Rights-First:

Principles for Digital Platform Regulation

Human Rights Law Centre

Acknowledgement of Country
The Human Rights Law Centre acknowledges the lands on which we work and live, including the lands of the Wurundjeri, Bunurong, Gadigal, Ngunnawal, Darug, and Wadawurrung people. We pay our respect to the Elders of these lands, waters, and skies, both past and present.

We recognise that Aboriginal and Torres Strait Islander people and communities were the first technologists and innovators on this continent, with deep knowledge systems that continue to shape our understanding of innovation, sustainability, land stewardship, and community care.

We recognise that this land always was and always will be Aboriginal and Torres Strait Islander land because sovereignty has never been ceded.

We acknowledge the role of the colonial legal system in establishing, entrenching, and continuing the oppression and injustice experienced by First Nations peoples and that we have a responsibility to work in solidarity with Aboriginal and Torres Strait Islander people to undo this.

With thanks
To everyone who assisted in the production of this report, including:

- Jackie Turner
- Australian Conservation Foundation
- This report was made possible by a grant from Mannifera. We appreciate Mannifera's commitment to addressing the harmful impacts of misinformation and disinformation on our democracy.

Please note:
This report aims to provide general information only. It is not intended to be legal advice and should not be relied upon as such. This report does not suggest any illegal or improper conduct on the part of any individual or organisation named.

Views expressed in this report are those of the named authors and do not necessarily reflect the official position of the Human Rights Law Centre, any affiliated organisations, or supporters.

This report was authored by David Mejia-Canales, Senior Lawyer at the Human Rights Law Centre, 4 October 2024.

Foreword

Since at least 2019, trans and gender diverse people have been subject to an incredibly vile and escalating attack on our rights, lives, and health care through a sophisticated and globally resourced disinformation campaign.

I have spent the last few years untangling the web of pseudo-scientific organisations, gender critical groups, and right-wing evangelists that make up the anti-trans lobby. With funding from right-wing think tanks in the US these groups have managed to muddy the water on proven and effective health care that is vital for trans and gender diverse people to thrive.

These campaigners have realised that it matters less what is correct and more what 'feels' correct. They have adapted to the platforms, stoking controversy and outrage to game the algorithm and make their accounts of any situation go viral.

As our public domains are increasingly becoming private spaces, our society takes for granted that the main platforms for civic engagement are administered primarily by massive private companies in the Global North. In addition to this, we have seen a massive concentration of power in news media over the last few decades.

These issues will affect all of us. Whether you are seeking accurate health information during the next pandemic, or considering how to vote on an important issue, a lot of the information that you base your decisions on will be delivered to you by the algorithms of Facebook, Instagram, Twitter, or TikTok. The sophisticated disinformation campaigns we are seeing to enable the genocide in Gaza, to attack trans rights, and to defeat the recent Voice campaign should serve as a wake-up call to us all.

More than ever, we need bold and visionary policies that take responsibility for the future we are building. Key to this is ensuring that corporations which have managed to monetise our attention spans, are forced to take responsibility for how they wield and profit from that incredible power.

I hope this framework is the beginning of a conversation about the interventions that governments need to take to ensure that digital spaces are administered in a way that protects all our communities, promotes human rights, and defends our democracy.

Jackie Turner
Trans Justice Project
transjustice.org.au/handbook

The Trans Justice Project runs training on combating anti-trans disinformation. For more information: transjustice.org.au/about-us/training

Executive Summary

Digital platforms like Meta, X, and Google must be regulated, and this regulation must be grounded in human rights law and principles.

Digital platforms have transformed how we communicate, share information, and connect with each other. However, they have also facilitated the spread of illegal and harmful content, disinformation, and hate speech, resulting in real-world harm to people and communities.

For example, in August 2024, misinformation and disinformation online stoked terrifying racist violence which lasted for six days in the United Kingdom, leading rioters to damage public buildings and set fire to hotels housing people seeking asylum.[1]

Australia once led the way in digital regulation. However, our current framework to prevent the spread of misinformation and disinformation relies on the digital platforms regulating themselves. Experience has demonstrated that self-regulation by powerful corporations is ineffective.

The supermarket voluntary Food and Grocery Code of Conduct (Code) was introduced to improve standards of behaviour in the industry and to govern the conduct of retailers and wholesalers towards their suppliers.[2] A review by the Treasury found that the Code was ineffective because of the power imbalance existing between large and powerful supermarket corporations and their smaller vendors and suppliers.[3] After repeated claims that large supermarkets were abusing their power, the Australian Government stepped in to regulate the sector by enforcing a mandatory code of conduct. As a result, corporations like Coles, Woolworths, and Metcash will be subject to multi-million-dollar penalties for serious breaches of the new code.

In the same way that we do not allow harmful industries like tobacco or gambling to regulate themselves, we should not allow digital platforms to do so either. Digital platforms have shown that they will continue to amplify and profit from harmful and illegal content on their services at the expense of our safety.

The European Union's (EU) Digital Services Act (DSA) provides a model for regulating digital platforms that Australia should follow.

The DSA requires platforms to assess the risks their systems and products cause, and also requires platforms to be transparent and accountable for reducing these risks or face significant penalties for non-compliance.

To ensure effective and balanced online regulation, protecting human rights must be at the core of any new regulatory framework, particularly as unregulated misinformation, disinformation and other harmful content is threatening the enjoyment of our human rights.

Harmful online content threatens fundamental human rights – including the right to be free from discrimination, the right to health, the right to participate in public affairs, and the right to vote – to name a few. Conversely, overly strict regulation that is not tethered to human rights principles can also infringe on users' rights – including the freedom of expression and the right to information. Striking the right balance is crucial to ensure online safety while upholding the rights and freedoms we all rely on.

By grounding regulatory efforts in human rights law, governments and regulators can develop clear and equitable rules to keep us safe online while also retaining the many benefits that digital platforms have brought to our lives. Human rights law also provides a well-established, legitimate, and globally recognised framework for managing and balancing rights, ensuring that regulations protect users' freedoms, while also addressing conflicts between competing rights fairly and equitably.

To achieve effective regulation of digital platforms, the Australian Government needs to legislate according to five key principles:

1. Digital regulation must be based on human rights law and principles;
2. Digital platforms should have a legal duty to make sure their products, systems, and services do not cause harm;
3. Content removal powers have a role to play in limited circumstances, but only as part of a broad, comprehensive regulatory framework;
4. Users should have control over how platforms collect and use their data; and
5. Court oversight is essential in a comprehensive regulatory framework.

Introduction

Australia was once a world leader in digital regulation, but we have quickly fallen behind.

The rise of internet technologies has greatly benefited our lives, including by amplifying diverse voices, enabling economic opportunities, democratising access to information and education, and providing new ways us to gather and connect.

The flipside is that these technologies have also enabled the rapid spread of illegal and abhorrent material like child exploitation content, hate speech, misinformation, disinformation and other harmful content. In turn, this has contributed to an erosion of trust in our democracy,[4] and is causing real harm to people and communities.[5] Children, young people, women, Aboriginal and Torres Strait Islander people, communities of colour, and LGBTQIA+ individuals are particularly at risk to harmful content, harassment, and hate speech online.[6]

Digital platforms are driving the spread of harmful and illegal content on an unprecedented scale. Large digital platforms in particular wield immense power over public discourse, amplifying and manipulating the information – and disinformation – that is shaping our decisions and beliefs. Worst of all, these platforms are profiting billions of dollars every year from fuelling the spread of harm and hate in our society.

That's why the Australian Government must prioritise strong and effective regulation of digital platforms.

Platforms must be accountable

The World Economic Forum's Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms (Toolkit: Typology of Online Harms) recognises the role users play in producing, distributing, and consuming harmful content, but it also acknowledges that technology itself facilitates behaviour that leads to harm.[7]

Digital platforms have consistently failed to rectify their use of addictive design features, recommender systems, invasive data harvesting practices, ineffective and vague content moderation policies, and opaque mechanisms for reporting misconduct or abuse online - the list goes on. Digital platforms will not voluntarily act to fix the systems that enable the spread of harmful content. These platforms lack a commercial incentive to regulate misinformation and disinformation because inflammatory content drives engagement which is ultimately boosting their profits.

In this context, digital platforms, the entities with the greatest ability and capacity to reduce the harmful content they host, must be made to take on more responsibility for mitigating the harms that they enable.

Australia was once a pioneer of digital platform regulation – we were the first country to legislate for online safety and to appoint an Online Safety Commissioner. We led the way in legislating negotiations between digital platforms and news outlets, and insights from the Australian Competition and Consumer Commission's Digital Platforms Inquiry Final Report continue to shape policy development nationally and abroad.

However, we are falling behind – and fast. Digital platforms have become too large, too influential, and have too much control over our lives, livelihoods, and our interactions with each other.

Image: Communities In Brighton protesting racist violence fuelled by misinformation – August 7, 2024 (Photo: Shutterstock/Edward Zorzi-Chapman).

Australia's current framework needs improvement

The Australian Government has conducted numerous reviews and inquiries into creating safer online environments. These include reviews of the Privacy Act 1998 (Cth) and the Online Safety Act 2021 (Cth) (Online Safety Act), as well as parliamentary inquiries into the influence of large online platforms, the role of social media in Australian society, and the impact of artificial intelligence.

In response, the Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024 (Cth) (the Bill) was introduced into the Australian Parliament on 12 September 2024. The Bill seeks to tackle the growing issue of harmful misinformation and disinformation on digital platforms. The Bill was released as this report was being finalised, the information about the Bill set out below is subject to change, as the Bill makes its way through Parliament.

The Bill proposes amendments to the Broadcasting Services Act 1992 (Cth) with the aim of increasing digital platform transparency and accountability by, among other things, requiring platforms to self-regulate at first instance, empowering the Australian Communications and Media Authority to create digital platforms rules in the event that self-regulation is ineffective, and requiring platforms to publish media literacy plans outlining the steps a platform is taking to combat misinformation and disinformation on their service.

The Bill's reliance on industry self-regulation is problematic. Digital platforms currently profit from the misinformation, disinformation, and other harmful content on their services. Without strong regulatory drivers to meaningfully address the problem, platforms have little incentive to meaningfully change their approach unless forced to.

The Bill defines a media literacy plan as a plan that outlines the measures a digital platform will take to enable users to identify misinformation and disinformation on the platform, including the ability for users to identify the source of content disseminated on the platform.[8] The intent is supported, as providing users with tools to better understand the provenance of online content is important. However, media literacy alone cannot compete with the pervasive and rapidly evolving nature of misinformation, especially in the era of generative artificial intelligence.

Additionally, the Bill fails to clarify how media literacy plans will be accessible to people who are not fluent in English, children, young people, and some people with disability, for whom such strategies may be inappropriate. The Bill does not currently require these plans to be in different languages or in an accessible or plain language format, which could limit their effectiveness.

Furthermore, the lack of a prescriptive mechanism indicates that there is no standardisation of media literacy plans across platforms leaving each platform to potentially adopt different media literacy approaches, leading to inconsistency and confusion among users.

At the time of writing, the Bill had been referred to the Senate's Environment and Communications with a final report due back from the Committee on 25 November 2024. The public was given only seven business days to provide feedback on the Bill, falling short of the human rights principles of participation and inclusion, which are essential for realising human rights.

Moreover, the Bill is not fully anchored in human rights law, making it a missed opportunity to not just protect the freedom of expression but all other fundamental rights as required under international human rights standards.

While the Bill is yet to be debated in Parliament, it is clear that the current proposal overemphasises self-regulation and does not take a comprehensive rights-based approach to combatting misinformation, disinformation and other harmful online content.

Definitions of key terms

For the purposes of this report, we adopt the following definitions:

Digital platform	A system or service that enables users to interact, share, search for, or receive content, exchange information or conduct transactions. Digital platforms include (and is not limited to) social media networks, e-commerce sites, online marketplaces, search engines, and hosting services.
Disinformation	False, inaccurate or misleading information that is designed, presented and promoted to intentionally deceive or secure economic or political gain, and which may cause public harm.
Misinformation	Unintentionally false or inaccurate information spread without malicious intent.
Online harm	Online harm encompasses a range of negative impacts resulting from digital interactions and content. We have relied on the Toolkit: Typology of Online Harms[9] which places these harms into three categories: » content harms, which involve harms arising from the production, distribution, and consumption of harmful online content;[10] » contact harms, which occur through interactions with others online;[11] and » conduct harms, which stem from harmful behaviours enabled by technology and digital platforms.[12] The EU's DSA expands on this understanding by including harms associated with exposure to illegal content online, harms to fundamental human rights, harm to democratic processes, and harm to public health and individual wellbeing.[13] Together, this understanding of online harm reflects the broad scope of harms that can be caused online and the need for comprehensive strategies to address and mitigate its various forms.
Recommender system	Digital platforms choose the content we see, the products we are suggested and the services we are offered based on our preferences, data profiling, past behaviour, and interests. This information filtering system, or "recommender system", is designed to keep us engaged and on the platform for as long as possible. An example of one of these systems is TikTok's "For You Page".

Definitions of human rights principles:

Human rights principles are the foundational guidelines that ensure human rights are protected and realised in practice. These principles establish a framework for the universal respect, protection, and fulfillment of rights.[14] They include:

Universality and inalienability	Human rights are inherent to every person, regardless of their location or circumstance. They cannot be voluntarily surrendered or taken away.
Indivisibility	All human rights – whether civil, cultural, economic, political, or social - are of equal importance and interrelated. No right holds greater value than another, as they collectively uphold the dignity of every human being.
Interdependence and interrelatedness	The fulfillment of one human right often depends on the realisation of others. For instance, to fully realise the right to peaceful protest other rights must also be realised and protected, like the freedom of association, the freedom of expression, and the freedom of movement.
Equality and non-discrimination	Every individual, regardless of race, gender, ethnicity, age, or other status, is entitled to the same rights and dignity at all times. Discrimination in any form violates the core principle of equality that human rights are built upon.
Participation and inclusion	All people have the right to actively participate in the decisions and processes that affect their civil, economic, social, cultural, and political rights. Meaningful participation is essential for realising human rights.
Accountability and rule of law	States and other duty-bearers are responsible for upholding human rights standards. When rights are violated, those affected have the right to seek redress through legal mechanisms, ensuring that justice and accountability are maintained.

Principle 1

Digital regulation must be based on human rights law and principles.

Human rights law provides a globally accepted and legitimate framework for protecting people's rights, whether online or offline, and offers a recognised method for managing conflicts between rights when they arise.

Digital spaces have become an integral part of our daily lives, making their protection essential. Our human rights do not vanish the moment we go online to connect with others or to seek information. Indeed, human rights law requires that the rights that we enjoy offline must be equally guaranteed in the digital world.[15]

While human rights obligations are primarily imposed on states, the United Nations' Guiding Principles on Business and Human Rights make it clear that digital platforms, like all business enterprises, have a responsibility to respect human rights wherever they operate.[16] This responsibility exists alongside Australia's legal obligation to guarantee and protect the human rights it has committed to under international law, including the obligation to protect everyone within Australia's jurisdiction from human rights abuses.[17]

To create safer online spaces, digital platform regulation must be rooted in human rights law. Human rights law provides a universal framework to protect fundamental rights and freedoms through internationally recognised laws, principles, and standards. Grounding regulation in this framework ensures that any restrictions on human rights are lawful, strictly necessary to achieving a legitimate objective, and proportionate.

Furthermore, by aligning regulations with human rights principles, we can effectively address harmful online content – such as misinformation, disinformation, and hate speech–while balancing the rights of all users. This approach is crucial as harmful online content can significantly undermine the enjoyment of various human rights, including:

The right to freedom of thought and conscience

The right to freedom of thought and conscience is contained in Article 18 of the International Covenant on Civil and Political Rights (ICCPR), to which Australia is a signatory, and gives all people the right to think freely and hold beliefs or opinions based on their conscience, religion, or other convictions. Not only are these rights the foundation for every free and democratic society,[21] they form a basis for the full enjoyment of other human rights.[22]

The freedoms of thought and conscience are absolute, meaning that the ICCPR does not permit them to be restricted, even in times of national emergency.[23] However, the UN Human Rights Committee has noted that to protect the freedom of thought and conscience, countries must protect people from undue interference from individuals or organisations which could hinder their ability to enjoy these freedoms.[24]

Despite this, digital platforms rely on powerful algorithms that prioritise viral, emotionally charged content which boosts user engagement and in turn increases their profits.[25] Because these algorithms are often opaque, users might think they are being served objective information, but in reality, the content they are seeing is shaped by many hidden factors, including the commercial imperatives of the platforms, which in turn can impact or influence how users form and develop their opinions.

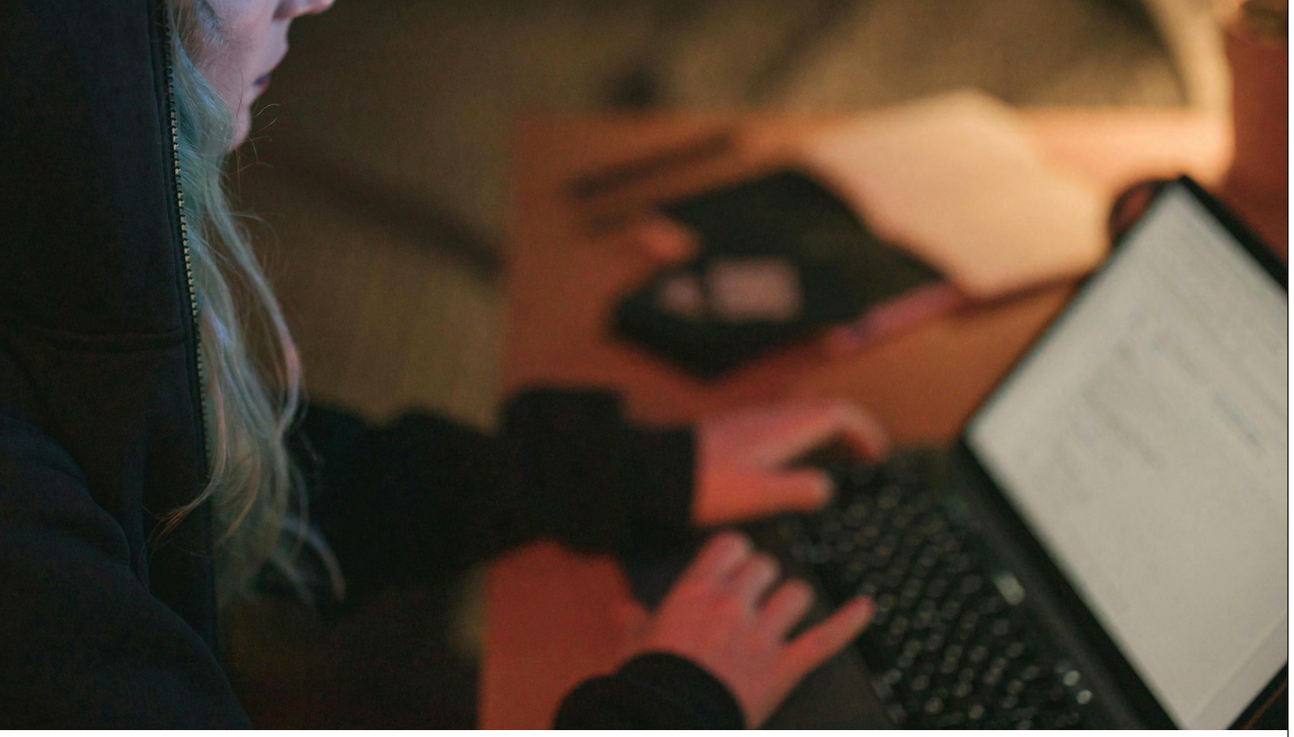
A 2015 study found that even the way internet search results were ranked on a page could change the voting preferences of undecided voters by 20 percent or more.[26]

The right to freedom of expression

Article 19 of the ICCPR guarantees the right to freedom of expression. The right allows people to seek, receive, and share information and ideas through any medium.[18] The freedom of expression allows people to share all types of information, not just information that is true, albeit subject to limited restrictions.

Under Article 19, any restriction on the freedom of expression must meet the following criteria: it must be necessary and proportionate, clearly defined by law, and aimed at protecting the rights or reputation of others, or safeguarding national security, or public order, or public health, or public morals.[19]

Human rights law offers a universally accepted framework for balancing the freedom of expression with other rights when conflicts arise. Effective regulation of digital platforms must be grounded in this legal framework to maintain this balance. Particularly as the freedom of expression must be interpreted in relation to other human rights, which ensures that one right does not unduly undermine or conflict with the protection of other rights. For example, human rights law requires that Article 19 be read alongside Article 20 of the ICCPR, which mandates that states must prohibit propaganda for war and speech that advocates national, racial, or religious hatred that incites discrimination, hostility, or violence.[20]



A misguided interpretation of the freedom of expression (including by free speech absolutists) has been weaponised to avoid accountability for the harms caused by abuses of the freedom of expression. However, it is also important to recognise that over-regulation, especially when it is not grounded in human rights law, can itself pose a legitimate threat to freedom of expression. Excessive or poorly designed regulations may suppress free speech, stifle public debate, or lead to censorship that is neither necessary nor proportionate.

A regulatory regime grounded in human rights law protects users from both extremes – ensuring accountability for harmful content while safeguarding individuals from undue restrictions on their rights. This balance is crucial to maintaining a healthy, open, and free digital space.

The right to privacy

Article 17 of the ICCPR guarantees the right to privacy, protecting individuals from arbitrary or unlawful interference with their privacy, family, home, or correspondence, and from unlawful attacks on their reputation. This right ensures that personal information, communications, and private activities remain protected from undue intrusion. The right to privacy is essential for preserving human dignity and autonomy, allowing people to freely express themselves, form relationships, and engage in private matters without fear of unjust surveillance or exposure.[27]

Currently, digital platforms are falling well short of meeting these minimum standards – contrary both to their own obligations under business and human rights frameworks, and Australia's legal duty to protect people from human rights violations. Personal data online is collected, sold, and used extensively by platforms in opaque algorithmic processes, often without a user's awareness or full consent. This widespread use of personal data threatens our right to privacy while generating significant profits for digital platforms, leaving individuals with limited control over what happens to their information.[28]

The right to participate in public affairs and the right to vote

Article 25 of the ICCPR guarantees the right of citizens to participate in public affairs, either directly or through freely chosen representatives, and the right to vote in genuine and free elections. This right is fundamental to a functioning democracy, ensuring that citizens have a say in the governance of their country and that elections are conducted fairly and transparently.

The United Nations Human Rights Committee has emphasised that to protect this right, States must ensure voters can form opinions independently, free from violence, coercion, or manipulation.[29] The Committee also highlights that the free exchange of information and ideas between citizens, candidates, and elected officials is essential for a functioning democracy.[30]

However, digital platforms and their algorithms are distorting public debate around elections and referendums, including by spreading disinformation, about voting and electoral processes which can undermine democratic participation by discouraging voting or campaigning, or by preventing users from accessing accurate information on which to form their views and beliefs.[31]

Likewise, the collecting and trading of personal data to manipulate voting behaviour through highly targeted advertisements can also influence public debate or how they form their opinions.[32] For example, a political consulting firm could purchase or collect data on users that is held by a digital platform, which identifies individuals who are likely to be undecided voters to target them with misleading advertisements about an opponent candidate's political position.

In conclusion

The need for digital regulation grounded in human rights law is vital to protecting our fundamental freedoms and ensuring a fair and safe online environment. As digital platforms continue to play an increasing role in shaping public discourse and social interaction in Australia, a human rights framework provides the only legitimate and globally accepted means to address the challenges posed by harmful content while safeguarding our rights and freedoms.

By grounding regulation in human rights law, we can protect individuals, promote accountability, and foster safer, fairer digital spaces for all.

Image: People in Brisbane lining up to vote.

Case Study: Digital platforms are putting people at risk: The experience of children online

Children and young people are regularly exposed to illegal, harmful, abusive, and exploitative content on digital platforms.[33] The 2023 Australian Online Safety Survey reveals that children are often victims of repeated unwanted contact, cyberbullying, and exposure to sexually inappropriate material, with these experiences being far more intense for children and young people with disability, or who identify as LGBTQIA+.[34] These harms not only threaten the safety of children online, but also their overall sense of security and well-being, as they navigate increasingly hostile digital environments.

Yet digital platforms have consistently resisted implementing truly meaningful safeguards to ensure their safety. In 2023, X (formerly known as Twitter) was fined $610,500 by Australia's eSafety Commissioner for failing to explain how it was addressing child abuse on its platform amid revelations that X had reduced its efforts to detect illegal material.[35]

Case Study: Digital platforms are fuelling serious real-world harm: Facebook in Myanmar.

In August 2017, Myanmar's military launched a campaign of ethnic cleansing and genocide against the Rohingya Muslim population in northern Rakhine State, where thousands of people were killed, tortured, and sexually assaulted. The violence caused over 700,000 Rohingya people to flee Myanmar into Bangladesh. This violence was fuelled by a long history of discrimination against Rohingya people and the deliberate spread of anti-Rohingya hate speech and disinformation on Facebook.[36]

In the years leading up to the atrocities, Facebook became a dominant platform in Myanmar, often referred to there as the internet itself. Facebook's algorithms systematically amplified hateful content against the Rohingya people, with military-linked actors and radical nationalist groups using the platform to incite violence and dehumanise Rohingya people. The spread of disinformation, portraying Rohingya people as threats to national security, directly contributed to the escalation of violence.[37]

Despite multiple warnings from local activists and international human rights groups, Facebook failed to adequately address the growing risk of violence. Internal documents revealed that the platform's content moderation was insufficient, and its algorithms prioritised inflammatory content to maximise user engagement and their advertising revenue.[38]

The United Nations and various human rights organisations, including Amnesty International, have criticised Facebook's role in enabling genocide. Although Facebook's parent company Meta has acknowledged its failures, it has not yet provided meaningful remedies. Legal cases seeking reparations from Meta are ongoing, and Rohingya people continue to demand justice for their communities.[39]

Image: Election Day at The Newseum's Campaign 2016 exhibit. Photo by Lorie Shaull, St. Paul, United States.

Case Study: Election interference: Russia's Internet Research Agency

The Internet Research Agency (Agency), a Russian company that operated with the direct approval and endorsement of Vladimir Putin,[40] engaged in online propaganda to influence public opinion and elections via the spread of misinformation and disinformation (among other things). The Agency existed to sow discord and division in countries that were not aligned with Russia's geo-politics and interests, as well as to undermine confidence in elections and democratic institutions. It did this by amplifying existing social divisions.[41] The Agency had approximately 400-600 staff at any one time.[42]

The inherent features of digital platforms, like data tracking and profiling of users, gave the Agency the ability to target specific audiences.[43] One of the Agency's key performance metrics was to transform online behaviour and beliefs into real world action, including provoking real-world violence between people and inflaming racial tensions in the United States of America and across Europe.[44]

Between 2014 and 2017, the Agency attracted 3.3 million-page followers on Facebook who generated 76.5 million engagements.[45] These included 30.4 million shares, 37.6 million likes, 3.3 million comments, and 5.2 million reactions.

Facebook estimated that the Agency's content was seen by 126 million Facebook users.[46] During that same period the Agency controlled 3,841 accounts on Twitter which generated 73 million engagements from approximately 1.4 million people. On Instagram, the Agency held 12 accounts with over 100,000 followers each, with its top accounts had up to 'tens of millions of interactions.'[47] The Agency also owned 17 YouTube channels and produced 1,107 videos.[48]

Principle 2

Digital platforms should have a legal duty to make sure their products, systems, and services do not cause harm.

A duty of care is a legal obligation to ensure the safety and well-being of others by taking reasonable steps to prevent harm or injury.

A duty of care on digital platforms is essential to protect users from harm, but it is equally critical to ensure that platforms are not allowed to regulate or co-regulate themselves. When left to their own devices, platforms tend to prioritise profit over well-being, as they benefit from the very systems that enable harmful content and behaviours. Allowing platforms to regulate themselves would also create conflicts of interest, undermining efforts to genuinely protect users.

To effectively prevent harm, regulation must require platforms to proactively design their systems to prevent harm in the first place. Imposing a duty of care on platforms would make digital platforms accountable for the harms they enable and profit from.

A duty of care should include requirements for digital platforms to:

1. Uphold and protect the fundamental human rights of their users;
2. Undertake comprehensive risk assessments to identify and analyse risks stemming from their products and systems;
3. Address identified risks though effective risk mitigation measures; and
4. Open up their assessment and mitigation measures for scrutiny by third parties to enable independent testing and verification.

There must also be effective mechanisms for redress for harms caused by a digital platform's breach of their duty of care.

EU Digital Services Act

The EU's DSA takes a risk-mitigation based approach to regulating digital platforms, underpinned by extensive transparency and accountability requirements that are imposed on the platforms. These include:

- **Comprehensive risk assessments:** Requirements for large digital platforms to undertake annual risk assessments to identify significant systemic risks arising from the functioning and use of their services. This encompasses algorithms, recommender systems, content moderation systems, user terms and conditions, advertising systems and data-related practices.[49]
- **Risk mitigation:** Where risk assessments undertaken by very large platforms identify systemic risks, these platforms are required to implement "reasonable, proportionate and effective mitigation measures."[50]

- **Transparency measures**:
  - » Risk assessments and mitigation measures are subject to independent audits;[51]
  - » Platforms must provide annual public transparency reports which are heavily prescriptive;[52]
  - » Large platforms must provide advertising repositories, which are openly searchable, and include details regarding the advertisements on their platforms, including information about who paid for the advertisement;[53]
  - » On request from the regulator, platforms are required to provide independent researchers with access to platform data to detect, identify and understand the systemic risks that have been reported by the platforms and to assess the adequacy, efficiency and impacts of the platforms' risk mitigation measures.[54]
- **Regulator oversight:** The DSA empowers the regulator, the European Commission, with strong enforcement powers, including the ability to impose significant penalties for violations. These include, applying fines of up to 6% of a platform's worldwide annual turnover for breaches of DSA obligations or applying periodic penalties of up to 5% of the average daily worldwide turnover for each day that a platform delays in complying with remedies.[55]

UK Online Safety Act 2023

The United Kingdom's Online Safety Act 2023 (UK) also takes a similar risk-mitigation approach to regulation while imposing duties on social media companies and search engines to protect users, particularly children, from online harm.[56] 60 Key provisions of the UK's Online Safety Act include:[57]

- **Risk Reduction:** Platforms must implement systems to minimise the risk of illegal activity on their services and actively remove illegal content when it appears. The strongest protections are for children, requiring platforms to block harmful and age-inappropriate content.

- **Transparency and Control:** Platforms must be transparent about the harmful content on their services and provide tools for users to control what they see online.

- **Regulator Oversight:** The Office of Communications, the independent regulator, is empowered to enforce the Online Safety Act, set safety standards, and ensure platforms comply with new duties. Significant penalties can be imposed for violations, including fines up to £18 million (over AUD$35 million) or 10% of a platform's global revenue- whichever is the highest.

- **Global Reach:** The Act applies to any service accessible in the UK, even if the company is based outside the country.

The UK's Online Safety Act introduces new criminal offences, like cyberflashing[58] and encouraging self-harm, while mandating that platforms address harmful algorithms that increase users' exposure to illegal or harmful content.[59]

Mitigating risk is a practical way to reduce potential harms while still allowing room for innovation and growth. This approach applies the strongest protections where the risks are highest, helping to build accountability and trust across industries without slowing progress. The EU's Artificial Intelligence Act (AI Act) is another good example of this strategy in action.

EU Artificial Intelligence Act

The AI Act uses a risk-based approach to regulate artificial intelligence, offering a clear legal framework that addresses the risks of AI systems while encouraging innovation and trust. Key features of the AI Act include:

Risk Management: The AI Act categorises AI systems into four levels of risk: unacceptable, high, limited, and minimal/no risk. Unacceptable-risk AI, such as systems that pose clear threats to fundamental human rights or public safety, are banned. High-risk AI systems, which include AI applications in areas like employment, critical infrastructure, and law enforcement, are subject to stringent requirements. These requirements include risk assessments, data quality measures, and human oversight to ensure that AI systems do not cause harm or discriminate unfairly.

Transparency and Accountability: For high-risk AI systems, providers must maintain detailed documentation and ensure transparency in the system's purpose and functioning. This is coupled with requirements for monitoring, reporting of serious incidents, and post-market surveillance. Importantly, AI-generated content, such as deepfakes or chatbots, must be labelled as such, ensuring users are aware when they interact with AI.

Enforcement and Oversight: The AI Act grants the EU AI Office, established in 2024, significant enforcement powers to ensure compliance with the regulation. The Office, working with national authorities, oversees the market and ensures ongoing quality and risk management for AI technologies. Failure to comply with the AI Act can result in significant penalties, ensuring that AI providers are incentivised to adhere to the regulations and are held accountable when they do not.

In conclusion

Imposing a duty of care on digital platforms represents a crucial step towards ensuring that technology and digital platforms serve the public good while minimising harm.

The EU's DSA and AI Acts and the UK's Online Safety Act demonstrate how a risk-based regulatory approach can foster accountability, enhance transparency, and promote user safety. These frameworks not only address existing risks but also anticipate and mitigate potential future harms, setting a high standard for digital regulation globally.

As governments around the world adopt similar duties of care to protect users and prevent harm, Australia has the opportunity to follow suit. By leveraging existing international frameworks as blueprints, our government can develop and implement robust regulations that ensure digital products and services are designed with safety and fundamental rights at their core.

The path forward is clear: Australia must step up and enact comprehensive legislation that mirrors the proactive and risk-based approaches seen in around the world, safeguarding users and fostering a trustworthy digital environment.

Case Study: Misinformation and disinformation are impacting our right to health: COVID

Coordinated disinformation campaigns pose a serious threat to our right to health, as highlighted during the COVID-19 pandemic.[60] During the pandemic, vaccines were one of the biggest topics of misleading health claims.[61]

It's estimated that misinformation and disinformation about COVID-19 vaccines alone cost the United States' economy up to USD $300 million per day due to hospitalisations, long-term illness, lives lost and economic losses from missed work.[62]

Image: "No New Normal Rally", Vancouver, 22 November, 2020 (GoToVan).

Principle 3

Content removal powers have a role to play in limited circumstances but only as part of a broad, comprehensive regulatory framework.

Regulation premised on content moderation or removal is reactive, inefficient, and too slow to keep up with the vast amount of harmful online content hosted on digital platforms.

Empowering an independent regulator to mandate the removal of harmful online content, like child exploitation material and image-based abuse, is important for fostering safer digital environments. However, a regulatory framework premised on content policing will soon become inadequate and can lead to other issues elsewhere. Overly strict content regulations may inadvertently suppress the freedom of speech or restrict access to legitimate information, while too lenient rules could allow harmful content to spread unchecked, undermining user safety and trust.

Australia's system, under the Online Safety Act, allows the e-Safety Commissioner to issue removal notices for harmful content, including cyberbullying, image-based abuse, and illegal material. While this is an important mechanism, it is not a long-term solution. Basing a regulatory approach on monitoring for harmful content creates a reactive "whack-a-mole" approach, where harmful content is taken down only to reappear elsewhere.[63]

Effective regulation of digital platforms must focus on preventing harm before it occurs, rather than relying on the removal of harmful content after the fact. By focusing on preventive measures, regulators can address the root causes of harmful content, rather than just reacting to its manifestations. Consequently, platforms should be held accountable for the risks they create by hosting, amplifying, and profiting from such content.

In addition, it is just as important that platforms ensure their own content moderation practices – be they manual or automated – align with human rights principles. Platforms must be required to implement transparent moderation systems that avoid arbitrary or overly restrictive decisions that could limit lawful speech.

Platforms should also be required to fully explain to their users, in clear and plain language, the reasons for removing or restricting content or accounts. Moreover, users must have access to simple, accessible mechanisms to seek redress and resolve disputes when they contest such decisions with the platforms.

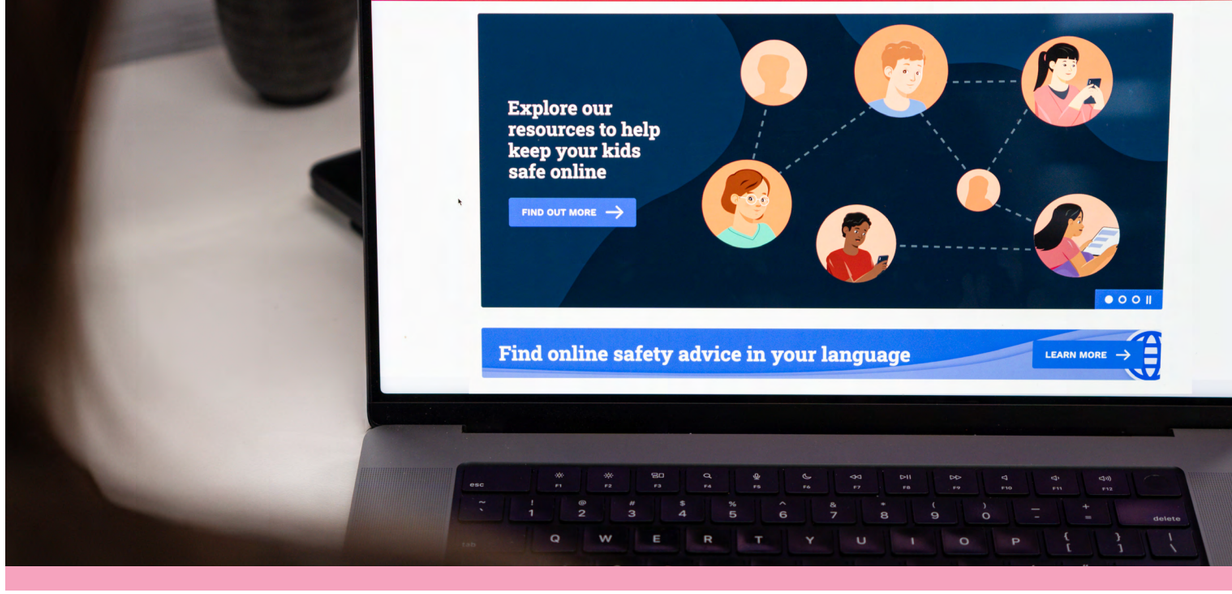


Image: A laptop displaying the Australian eSafety Commissioner's website.

The DSA offers a blueprint for increasing transparency and accountability in content moderation. It mandates clear communication about content decisions and provides robust mechanisms for users to challenge and appeal such decisions. Under the DSA:

- Platforms must publish clear and specific statements of reasons for their content moderation decisions, including detailed explanations when removing or restricting content and closing accounts.[64]
- The DSA's Transparency Database, launched by the European Commission, makes all statements of reasons provided by platforms for their content moderation decisions accessible to the public, enhancing scrutiny and accountability.[65]
- Users are empowered to seek redress for content moderation decisions through various means, such as internal complaint-handling mechanisms, out-of-court dispute resolution, and judicial processes.[66]
- Platforms are required to implement measures for easier reporting of illegal content, including mechanisms for users to flag such content.[67]

In conclusion

Australia has made significant progress with the e-Safety Commissioner's content removal powers, but further development is necessary to enhance online safety. A robust regulatory framework must go beyond just "whack-a-mole" content removal powers and instead prioritise harm prevention, by among other things, legislating a duty of care for digital platforms.

Case Study: Digital platform transparency in action: The DSA Transparency Database

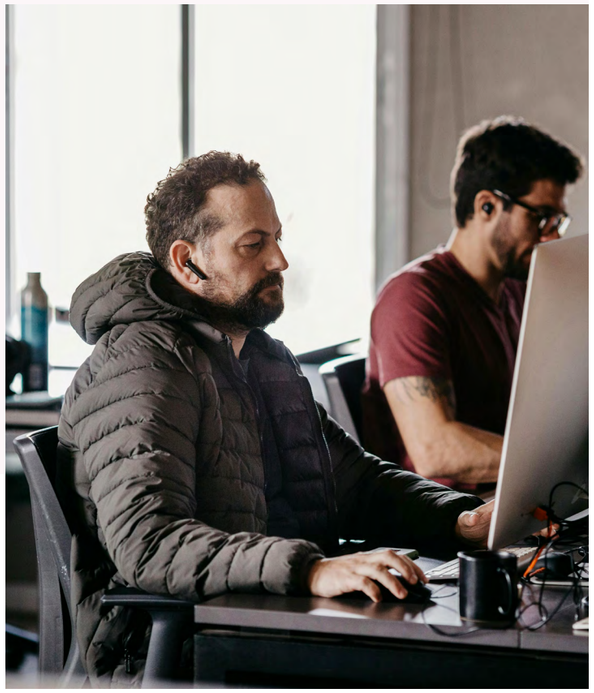
The DSA mandates increased transparency in content moderation by requiring digital platforms to provide clear and specific explanations, known as statements of reasons, when they remove or restrict access to user content. This requirement is detailed in Article 17 of the DSA and aims to empower users by clarifying the reasons behind a platform's content moderation decisions.

To enhance this transparency, Article 24 (5) of the DSA mandates that all statements of reasons from online platforms be submitted to the DSA Transparency Database. This database is publicly accessible and machine-readable, providing a valuable resource for researchers and the general public to monitor and scrutinise a platform's content moderation practices and the types of content they are taking action over. A statement of reasons includes details on the type of restriction imposed by the platform over the offending content, the grounds for the decision, and the context of the moderation action.

As of six months prior to September 2024, the database had accumulated over nine billion statements of reasons. The most frequently reported violations involve issues related to content that falls outside the scope of a platform's service, illegal or harmful speech, and unsafe or illegal products. The predominant restrictions imposed by platforms include disabling access to the offending content, content removal, and suspending or terminating accounts. Notably, 59% of the database entries pertained to decisions made through platform's

On 27 July 2024 alone, the database recorded significant daily activity, including 1,268,499 statements related to illegal or harmful speech, 466,037 concerning unsafe or illegal products, 159,187 related to violence, 215,976 statements for scams or fraud, 48,676 for self-harm, 69,913 for negative effects on civic discourse or elections, 3,669 statements for non-consensual behaviour, and 6,960 statements for risks to public security.[69]

Just the availability of the statements in the database alone demonstrates how robust regulations can force platforms to publicly share details about the risks present in their products and systems. This level of transparency provides a clearer view of the challenges and dangers tied to online platforms, encouraging a more informed dialogue on how to address these issues.



Case study: Digital platforms must be transparent in account and content decisions: The Australian Conservation Foundation

In August 2024, the Australian Conservation Foundation's X account (@AusConservation) was suspended, allegedly due to targeted attempts by pro-nuclear groups seeking to suppress critical commentary about the real dangers of nuclear energy in Australia. The suspension occurred following ACF's increased posting of genuine, evidence-based content critical of nuclear energy, in response to the Coalition's proposal to lift Australia's nuclear energy ban.[70]

ACF has not posted any content that could be considered in breach of X's rules. Like any major not-for-profit organisation, we use our X account for advocacy, to share news, views and factual content.

ACF has more than 34,000 followers on X and we have, until recent times, enjoyed mostly positive and productive engagement via this platform. In recent months, following the release of the Coalition's nuclear policy, ACF has been posting more factual information about nuclear power related to the time, cost, delay and its unsuitability for Australia.

We believe the unfair suspension of our account was directly connected to this content. X's Support Team contacted ACF with an apology and a brief explanation for the suspension, citing that we were mistakenly flagged by their automated system as a spam account.

This confirms our suspicions that we have been the target of a critical mass of false reports. Such tactics are a concerning development in the Australian social media landscape.
In recent weeks the removal of evidence-based information on renewables, nuclear power and electric vehicles has been reported across X, Facebook and TikTok.

We are pleased that our X account is now back up and running. While those who tried to shut us down were temporarily successful, this has ultimately backfired and has resulted in ACF gaining more than 3000 new followers.

We appreciate the apology and swift reinstatement of our account by X and look forward to our ongoing engagement on the platform.

It is clear that we are facing ongoing attempts to silence us. And we anticipate this will happen again.

Jane Gardner
Director of Engagement
Australian Conservation Foundation
acf.org.au

Image top right: "I love my grandchildren - Reef Not Coal snap action" (Julian Meehan).

Principle 4

Users should have control over how platforms collect and use their data.

All users must be able to know, easily and simply, how their data is being used so that they can maintain their autonomy and privacy online.

Providing all users with comprehensive control over their data is crucial, as user data, including personal or sensitive data, is a commodity that digital platforms sell to advertisers, and to other companies for data profiling. Data profiling – the use of personal information to create detailed user profiles for advertising or tracking – should not be the default privacy setting for anyone.

Digital platform users should be required to actively opt in if they wish to be profiled and tracked. Opting in should be a freely made, informed choice by an adult user that can be revoked easily at any time.

Anyone under 18 years of age should not be able to opt in to data profiling at all.

Under the DSA, digital platforms must offer greater transparency and control over the content their users see and how their data is used. Users must be able to opt out of personalised recommendations, with larger platforms required to provide a non-personalised content option by default.[71]

Digital platforms must also clearly label advertisements on their platforms as such, and provide – in real time – clear information about the advertisement, including information about what is being advertised, who paid for it, and the parameters that were used to determine why an advertisement was shown to a particular user.[72] Some large platforms are also required to maintain an advertising repository detailing all paid advertisements on their service.[73]

The DSA prohibits targeting users with advertisements based on the profiling of sensitive user data, like a user's religious beliefs or their sexuality.[74] Under the DSA, advertisements cannot be targeted towards minors, and platforms accessible to minors must guarantee their privacy, security, and wellbeing when using the platform, including by mandating special privacy and security settings by default.[75]

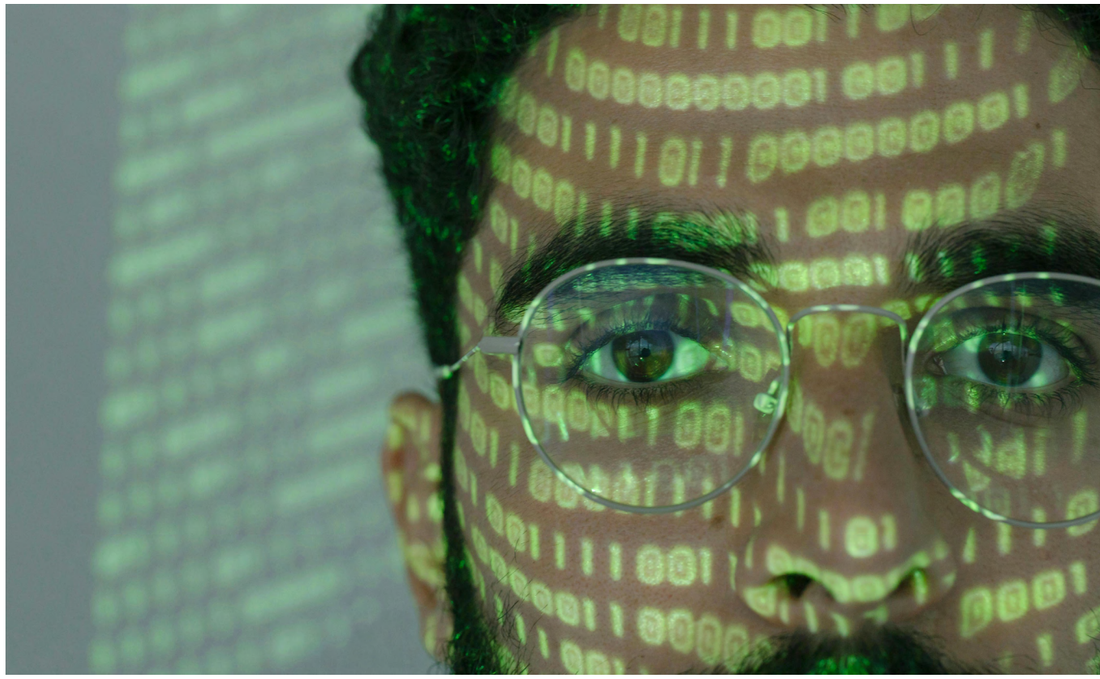
The DSA also prohibits the use of "dark patterns".[76] Dark patterns are design choices that manipulate or trick users into taking actions they might not otherwise take. Banning dark patterns can ensure that a decision a user makes, like opting out of profiling, is an informed, deliberate choice.

Furthermore, the DSA mandates that platforms provide clear user terms and conditions. If their service is directed towards minors, these terms and conditions must be easily understandable to that age group. The terms and conditions must also explain the main parameters of a platforms' recommender systems.[77]

In conclusion

Ensuring that all users can easily understand and control how their data is used is fundamental to protecting their autonomy and privacy online. The current practice of default data profiling and tracking by digital platforms undermines users' control over their personal information and allows their data to be exploited for commercial gain without their explicit consent.

The DSA offers a comprehensive framework for addressing these issues, emphasising transparency, user choice, and the prohibition of manipulative practices. Australia should adopt similar regulations that empower users with clear, straightforward choices about how platforms use their personal data, while also giving them the ability to opt out of profiling.



Principle 5

Judicial oversight is essential in a comprehensive regulatory framework.

Court oversight is essential for good digital platform regulation, particularly in resolving complex regulatory decisions.

Courts are established and trusted bodies that are designed to consider and resolve intricate legal questions fairly and consistently. Furthermore, the judiciary conduct their work publicly, providing an extra layer of transparency and scrutiny, which in turn reinforces trust in the regulatory process.

Judicial oversight ensures that significant or contentious decisions made by regulators can be challenged. By making these decisions open to judicial review, the integrity of the regulatory process is reinforced, acting as a safeguard against regulator overreach or misuse of power.

However, if courts are to have such oversight, it is important that rules and processes are in place to prevent well-resourced litigants, like large social media companies, from weaponizing the courts through endless litigation. A balance must be struck between ensuring courts can resolve complex issues without them being misused by powerful actors.

The standards set by regulators for digital platforms should be subject to judicial review. This review should be able to assess how these standards are applied in practice. The DSA has established a thorough enforcement framework, giving the Commission broad investigative and sanctioning powers to ensure compliance, while also requiring court oversight over significant regulatory decisions.

Under the DSA, the Commission plays a central role in supervising and enforcing compliance, particularly among large digital platforms. If the Commission suspects a breach of the DSA it can initiate investigations, request information, conduct interviews with, or inspect the premises of, digital service providers.

Importantly, the Commission's decisions are themselves subject to the review and oversight of a range of courts. The Court of Justice of the European Union (CJEU) has unlimited jurisdiction to review the Commission's decisions to impose fines or penalty payments.[78] The CJEU is able to cancel or reduce any fine or penalty payment imposed by the Commission, but may also increase the fine if the Court deems it to be inadequate.[79] The imposition of a fine can also be appealed before EU courts.[80]

The Commission's decision to inspect a platform's premises or impose interim measures may also require judicial authorisation from a Member State's courts.[81] Additionally, the CJEU is able to review a decision of the Commission to conduct an inspection of the premises of a very large online platform or online search engine.[82] In cases of severe non-compliance, the Commission may seek orders from a judge in an EU Member State to temporarily suspend a platform's services.[83]

The DSA also mandates that the Commission inform platforms of their rights when decisions are made that affect them, including mechanisms for administrative complaint handling and avenues of judicial redress.[84] In this regard, the DSA simultaneously provides for oversight of the Commission while also empowering affected parties to understand their rights of review and appeal, thereby ensuring the integrity and accountability of the Commission's decisions.

In conclusion

In Australia, it is crucial that any stronger regulatory approach strikes the right balance between providing adequate enforcement powers for regulators and ensuring appropriate checks and balances, including judicial oversight.

Australia's courts must play a key role in overseeing digital platform regulations, but safeguards must be in place to prevent deep-pocketed corporations from abusing the legal system. By balancing judicial review with protection against litigation abuse, Australia can ensure that its regulatory framework is both effective and fair.

Case study: Court oversight in action: X in Brazil.

Authorities in Brazil, the country with the world's fifth-largest number of internet users, have banned the social media platform X (formerly known as Twitter). The ban followed a lengthy legal battle between X's owner, Elon Musk, and Brazil's Supreme Court.[85]

Brazil's decision to ban X was not an isolated move. It was part of a broader effort to address the misuse of social media platforms, especially in relation to far-right users and their spread of disinformation which was undermining Brazil's democracy.[86] Between 2020 and 2023, the Brazilian Supreme Court initiated three key criminal investigations focusing on social media activity. These inquiries targeted fraudulent news, organised digital groups that manipulated discourse (known as the milícias digitais- digital militias), and individuals involved in the 2023 attack on Brazil's Congress after former president Jair Bolsonaro's electoral defeat.[87]

In April 2024, Justice Alexandre de Moraes ordered Musk to shut down several far-right accounts spreading disinformation about the 2022 election.[88] While X had previously complied with similar orders, Musk refused while also removing X's legal representative in Brazil. This led Justice de Moraes to set a deadline for Musk to appoint a new legal representative, a legal requirement for foreign companies operating in Brazil.[89] When Musk failed to meet this deadline, X was banned in Brazil, and the financial accounts of Musk's Starlink service were also frozen.

As of 10 September 2024, the ban remains in place until Musk complies with all court orders, including appointing legal representation and paying fines totalling A$4.85 million. In reviewing Justice De Moare's decision, a five-member panel of Brazil's Supreme Court unanimously upheld the ban, ensuring X's suspension will continue indefinitely.[90]

Image: X logo with a Brazilian flag in the background.

Conclusion

Digital platforms have become central to modern life, amplifying both positive and negative impacts on society. However, without proper regulation, these platforms enable harmful content that can result in real-world consequences, particularly for those of us most at risk.

Grounding regulation in human rights law is essential to ensuring that the rights of all users are protected while also fostering a safe, accountable, and transparent online environment. By adopting a risk-based, proactive regulatory framework, as seen in the EU's Digital Services Act and similar international models, Australia can strike the necessary balance between innovation, user protection, and fundamental rights.

Self-regulation has consistently proven inadequate in addressing the serious harms posed by misinformation and disinformation. Digital platforms, driven by profit incentives, often lack the transparency and accountability required to meaningfully tackle these issues. Time and again, we've seen self-regulation fail, with platforms either acting too late or not at all to prevent the spread of harmful content. Allowing platforms to govern themselves will only prolong the very problems we aim to solve.

Australia cannot afford to take this ineffective path. If we are to invest political capital and resources into regulating digital platforms, it is imperative that we do it right the first time. Human rights must form the foundation of any regulatory framework, ensuring that our approach protects the rights and safety of all users. A robust, rights-based regulatory model will not only be more effective but will also ensure that the needs of the most at risk from online harm are safeguarded.

Australia must adopt strong, enforceable regulations that hold platforms truly accountable from the outset. This approach is crucial to addressing the harmful effects of digital platforms while ensuring that our regulatory efforts are equitable, transparent, and sustainable.

Image: David Mejia-Canales, Senior Lawyer at the Human Rights Law Centre.

Endnotes

1 Armani Syed, 'How online misinformation stoked anti-migrant riots in Britain', *Time Magazine* (Online, 7 August 2024) <https://time.com/7007925/misinformation-violence-riots-britain/>

2 The Treasury, *'Independent Review of the Food and Grocery Code of Conduct'*, (Final Report, June 2024) 10.

3 Ibid 19.

4 Gabriel R Sanchez, Keesha Middlemass, 'Misinformation is eroding the public's confidence in democracy', *The Brookings Institute* (Online, 26 July 2022) <https://www.brookings.edu/articles/misinformation-is-eroding-the-publics-confidence-in- democracy/>.

5 Social Media and Online Safety, *'Social Media and Online Safety'*, Report of the House of Representatives Select Committee on Social Media and Online Safety, (March 2022), 11 - 12.

6 Ibid 29-44.

7 World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms*, 2023, 12.

8 Schedule 1, Clause 2, *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024* (Cth), as introduced and read a first time, 12 September 2024.

9 World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms,* 2023.

10 Ibid 5.

11 Ibid.

12 Ibid.

13 *Digital Services Act*, recitals 80-83.

14 The United Nations Sustainable Development Group, *The Human Rights Based Approach to Development Cooperation Towards a Common Understanding Among UN Agencies*, (Online, September 2023) <https://unsdg.un.org/sites/default/files/6959-The_Human_Rights_Based_Approach_to_Development_Cooperation_Towards_a_Common_Understanding_among_UN.pdf>.

15 UN Human Rights Council Resolutions (2012-2018), *The promotion, protection and enjoyment of human rights on the Internet,* UN Doc A/HRC/RES/38/7 (5 July 2018), A/HRC/RES/32/13 (1 July 2016), A/HRC/RES/26/13 (26 June 2014), A/HRC/RES/20/8 (5 July 2012).

16 United Nations Office of the United Nations High Commissioner for Human Rights, *'Guiding Principles on Business and Human Rights'* (2011) <www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf>

17 Kate Jones, *'Online Disinformation and Political Discourse Applying a Human Rights Framework'*, (Report, November 2019) <https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf> 30.

18 International Covenant on Civil and Political Rights, art 19(2).

19 ICCPR Article 19. See also: UN Human Rights Committee, *General Comment No. 34* (2011) 22.

20 ICCPR Article 20.

21 UN Human Rights Committee, *General Comment No. 34: Article 19 (Freedoms of Opinion and Expression),* Human Rights Committee 102nd session, UN Doc CCPR/C/GC/34 (12 September 2011) 2.

22 Ibid 4.

23 Ibid 5.

24 Ibid 7.

25 Ibid.

26 Kate Jones, *'Online Disinformation and Political Discourse Applying a Human Rights Framework'*, (Report, November 2019) <https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf> 35.

27 Kate Jones, *'Online Disinformation and Political Discourse Applying a Human Rights Framework'*, (Report, November 2019) <https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf> 37-38.

28 Ibid 38.

29 UN Human Rights Committee, *General Comment No. 25* (1996) 19.

30 Ibid 25.

31 Kate Jones, *'Online Disinformation and Political Discourse Applying a Human Rights Framework'*, (Report, November 2019) <https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf> 49.

32 Ibid.

33 eSafety Commissioner, *Inappropriate Content: Factsheet,* Australian Government (Website, April 2020) <https://www.esafety.gov.au/sites/default/files/2020-04/Inappropriate%20content%20Factsheet.pdf>.
Research suggests that a high proportion of young people aged 12-17 in Australia have encountered inappropriate or hateful content online: 57% have seen real violence that was disturbing, 33% have seen images or videos promotion terrorism, and nearly half of children aged 9-16 experience regular exposure to sexual images. See also Antonia Quadara, Allisar El-Murr and Joe Latham, *'Online pornography: Effects on children & young people'*, Australian Institute of Family Studies (Australian Government) < https://aifs.gov.au/sites/default/files/publication-documents/online_pornography-effects_on_children_young_people_snapshot_0.pdf>.

34 Shane Compton, Alison Eglentals, Aine Donohoe, *2023 Online Safety Issues Survey- Summary Report*, (Report, June 2023) <https://www.infrastructure.gov.au/sites/default/files/documents/2023-online-safety-issues-survey-summary-report-june2023.pdf>.

35 eSafety Commissioner, *'eSafety initiates civil penalty proceedings against X Corp'* (21 December 2023) <https://www.esafety.gov.au/newsroom/media-releases/esafety-initiates-civil-penalty-proceedings-against-x-corp>.

36 Amnesty International, *The Social Atrocity- Meta and the right to Remedy for the Rohingya* (Report, 29 September 2022) <https://www.amnesty.org/en/documents/ASA16/5933/2022/en/> 6-10.

37 Ibid.

38 Ibid.

39 Ibid.

40 Morgan Saletta, Richard Stearne, *'Understanding Mass Influence'* (Report, 11 February 2021) <https://www.unsw.edu.au/content/dam/pdfs/unsw-adobe-websites/canberra/research/defence-research-institute/2023-02-Understanding-Mass-Influence---A-case-study-of-the-Internet-Research-Agency.pdf> 5.

41 Ibid

42 Ibid

43 Ibid.

44 Ibid

45 Kate Jones, *'Online Disinformation and Political Discourse Applying a Human Rights Framework'*, (Report, November 2019) <https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf> 14.

46 Ibid 15.

47 Ibid.

48 Ibid.

49 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (*Digital Services Act*), art 34.

50 Ibid art 35.

51 Ibid art 37.

52 Ibid art 17.

53 Ibid art 39.

54 Ibid art 40.

55 European Commission, *'The Enforcement Framework Under the Digital Services Act'* (Website, 30 April 2024) <https://digital-strategy.ec.europa.eu/en/policies/dsa-enforcement>.

56 Department for Science, Innovation & Technology (UK), *'Guidance- Online Safety Act: Explainer'* (Website, 8 May 2024) <https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer>.

57 Ibid.

58 The sending of an unsolicited sexual image to someone else via social media, dating apps, or data sharing services such as Bluetooth and Airdrop.

59 Department for Science, Innovation & Technology (UK), *'Guidance- Online Safety Act: Explainer'* (Website, 8 May 2024) <https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer>.

60 Kirsten Pickles, *'Covid-19 misinformation trends in Australia: prospective longitudinal national survey'* (2021) 23(1) Journal of Medical Internet Research <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7800906/>.

61 Monica Wang, Olivia Britton, Jennifer Beard, *'The call for science communication and public scholarship'* (2023) 13(3) Translational Behavioural Medicine, 156-159 <https://doi.org/10.1093/tbm/ibac096>.

62 Ibid 156-159.

63 Reset.Tech Australia, *A duty of care in Australia's Online Safety Act* (Policy Briefing), April 2024, 2.

64 Digital Services Act, art 17(1).

65 Ibid art 24(5)

66 Ibid art 17(3-4).

67 Ibid art 16, art 22.

68 European Commission, *DSA Transparency Database* (Website, accessed 10 September 2024) <https://transparency.dsa.ec.europa.eu/>.

69 Ibid.

70 Graham Readfearn, *'Australian Conversation Foundation's X account suspended after apparent 'report bombing'*, The Guardian, (Online, 5 August 2024) <https://www.theguardian.com/australia-news/article/2024/aug/05/australian-conservation-foundation-acf-x-account-suspended-report-bombing>.

71 European Commission, *'Shaping Europe's Digital Future'* (Website, 25 July 2024) <https://digital-strategy.ec.europa.eu/en/factpages/safer-fairer-online-environment>.

72 *Digital Services Act*, art 26.

73 Ibid

74 Ibid art 26.

75 Ibid arts 28, 34-35.

76 Ibid arts 25, Art 31.

77 Ibid art 71.

78 *Digital Services Act*, Art 81.

79 Ibid.

80 European Commission, *'The Enforcement Framework Under the Digital Services Act'* (Website, 30 April 2024) <https://digital-strategy.ec.europa.eu/en/policies/dsa-enforcement>.

81 Ibid.

82 *Digital Services Act*, art 69.

83 European Commission, *'The Enforcement Framework Under the Digital Services Act'* (Website, 30 April 2024) <https://digital-strategy.ec.europa.eu/en/policies/dsa-enforcement>.

84 *Digital Services Act*, recital 39.

85 Tariq Choucair, *'Brazil just banned X. Could other countries follow suit'*, The Conversation (Website, 2 September 2024) <https://theconversation.com/brazil-just-banned-x-could-other-countries-follow-suit-237960>.

86 Ibid.

87 Ibid.

88 Ibid.

89 Ibid.

90 João da Silva, Vanessa Buschschlüter, *'Top Brazil court upholds ban of Musk's X'*, BBC (Online, 3 September 2024) <https://www.bbc.com/news/articles/crkmpe53l6jo>.

Human Rights Law Centre.