encrypted communications and Wikimedia's

---

71 United Nations Office of the Secretary-General's Envoy on Technology (undated), 'Global Connectivity', https://www.un.org/techenvoy/content/global-connectivity.
72 Russell, J. (2014), 'India's Government Asks ISPs To Block GitHub, Vimeo And 30 Other Websites', TechCrunch, 31 December 2014, https://techcrunch.com/2014/12/31/indian-government-censorsht.
73 Hall, W. and O'Hara, K. (2018), *Four Internets: The Geopolitics of Digital Governance*, paper, Waterloo, ON: Centre for International Governance Innovation, 7 December 2018, https://www.cigionline.org/publications/four-internets-geopolitics-digital-governance.

Rebecca Mackinnon on age verification.[74] Meanwhile, US inaction exasperates regulators on the other side of the Atlantic. Countries outside of traditional multilateral forums feel frustrated and unable to influence the technological landscape that their citizens increasingly depend on. While global regulatory alignment is unlikely, better cooperation and dialogue between countries reliant on shared digital infrastructure are essential. Threats made by both companies and governments to withdraw services or raise barriers should not be taken lightly.

74 Newman, C. (2023), 'Online Safety Bill debate: Could it lead to 'unprecedented paradigm-shifting surveillance'?', interview with Meredith Whittaker, Channel 4 News, 3 July 2023, https://www.channel4.com/news/online-safety-bill-debate-could-it-lead-to-unprecedented-paradigm-shifting-surveillance.

04
Establishing global frameworks: the potential for a human rights-based approach

Human rights provide a well-established set of rules, norms and approaches to complex governance issues like digital platform regulation. However, they are not a simple, catch-all solution. Rather, they should be looked to for guidance and policy innovations.

Tomorrow's platform regulation may be led by efforts in Beijing and Brussels, or by decisions made in London or Washington. But hope exists for a more collaborative international approach. Internet governance has, for the past three decades, been characterized by unique multi-stakeholder bodies, from those responsible for setting web standards such as the Internet Corporation for Assigned Names and Numbers (ICANN); processes including the Paris Call, the Global Digital Compact or the Global Network Initiative; and multilateral, UN-led convening forums like the Internet Governance Forum (IGF). But in the context of platform regulation, similar processes have not yet been practically applied at a global level.

Multi-country commitments to platform regulation have been made in multiple forums. For instance, in response to livestreams broadcast online during the March 2019 terrorist attacks in New Zealand, the Christchurch Call led by France and New Zealand brings together over 100 governments and organizations in demanding platforms to take steps to eliminate terrorism-related content.[75] UNESCO is consulting on guidelines for regulating digital platforms.[76] The challenge for these global efforts is to agree on a unified global framework – existing or new – through which to approach the questions raised by platform regulation.

Human rights should underpin at least part of this framework. Yet, the adoption of values-led or human rights-led thinking in platform regulation remains somewhat complicated, and is actively avoided by some states. The data collected by researchers for this paper attest to two realities:

- Human rights have been largely overlooked in attempts to define principles around the governance of digital platforms, with the possible exception of the EU's DSA and early UN efforts; and
- Translating human rights principles into effective platform regulation is in itself a challenge.

Why the concept of human rights remains relevant

Human rights embody the idea that individuals must be protected against certain abuses perpetrated by their own governments and states, as well as by individuals and private entities.[77] The concept of human rights is recognized in international, regional and domestic legal frameworks. Their exact definition and scope of protection vary in each. But IHRL is quasi-universal, flexible and already binding on most states. IHRL therefore provides a pre-existing and widely accepted set of principles, rules and definitions that could be adopted as part of a global framework for online platform governance.

While human rights may not have all the answers, they provide a well-established foundation for sound governance of technology.[78] Despite nuances in its interpretation and implementation in different jurisdictions, IHRL is regarded by many as developed and adaptive enough to address regulatory issues and gaps. It provides a clear and robust framework to mitigate risks of human rights violations by imposing binding obligations on states to respect, protect and ensure a range of fundamental rights, as well as the establishment of monitoring and oversight processes and an ecosystem of safeguards and accountability

75 The Christchurch Call (undated), 'Our Work', https://www.christchurchcall.com/our-work (accessed 19 May 2023).
76 UNESCO (2022), 'Guidelines for Regulating Digital Platforms', programme and meeting document, https://unesdoc.unesco.org/ark:/48223/pf0000382948.
77 Klabbers, J. (2017), *International Law* (2nd edition), Cambridge: Cambridge University Press, p. 120.
78 Jones, K. (2023), *AI governance and human rights: Resetting the relationship*, Research Paper, London: Royal Institute of International Affairs, https://doi.org/10.55317/9781784135492.

mechanisms.[79] Applying this existing framework to platforms may be a viable route to tackling harms online, while also balancing the rights of individuals with the interests of governments and corporations that hold power over digital space.

Assumptions and misconceptions

Yet human rights have not yet been used to their full potential. To date, platform regulations around the world have not drawn heavily on existing human rights frameworks. As examined in this paper, just one in five of the regulations in place demand that platforms carry out human rights due diligence assessments.

Online content regimes largely focus instead on other concepts, most commonly that of harm and its prevention.[80] Discussions on platform regulation also tend to omit human rights expertise.[81] This omission often leads to misleading assumptions and misconceptions in the tech sector, such as that human rights are only a concern for governments and not for companies, for whom they remain mere ethical considerations and not legally relevant.[82]

The sidelining of human rights in approaches to digital platform governance contrasts with the rich literature discussing in detail the importance of upholding the freedom of expression in digital platforms, identifying the opportunities and pitfalls of such an approach, exploring avenues for progress and including the private sector.[83] It also contrasts with the human rights emphasis of regional and international discussions and decisions, including in the context of the Council of Europe, UNESCO and the UN's Human Rights Council.[84]

---

79 McGregor, L., Murray, D. and Ng, V. (2019), 'International Human Rights Law as a Framework for Algorithmic Accountability', *International & Comparative Law Quarterly*, 68(2), pp. 309–43, https://doi.org/10.1017/S0020589319000046.

80 These areas of focus in regulating digital platforms were discussed during Chatham House's workshop at the 2022 Internet Governance Forum held on 1 December 2022 in Addis Ababa, Ethiopia. For a summary of the discussions, see Internet Governance Forum (2022), 'IGF 2022 WS #458 Do Diverging Platform Regulations Risk an Open Internet?', meeting summary, 1 December 2022, https://www.intgovforum.org/en/content/igf-2022-ws-458-do-diverging-platform-regulations-risk-an-open-internet.

81 Jones (2023), *AI governance and human rights*.

82 Ibid.

83 See, for example, Sander, B. (2020), 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation', *Fordham International Law Journal*, 43(4), pp. 939–1006, https://dx.doi.org/10.2139/ssrn.3434972; Kaye, D. and Shaffer, G. C. (2021), 'Transnational Legal Ordering of Data, Disinformation, Privacy, and Speech', *UC Irvine International, Translational & Comparative Law*, 6(1), https://scholarship.law.uci.edu/ucijil/vol6/iss1/2; Kaye, D. (2022), 'Human rights standards should guide company decisions', research paper, Irvine, CA: University of California, Irvine – School of Law, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4246044; and Pasquale, F. (2016), 'Platform neutrality: enhancing freedom of expression in spheres of private power', *Theoretical Inquiries in Law*, 17(2), https://doi.org/10.1515/til-2016-0018.

84 In the context of the Council of Europe, for example, a formal motion for a resolution on the public regulation of the freedom of expression in digital platforms was submitted by a number of members of the Parliamentary Assembly, see Katrougalos, G. (2022), 'Public regulation of the freedom of expression in digital platforms', Parliamentary Assembly of the Council of Europe, 21 June 2022, https://pace.coe.int/en/files/30118. UNESCO explicitly works to promoting freedom of expression online, while advocating for 'greater transparency and accountability of digital platforms' in the light of the Windhoek +30 Declaration: see UNESCO (undated), Freedom of Expression Online', https://www.unesco.org/en/freedom-expression-online (accessed 30 January 2023). In the context of the UN Human Rights Council, the then Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, presented a report specifically on the regulation of user-generated online content, see UN Office of the High Commissioner for Human Rights (2018), *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*.

This paradox may not only attest to the attitude of governments and their diverging approaches to digital platform governance. It also reflects more generally the discrepancies between interests at the international level, and at least certain regional contexts, and political will at the national level.

Where human rights are foregrounded in digital platform regulation, primacy tends to be given to one right over others. Freedom of expression is routinely presented as the main (and, sometimes, only) focus by platforms in their policies and governments in regulatory tools. This is most notable in the context of online content moderation – or even, in certain cases, arguably used as a 'laissez-passer', a pretext to quieten dissenting voices. Meanwhile, the rights to privacy, freedom of thought and access to information and the media – all inherently part of the freedom of expression – and other rights are rarely given equal status. This prominence of one right over others not only contradicts the idea that human rights form a single, indivisible body of rights. It also raises questions over the potential for human rights to be placed into a hierarchy – or for individual rights to be graded against one another.

The debate of universality vs prioritization is not a new one in the context of human rights. But, in digital platform regulation, the priority given to freedom of expression may undermine attempts to follow a holistic rights-based approach. Decision-makers must exercise caution in the way they manage conflicting interpretations of human rights principles and obligations, and more broadly in approaching human rights through political prioritization.

In practice

A human rights-based approach must not be considered as the complete remedy to abusive and exploitative platforms. Rather, it should be applied alongside other relevant legal regimes (e.g. criminal law for liable offences), as well as standards, regulations and other 'soft' law tools. This is of particular importance in light of the largely private ownership of digital platforms, as IHRL remains, essentially, binding on states only.[85] Nevertheless, this is not to discount the responsibilities that the wider human rights framework may confer on companies and other non-state actors. For example, in certain domestic contexts, corporations have due diligence duties to identify, mitigate and remedy human rights risks, and may be held liable for failing to do so.[86] This would be the case, for example, under the directive on corporate sustainability due diligence within EU law adopted by the European Commission in 2022.[87]

Despite these limiting factors, a human rights-based approach could still provide a strong underpinning for an approach to platform governance. Human rights and IHRL provide states, platforms and multi-stakeholder coalitions with an appropriate language for online content governance. They also set out how human

85 Jørgensen, R. F. (2017), 'What Platforms Mean When They Talk About Human Rights', *Policy & Internet*, 9(3), pp. 280–96, https://doi.org/10.1002/poi3.152.3.
86 Jones (2023), *AI governance and human rights*.
87 European Commission (2012), 'Just and sustainable economy: Commission lays down rules for companies to respect human rights and environment in global value chains', press release, 23 February 2022, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1145.

rights may be respected, protected and ensured in different contexts – online and offline. If the Brussels effect is indeed felt around the world, human rights may well become the framework of choice. But efforts by international bodies to strengthen the adoption of such approaches through multi-stakeholder dialogue must continue, if the trend towards multiple competing and conflicting national regulatory approaches is to change.

As noted earlier in this paper, the right to seek, receive and impart ideas and information of all kinds – i.e. the freedoms of expression and information – are of particular importance in the context of digital platforms.[88] Online content is the expression of ideas or information, which may be sought by billions of internet users worldwide. But other human rights also apply online and deserve equal, if not greater, protection in different contexts. For example, states must, and platforms should, protect the lives and health of individuals from the threats to public health posed by disinformation of the kind seen during the COVID-19 pandemic.[89] IHRL, including international and regional human rights treaties, already provides the tools to navigate these conflicts between rights.

Human rights and IHRL provide states, platforms and multi-stakeholder coalitions with an appropriate language for online content governance.

For instance, Article 19(3) of the International Covenant on Civil and Political Rights (ICCPR) provides that the exercise of the rights to freedom of expression and information carries with it special duties and responsibilities.[90] This means that those rights may be limited by law to safeguard a legitimate aim, and insofar as necessary and proportionate in the circumstances: the so-called 'tripartite test' of legality, legitimacy, and necessity and proportionality.[91] Article 19(3) ICCPR identifies as legitimate aims that may justify limitations to freedom of expression and information: the respect for the rights or reputations of others; the protection of national security; or the protection of public order, public health or morals. Similar provisions are found in Article 10(2) of the European Convention on Human Rights[92] and Article 13(2) of the American Convention on Human Rights.[93]

---

88 See UN Human Rights Council (2021), 'Disinformation and freedom of opinion and expression Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan', 13 April 2021, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf?OpenElement, para 37.
89 See Urs, P., Dias, T., Coco, A. and Akande, D. (2023), *The International Law Protections against Cyber Operations Targeting the Healthcare Sector*, Oxford Institute for Ethics, Law and Armed Conflict, February 2023, pp. 221–22, https://www.elac.ox.ac.uk/wp-content/uploads/2023/04/ELAC-Research-Report_International-Law-Protections-against-Cyber-Operations-Targeting-the-Healthcare-Sector.pdf.
90 International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171.
91 See Aswad, E. (2020), 'To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?', *Washington and Lee Law Review*, 77(2), pp. 609, 618 and 622.
92 Convention for the Protection of Human Rights and Fundamental Freedoms (adopted on 4 November 1950, entered into force 3 September 1953) ETS No 5.
93 Organization of American States (OAS), American Convention on Human Rights, 'Pact of San Jose', Costa Rica, 22 November 1969.

In the context of digital platforms, including private messaging apps, search engines and social media, these provisions have three significant implications. First, states must enact legislation or regulation and companies should adopt policies that define: i) what kinds of online content may be limited; ii) for what purpose they may be limited; and iii) how they may be limited.[94] Legislation and company policies should be sufficiently clear, accessible and transparent.[95] Second, necessity and proportionality require limitations on speech to be balanced against the importance of the rights or interests at stake (e.g. public health, morals and the rights or reputations of others).[96] A non-binary approach to online content governance would ensure that these provisions are upheld.[97] Content that may be restricted should not be simply taken down or left in place. Other measures should be available, such as labelling or deprioritizing content, or directing users to other sources of information.[98] Finally, states must ensure that platforms put in place user redress or review mechanisms as a safeguard against wrongful content moderation decisions.[99] Errors are unavoidable in an environment where machine-learning algorithms sift through billions of posts every day.[100] But decisions and the processes behind them must be open to challenge.

Neither consensus-building nor the promotion of a human rights-based approach are straightforward, complete solutions. By uncovering the patterns and commonalities of current platform regulatory regimes, this paper has shown how regional and international agreement can become more attainable. It remains unclear whether agreement will be based on countries being allowed a measure of difference in their approaches or on broader alignment. Despite being overlooked by many in the tech community, human rights will always be a crucial part of this dialogue as a well-established international rulebook for greater transparency, accountability and remedy. International cooperation and alliances are achievable, and human rights must remain universal, even in the digital world.

---

94 Dias, T. (2022), 'Tackling Online Hate Speech through Content Moderation: The Legal Framework Under the International Covenant on Civil and Political Rights', in Bahador, B., Hammer, C. and Livingston, L. (eds.) (forthcoming), *Countering online hate and its offline consequences in conflict-fragile settings*, SSRN, p. 17, https://ssrn.com/abstract=4150909; Dias, T. (2021), 'Hate Speech and the Online Safety Bill: Ensuring Consistency with Core International Human Rights Instruments', Evidence Submission to the House of Commons Digital, Culture, Media and Sport Sub-committee on Online Harms and Disinformation, September 2021, pp. 6–8, 9–15, https://committees.parliament.uk/writtenevidence/38393/pdf. See also UN Human Rights Council (2021), 'Disinformation and freedom of opinion and expression Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan', paras 40–41; UN General Assembly (2019), 'Promotion and protection of the right to freedom of opinion and expression: Note by the Secretary-General', https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/308/13/PDF/N1930813.pdf?OpenElement', 9 October 2019, paras 31–32.
95 UN Human Rights Committee (2011), 'General comment No. 34 - Article 19: Freedoms of opinion and expression', 12 September 2011, https://daccess-ods.un.org/tmp/39093.7086194754.html, paras 25, 33–36; UN General Assembly (2019), 'Promotion and protection of the right to freedom of opinion and expression', paras 6(a), 20, 31–33.
96 UN Human Rights Committee (2011), 'General Comment No. 34', paras 33–36.
97 Dias (2022), 'Tackling Online Hate Speech through Content Moderation', p. 18.
98 Ibid.
99 Ibid, pp. 15, 22–24 and 27; UN General Assembly (2019), 'Promotion and protection of the right to freedom of opinion and expression', paras 7, 35, 55 and 57(e).
100 Douek, E. (2021), 'Governing Online Speech: From "Posts-as-Trumps" to Proportionality and Probability', *Columbia Law Review*, 121(3), pp. 763–834, https://columbialawreview.org/content/governing-online-speech-from-posts-as-trumps-to-proportionality-and-probability.

05 Conclusion and recommendations

This research paper has highlighted the challenges of achieving global platform regulation and underscored the gulf between the current technological reality of a near-global internet and the divergent national approaches to its governance.

Current responses at the national level are defined by deep cultural and political divisions, and attempts to reach international agreement on platform governance are likely to be difficult. In response to these realities, the following proposals present possible next steps for policymakers and organizations seeking to tackle national divergence, invest in international cooperation and preserve an open, global internet.

Prioritize a shared lexicon and cooperation through networks of regulatory bodies

Despite divergent national approaches, there is significant harmony in the language used by governments when approaching platform regulation, most often on questions of security, mitigating harms and protection for users. Governments must empower their respective regulatory bodies to seek international consensus on:

- Articulating clearly the values that underpin regulatory approaches and identifying consensus across emergent approaches around the world;
- Strengthening informal and formal international networks of regulators to boost information-sharing and exchange of technical expertise. (The Global Online Safety Regulators Network provides a useful starting point);
- Ensuring that regulatory bodies cooperate on the design and execution of regulation within the bounds of domestic legislation;

— Supporting technical transparency among regulators to create opportunities for joint adoption of regulatory instruments and ensure greater consistency between regulatory approaches; and

— Investing in technologies that build and enhance institutional consensus within regulatory bodies and their networks.

Build on human rights language to formulate human rights-based platform regulation

Regulatory approaches should be tested against human rights frameworks, particularly those recognized in the International Covenant on Civil and Political Rights (ICCPR) and other existing international and regional human rights instruments. Content policies or standards should equally reflect the international human rights legal framework. This means that:

— States must enact sufficiently clear, accessible and transparent laws/regulations for online content governance. The same should apply to platforms' content moderation policies. Laws, regulations and policies should carefully balance the rights of users and the general public to seek, receive and impart information and ideas of all kinds with other competing rights or interests online, such as non-discrimination, privacy and health.

— Achieving this balance will require a non-binary approach to content governance or moderation. Limitations on different types of online content should go beyond a 'take down/leave up' binary to include other measures, such as labelling, deprioritization and digital 'nudges' towards other sources of information. Review and redress mechanisms are an essential safeguard against erroneous content moderation decisions and flawed processes behind those decisions.

— Support must be given to continued development of 'systems first' regulatory approaches that focus on the intents, outcomes and processes employed by technology platforms, rather than on individual pieces of content or instances of user behaviour.

Support and secure the work of global internet standards bodies

Long-standing institutions for global internet governance already exist – for example, the IGF and the handful of standards bodies responsible for making technical decisions on the internet's architecture. Stronger national sovereignty over the internet will come from working within and alongside these institutions, rather than against them – a principle currently better understood by China than by other major digital powers.

Internationally harmonized and consistent regulation is only viable when what is being regulated is consistent across borders. To this end, governments must:

— Ensure sufficient government expertise to support political decision-makers in understanding the possible technical routes to achieving policy aims;

— Work with partners to ensure existing multi-stakeholder bodies remain politically neutral, amid a growing threat of state and institutional capture.

- Invest in their capacity to participate actively in existing multi-stakeholder bodies and engage with other national communities doing the same. This includes investing in the technical expertise required;
- Encourage national participation in standards bodies through:
  - Industry secondments from businesses sharing government values; and
  - Philanthropic and/or government support for civil society participation;
- Demand consistent reporting and discussion within existing multilateral bodies (e.g. EU–US Trade and Technology Council, G7, G20, UN, WTO) on questions of digital trade to underscore the importance of neutrality and independence for internet standards bodies; and
- Highlight efforts by hostile nations and corporate monopolies to undermine this neutrality and independence.

Significantly increase investment in bilateral and multilateral software cooperation

Global cooperation on software and digital regulation remains limited. Alongside regulatory efforts to affect the current landscape, it is critical that governments take seriously the requirement to design and deploy what comes next. These efforts should entail:

- Embracing a modular approach to cross-border collaboration on common processes and codes of practice;[101]
- Joint funding of sovereign technology investment where mutual societal requirements can be identified and met through a single project;
- Joint financing for independent technology funds to support the development of technologies outside of current US-centric investment models; and
- Strengthening formal and informal networks of digital collaboration across national governments to support the development of, and cooperation on, multilateral technology programming.

101 Riley and Ness (2022), 'Modularity for International Internet Governance'.

Appendices

Appendix 1: Methodology

Identification of laws and proposals for analysis

The scope of laws and policies around the world that apply to online platforms in some way is huge, including, among others, consumer protection regulations, competition law, media and broadcasting regulations and data protection legislation. For the purposes of this study, research focused on laws that relate specifically to the question of platforms' intermediary liability for user-generated content: namely, on laws that impose legal requirements on how platforms should moderate content. Narrowing the scope of the mapping of platform regulations to focus specifically on laws that introduce such requirements allowed us to explore in greater depth how regulators are currently grappling with the novel issues that platforms pose compared to other types of businesses. More developed regulatory systems will include provisions impacting platforms elsewhere.[102]

To map out relevant laws and policies, researchers conducted an initial mapping of laws and proposals that might potentially be in scope as of October 2022, drawing on:

- Global Partners Digital's existing body of research and monitoring of platform regulation laws around the world;
- Published mappings or reviews of global platform regulations or intermediary liability legislations;[103]
- Feedback from regional and local experts; and
- Desk research using national legal gazettes and records of legislation.

In some cases, translation tools were used to assist with conducting the mapping, which included laws currently in force as well as draft legislation and proposals. (Voluntary codes of practice or self-regulatory initiatives were not included.)

---

102 For instance, the EU's Digital Markets Act and AI Act will have major ramifications for social media platform moderation practices.

103 Including Stanford Law School Center for Internet and Security (undated), 'World Intermediary Liability Map', https://wilmap.stanford.edu; The Global Network Initiative (undated), 'Country Legal Framework resource', https://clfr.globalnetworkinitiative.org/compare2; and Mchangama, J. (2020), 'The Digital Berlin Wall Act 2:
How the German Prototype for Online Censorship went Global – 2020 edition', Justitia, 1 October 2020, https://justitia-int.org/en/the-digital-berlin-wall-act-2-how-the-german-prototype-for-online-censorship-went-global-2020-edition.

Through this initial exercise, researchers identified 137 laws and proposals across 95 jurisdictions as potentially including requirements relating to how online platforms moderate online content. These examples were taken forward for further examination.

Of this initial group of 137 laws, 82 were excluded as not in scope for further analysis for the following reasons:

- Five laws were excluded because they had already been repealed and were no longer in force;
- 11 proposals were excluded because they had been stalled for some time or had been identified by local researchers as highly unlikely to pass into law due to lack of support;
- 21 laws and proposals were excluded because they contained only a simple intermediary liability clause exempting platforms from liability for any user-generated content and or any content moderation decisions;[104]
- 17 laws and proposals were excluded because their requirements on online platforms did not relate specifically to the moderation of online content (for example, laws which focused on access, data privacy or anti-monopoly);
- 17 laws and proposals were excluded because, while they did relate to management of content online, the requirements were applicable only to internet service providers, media and press outlets or regulators rather than online platforms themselves; and
- 11 were excluded because it was not possible to find publicly available versions or reliable translations at the time of research, to be able to analyse whether the law included requirements for platforms' content moderation.

The remaining group of 55 laws and proposals (spanning 41 jurisdictions) were taken to be in scope and were analysed in full. Where laws or proposals were amendments to or regulations under an existing law, these were investigated in tandem as one holistic regulatory framework.[105] Of the 55 regulatory frameworks examined, 35 were already in force as of October 2022, 11 had been introduced as bills but not yet passed, and nine were draft proposals or frameworks. Appendix 2 contains a full list of regulations considered.

This list is intended to be a robust snapshot of existing laws and proposals placing requirements on how platforms should moderate content as of October 2022. But it is by no means exhaustive, and it is important to acknowledge that there may be other laws and proposals that include requirements on how platforms moderate online content which are outside of this dataset.

---

104 For example, Brazil's Marco Civil da Internet, or Section 230 of the United States Code enacted as part of the Communications Decency Act of 1996.
105 For example, Indonesia's Ministerial Regulation No. 5, 2020 was issued under Government Regulation No. 71, 2019, and therefore these two pieces of legislation were analysed together as one of the 55 regulatory frameworks.

Analysis of laws and proposals

The laws and proposals identified for analysis vary considerably in scope, approach and implementation. Some are hundreds of pages long and specifically focused on online safety and platform regulation, whereas others are just clauses in a broader piece of legislation. To be able to conduct quantitative analysis and to compare approaches across the whole group of regulations, researchers developed a taxonomy for analysis that could be applied to each law or proposal through a series of yes/no questions. The taxonomy was designed to capture trends and variation across platform regulations in terms of:

- The types and sizes of platforms the law includes in its scope;
- The nature and categories of online content the law relates to;
- The way that the regime is enforced, including through penalties and an independent regulator;
- The model of intermediary liability applied to platforms for user-generated content;
- The types of duties they most commonly introduce for platforms with relation to content moderation or other relevant processes; and
- The degree of protection they provide for freedom of expression.

The taxonomy consisted of 29 yes/no questions drawn from a preliminary evidence review, Researchers grouped these questions into six broad themes: 1) scope and governance; 2) penalties and sanctions; 3) content-based duties; 4) business-based duties; 5) considerations for freedom of expression; and 6) protections for users.

Scope and governance

- Does the regulation differentiate between types of digital platforms? (For example, between video streaming services and social network platforms?)
- Does the regulation differentiate between sizes of digital platforms? (For example, as measured by annual revenue or number of users or employees?)
- Is the regulation enforced by an independent authority?
- Does the regulation require a multi-stakeholder approach to platform governance?

Penalties and sanctions

- Does the regulation impose fines?
- Does the regulation threaten platforms with restrictions or blocking for non-compliance?
- Does the regulation threaten prison sentences for platform employees for non-compliance with content moderation requirements?

Content-based duties

- Does the regulation require platforms to remove prohibited content when ordered to do so by a court?
- Does the regulation require platforms to remove prohibited content whenever it is notified of such content?
- Does the regulation require platforms to proactively monitor for prohibited content?
- Does the regulation require platforms to remove prohibited content within a specific timeframe?
- Does the regulation tackle content which is already designated as illegal under other legislation?
- Does the regulation designate new types of content as illegal?
- Does the regulation require platforms to remove or deal with content that is not illegal?

Business-based duties

- Does the regulation require platforms to register its services with authorities?
- Does the regulation require platforms to establish a local office or local contact?
- Does the regulation require platforms to carry out human rights risk assessments?
- Does the regulation require platforms to report regularly on the performance of their content moderation systems?
- Does the regulation require platforms to report regularly on advertising revenue?
- Does the regulation require platforms to submit to independent audit?
- Does the regulation require platforms to store data locally?

Considerations for freedom of expression

- Does the regulation explicitly mention freedom of expression?
- Does the regulation reference platforms' responsibility to consider freedom of expression in their operations?
- Are there regulatory exemptions for journalistic, scientific or public interest content?
- Are there limitations on the powers of regulators in line with freedom of expression safeguards?

Considerations for user capacities

- Does the regulation require platforms to publish terms of service?
- Does the regulation require platforms to implement complaints mechanisms?
- Does the regulation require platforms to implement appeals mechanisms?
- Does the regulation require platforms to notify users of ongoing complaints or appeals?

These questions were designed to capture the variation in approaches looked at in a quantifiable way and to make clear commonalities and differences across jurisdictions and regimes. However, the limitations of this method were that:

- The nature of the questions themselves was informed by the researchers' own experience and understanding of platform regulations. For example, the focus was informed more by expertise on freedom of expression standards and safeguards than by expertise on children's rights or minority rights.
- Not all features or relevant details of each platform regulation are represented in the 29 questions. For example, the taxonomy did not capture whether the platform regulation includes requirements to trace the first sender of a message or to monitor private or encrypted communications.
- The binary nature of the 29 questions, while necessary in order to aggregate data, does not capture qualitative details that might also be relevant. For example, while the taxonomy shows whether a platform regulation differentiates between types or sizes of platforms, it does not capture or compare what those categories are across different legislative frameworks.
- The taxonomy does not differentiate between proposals not yet passed and laws already in force.
- In some cases, where official English translations were not available, translation tools were required to interpret relevant clauses, which could have introduced some errors in analysis.
- Only the text of each law was analysed, rather than any implementation or enforcement in practice.
- Due to the fast pace of change of this regulatory area, by the time of publication some of the draft or proposed laws may have been amended since the analysis was conducted. A small percentage of the responses in the dataset may therefore no longer be accurate.

Despite these caveats and limitations, this dataset still serves as a useful starting point of analysis for core elements of the 55 laws and proposals considered.

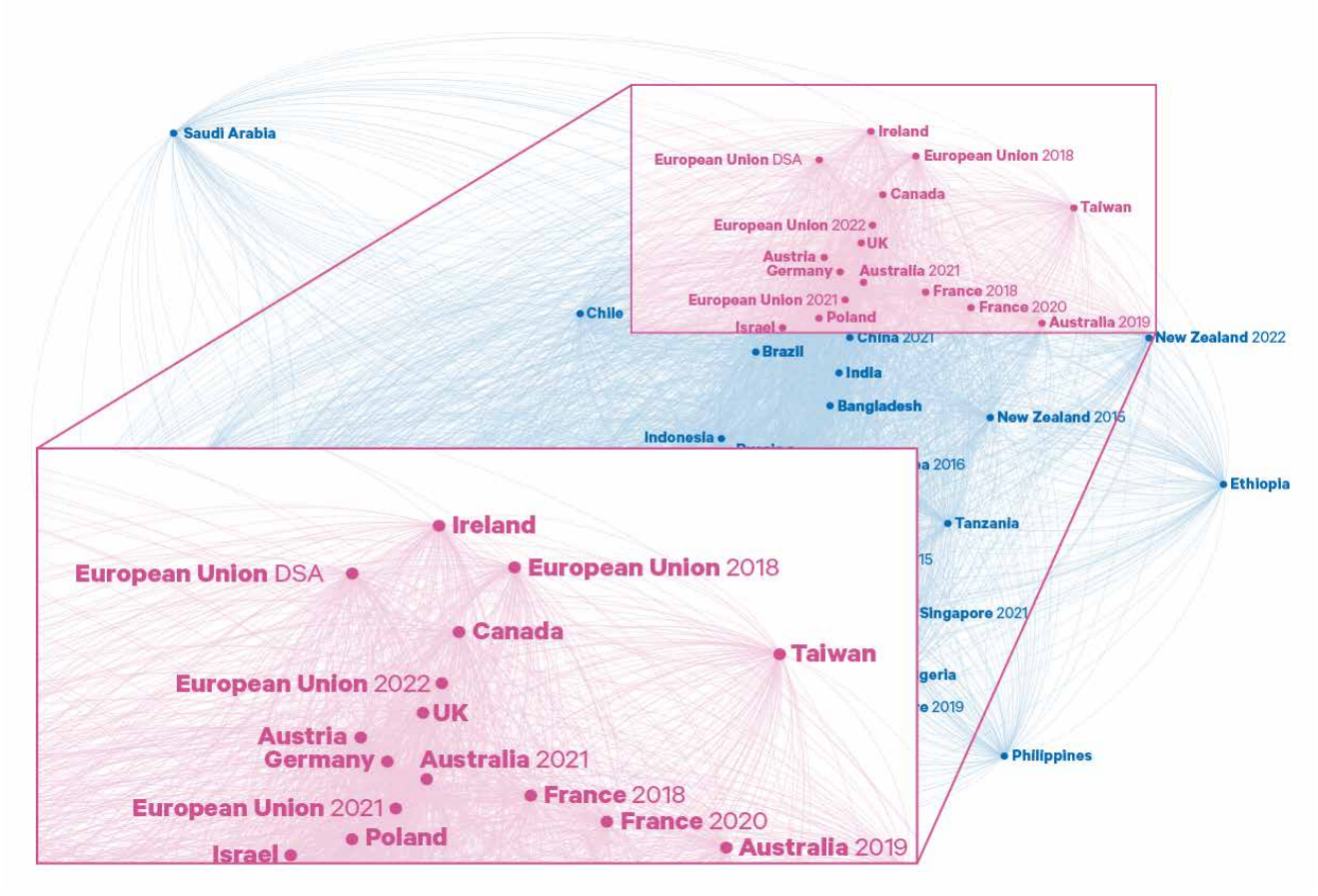
Thematic analysis of approaches to platform regulation

With 29 data points for each of the 55 laws and proposals (and thus 1,595 data points in total), data analysis tools were used to identify and draw out similarities and differences between regulations from the raw data and to provide insights for further analysis.

Two regulations that had matching answers to the 29 yes/no questions were deemed 100 per cent similar, while two regulations that differed on all 29 questions were deemed 0 per cent similar. Between these two extremes, most regulations had at least some answers in common. Regulations were mapped by measuring their similarity using a Jaccard Similarity score across the 29 binary attributes, with the similarity score used as an edge weight in a simple clustering software package (Gephi) using a default graph layout algorithm (ForceAtlas2).[106–108] This allows

for an at-a-glance analysis to see how similar or different the various regulations were: the further apart two regulations were on the map, the less similar they were.

Figure A1. Overall network map, showing detail of some similar regulations around European approaches



There are many ways to visualize relationships, similarities and differences between regulatory approaches, and the addition or subtraction of data could reshape the mapping significantly. Nevertheless, the approach taken for this study suggests that on the metrics chosen, there is some variety around regulatory approaches, and that differentiating by approach is a helpful way to look at global trends.

For instance, the cluster of regulations highlighted in Figure 1 around Bangladesh, Nigeria and Singapore tend to include provisions for prison sentences for platform employees who do not comply with content moderation duties, a provision that is not significantly present elsewhere. The cluster of regulations centred around approaches to platform regulation in the European countries and the UK tend to include provisions for an independent regulator, and in this way differentiate

106 Karabiber, F. (undated), 'Jaccard Similarity', https://www.learndatasci.com/glossary/jaccard-similarity.
107 Gephi (undated), 'The Open Graph Viz Platform', https://gephi.org.
108 Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014), 'ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software', *PLOS ONE*, 9(6), https://doi.org/10.1371/journal.pone.0098679.

themselves from other regulatory approaches found around the world. These differences underpin the five approaches that, in consultation with experts, this paper presents as contrasting global approaches.

In some cases, the provisions of a specific regulation could overlap with multiple approaches. For example, New Zealand's regulatory approach threatens employees at platforms with prison sentences,[109] but also includes provisions for regulatory independence and provisions for the protection of freedom of expression. The overlapping nature of New Zealand's approach is reflected by its inclusion under both approach 1 – *strict custody* – and approach 2 – *independent regulation*.

The five approaches set out in this paper are primarily illustrative and intended to support a discussion of major divergences between platform regulation that are difficult to group regionally or linguistically, and should be treated as such. Where analysts agreed the threshold for inclusion was not met, certain regulations were not grouped under any of the five approaches. Thailand's Draft Decree Regulating Digital Platforms, for instance, focused primarily on the regulation of e-commerce while making light-touch provisions for content regulation, business practices and government powers.[110]

Validation of research findings

In addition to desk research, the data and insights generated were tested with a range of stakeholders through individual consultations, as well as two focused discussions hosted by Chatham House and Global Partners Digital.

The research team undertook several consultations with individual experts and practitioners in the field of platform governance. The consultations were held both online and offline in a semi-structured interview format. The discussions presented an opportunity not only to collect insights on the subject; researchers were also able to confirm the veracity of data and stress-test the insights generated through this process. By including experts from multiple regions of the world, researchers strived to address, as much as possible, issues that could have eventually stemmed from geographical biases and language barriers.

In addition, researchers convened two international roundtable discussions, presenting initial findings and inviting feedback from participants to complete and strengthen this research paper. The first roundtable was hosted online in November 2022, held under the Chatham House Rule and focusing on Latin America. By convening experts and practitioners from that region, researchers were able to collect insights and perspectives that were not only reflected in this paper but placed into context, helping to overcome the Euro- and Western-centric biases that generally tend to dominate this space. Researchers adopted the same approach in December 2022, when Chatham House and Global Partners Digital convened a hybrid roundtable at the Internet Governance Forum (IGF) in Addis

109 Imprisonment is possible for online content hosts who are natural persons who refuse to comply with a court order to remove content. Given that the definition of online content host is 'the person who has control over the part of the electronic retrieval system, such as a website or an online application, on which the communication is posted and accessible by the user', this could conceivably result in imprisonment of an owner of a platform. See New Zealand Parliamentary Counsel Office (2015), 'Harmful Digital Communications Act 2015', https://www.legislation.govt.nz/act/public/2015/0063/latest/whole.html#DLM5711855, article 21(2(a)).
110 Draft Decree Regulating Digital Platforms, Thailand.

Ababa, Ethiopia. As part of those discussions, panellists representing the African, European, Latin American and South Asian perspectives shared insights into digital platform governance and the regulatory approaches in their respective regions. An industry representative and participants from across the globe were also present. Researchers also used this opportunity to present their initial findings to panellists and participants and seek feedback. Chatham House and Global Partners Digital ensured that suggestions and comments received were taken into account and reflected in the paper.

Appendix 2: Relevant laws and regulations

Laws and regulations approved and in force

Country	Legislation (accessed September 2023)	Date passed
Albania	Council of Ministers decision 465 on Measures to protect children from access to content illegal and/or harmful on the internet [made under the Law on the Rights and Protection of Children, 2017]	2019
Australia	Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act	2019
	Online Safety Act	2021
	Online Safety (Basic Online Safety Expectations) Determination	2022
Austria	Federal law enacting a communication platforms law and amending the KommAustria law	2020
Azerbaijan	Law on Information, Informatisation and Protection of Information	1998 [key amendments passed in 2017]
Belarus	Law of the Republic of Belarus On Mass Media 427-Z	2008 [key amendments passed in 2018]
China	Provisions on the Governance of the Online Information Content Ecosystem	2019
Ecuador	Communications Law	2013
Ethiopia	Hate Speech and Disinformation Prevention and Suppression Proclamation No. 1185/2020	2020
EU	Directive 2018/1808 of the European Parliament and of the Council amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities	2018
	Regulation 2021/784 of the European Parliament and of the Council on addressing the dissemination of terrorist content online	2021
	Regulation 2022/2065 of the European Parliament and of the Council on a Single Market For Digital Services and amending Directive 2000/31/EC ('The Digital Services Act')	2022
France	Law n° 2018-1202 of 22 December 2018 relating to the fight against the manipulation of information	2018
	Law n° 2021-1109 of 24 August 2021 consolidating respect for the principles of the Republic [amending Law n° 2004-575 on Confidence in the Digital Economy and Law n° 1986-1067 on Freedom of Communication]	2021
Germany	Network Enforcement Act ('NetzDG')	2017

Country	Legislation (accessed September 2023)	Date passed
India	Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules ('The IT Rules')	2021
Indonesia	Government Regulation Number 71 of 2019 concerning Implementation of Electronic Systems and Transactions	2019
	Regulation of the Minister of Communication and Information Technology Number 5 of 2020 concerning Private Electronic System Operators ('MR-5')	2020
Iran	Law No. 71063 on Computer Crimes	2009
Ireland	Online Safety and Media Regulation Act	2022
Kazakhstan	Law on Informatisation	2015
Kenya	Guidelines for Prevention of Dissemination of Undesirable Bulk Political SMS and Social Media Content via Electronic Communications Networks	2017
Malawi	Electronic Transactions and Cyber Security Act	2017
Mali	Law No 2019-056 on the Suppression of Cybercrime	2020
New Zealand	Harmful Digital Communications Act 2015	2015
	Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Act	2021
Pakistan	Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules	2021
Russia	Federal Law No. 272-FZ "On Measures to Influence Persons Involved in Violations of Fundamental Human Rights and Freedoms, Rights and Freedoms of Citizens of the Russian Federation"	2012 [key amendments passed in 2020][111]
	Federal Law of 27 July 2006 No. 149-FZ "On Information, Information Technologies and Protection of Information"	2006 [including amendments up to 2022][112]
Saudi Arabia	Draft Digital Content Platforms Regulations	2022
Singapore	Protection from Online Falsehoods and Manipulation Act	2019
	Foreign Interference (Counter-measures) Act	2021
	Online Safety (Miscellaneous Amendments) Act	2022
South Korea	Act No. 14080 on the Promotion of Information and Communications Network Utilization and Information Protection	2016
Syria	Law on Cybercrime	2022

111 Analysis does not include amendments made in June 2023.
112 Analysis does not include amendments made in July 2023.

Country	Legislation (accessed September 2023)	Date passed
Tanzania	Electronic and Postal Communications (Online Content) Regulations; Electronic and Postal Communications (Online Content) (Amendment) Regulations [made under The Electronic and Postal Communications Act of 2010]	2020; 2022
Turkey	Law No. 5651 on the Regulation of Publications Made in the Internet Environment and Combatting Crimes Committed through these Publications (including amendments up until October 2022)	2007
Venezuela	Constitutional Law Against Hatred, for Peaceful Coexistence and Tolerance	2017
Vietnam	Decree No. 72/2013/ND-CP of July 15, 2013 on the management, provision and use of Internet services and online information	2013
	Law on Cybersecurity and Decree 53/2022 Elaborating a Number of Articles of the Law on Cybersecurity of Vietnam	2018

Proposed laws and regulations

Country	Bill or draft bill (accessed September 2023)	Date proposed
Bangladesh	Telecommunication Regulatory Commission Regulation for Digital, Social Media and OTT Platforms [draft bill]	2021
Brazil	Draft Bill 2630/2020 Law on Freedom, Responsibility and Transparency on the Internet ('The Fake News Bill')[113]	2020
Canada	Proposed Approach to Online Safety (Discussion Guide and Technical Paper)	2021
Chile	Proposed Law to Regulate Digital Platforms	2021
China	Draft Guidelines on the Classification of Internet Platforms	2021
	Draft Guidelines for Implementing Subject Responsibilities on Internet Platforms	2021
	Draft Regulations on the Protection of Minors Online	2021
EU	Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse	2022
Iran	Cyberspace Users Rights Protection and Regulation of Key Online Services [bill]	2021
Israel	Social Networks Bill	2021
Nigeria	Bill for an Act to Make Provisions for the Protection from Internet Falsehood and Manipulation	2019
	Draft Code of Practice for Interactive Computer Service Platforms and Internet Intermediaries	2022

113 Analysis based on the original bill as proposed in 2020.

Country	Bill or draft bill (accessed September 2023)	Date proposed
Philippines	Act Prohibiting The Publication and Proliferation of False Content on the Philippine Internet [draft bill]	2019
Poland	Law of 2021 on the Protection of Freedom of Speech on Social Networking Sites [draft bill]	2021
Taiwan	Draft Digital Intermediary Service Act	2022
Thailand	Draft Decree Regulating Digital Platforms	2021
UK	Online Safety Bill	2022[114]

114 Analysis based on the June 2022 version of the bill, as amended in the public bill committee. The analysis does not cover changes made to the bill in late 2022 and 2023. (For the most recent version, see UK Parliament (2023), 'Online Safety Act 2023', https://bills.parliament.uk/bills/3137/publications.)

About the authors

Yasmin Afina is a researcher in the Security and Technology Programme at the United Nations Institute for Disarmament Research (UNIDIR) and a former research fellow with the Digital Society Initiative (DSI) at Chatham House.

Marjorie Buchser is executive director of the DSI, with expertise in economic, legal and policy issues relating to the digital economy, digital trade and artificial intelligence regulation, innovation and R&D.

Alex Krasodomski is a senior research associate with the DSI, leading efforts on digital public infrastructure, open-source sustainability and articulating, measuring and advocating for an internet compatible with democracy.

Jacqueline Rowe is policy lead at Global Partners Digital. She provides analysis and expertise to the Global Partners Digital team, projects and partners on issues including online hate speech, disinformation and platform governance.

Nikki Sun is an academy associate with the DSI, and her research explores the impact of emerging technologies on the future of work and labour relations.

Rowan Wilkinson is a programme coordinator with the DSI and the International Law Programme at Chatham House.

Acknowledgments

This research paper is based in part on data and analysis from Global Partners Digital (GPD), expert interviews and group discussions held under the Chatham House Rule, in collaboration with the Digital Society Initiative.

This publication is supported by data and analysis from GPD, a social purpose company working at the intersection of human rights and digital technologies, including platform regulation.

The authors wish to thank all the participants for their valuable contributions. The authors would also like to thank Meta for supporting this project. Thanks are due to GPD and the Chatham House publications team for their attentive assistance and guidance. Finally, they would like to thank the anonymous reviewers for their detailed feedback. The paper is much better for their involvement.

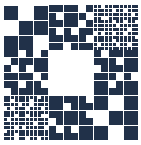
Cover image: A man looks at his mobile phone in front of a mural depicting people walking and using mobile devices at Tower Hill Underground station in London, 16 November 2023.
Photo credit: Copyright © Mike Kemp/In Pictures/Getty Images

ISBN 978 1 78413 593 5
DOI 10.55317/9781784135935

Cite this paper: Afina, Y. et al. (2024), *Towards a global approach to digital platform regulation: Preserving openness amid the push for internet sovereignty*, Research Paper, London: Royal Institute of International Affairs, https://doi.org/10.55317/9781784135935.

This publication is printed on FSC-certified paper.
designbysoapbox.com

Independent thinking since 1920



GLOBAL PARTNERS DIGITAL

The Royal Institute of International Affairs
Chatham House
10 St James's Square, London SW1Y 4LE
T +44 (0)20 7957 5700
contact@chathamhouse.org | chathamhouse.org

Charity Registration Number: 208223