WORLD ECONOMIC FORUM

Global Principles on Digital Safety: Translating International Human Rights for the Digital Context

WHITE PAPER

JANUARY 2023

Images: Getty Images

Contents

Foreword

The growth and scale of digital connectivity brings benefits to the global community, but the internet needs to be a safe and secure space for all.

Courtney Gregoire
Chief Digital Safety Officer, Microsoft

Iain Drennan
Executive Director, WeProtect Global Alliance

Cathy Li
Head, Shaping the Future of Media, Entertainment and Sport, World Economic Forum

Minos Bantourakis
Project Lead, Global Coalition for Digital Safety, World Economic Forum

More than 5 billion people use the internet worldwide. Safety challenges related to harmful content and conduct caused by bad actors, including issues such as child sexual exploitation and abuse, violent extremism and terrorism, hate speech, self-harm and suicide, and mis- and dis-information, can be amplified in the digital world. Urgent action is needed to minimize the potential harm to all people, with an emphasis on society's most vulnerable groups, including children.

The World Economic Forum's Global Coalition for Digital Safety is bringing together a diverse group of leaders to accelerate public–private cooperation to tackle harmful content and conduct online. Members of the coalition have worked together to develop the Global Principles on Digital Safety, intended to answer the fundamental question: "How should human rights translate in the digital world?". The principles aim to advance digital safety in a rights-respecting way, drive multistakeholder alignment and enable positive behaviours and actions across the digital ecosystem.

The principles are the result of intensive discussions, expert interviews and consultations among a diverse group of global experts, including policy-makers, major social media and tech platforms, safety tech companies, civil society organizations and academics. They build upon existing international human rights principles and frameworks and apply them to digital safety (see Appendix: Digital safety resources for an overview).

These principles are intended to serve as a guide for all stakeholders in the digital ecosystem to advance digital safety by informing and enabling regulatory, industry and societal efforts and innovations. They recognize the key roles played by governments, online service providers and civil society and provide a framework for applying rights-respecting approaches to online safety across an activity, from regulation to product development. Critically, the principles also encourage deeper collaboration and cooperation, recognizing that we all have responsibilities to help build a safe, rewarding and innovative digital world.

Introduction

Internet connectivity and online services play a critical role in enabling and empowering individuals to enjoy their human rights.

Digital services are at the heart of economic, educational, social and political affairs across the globe. Indeed, internet connectivity and online services play a critical role in enabling and empowering individuals to enjoy their human rights. Ensuring that everyone can engage safely online is essential to promoting healthy societies, supporting the realization of fundamental rights and engendering trust in an open, global internet.

The international community has enshrined fundamental rights and freedoms in the Universal Declaration of Human Rights and other foundational covenants, such as the International Covenant on Civil and Political Rights. Efforts such as the United Nations Guiding Principles on Business and Human Rights have also established a shared understanding that private actors are responsible for respecting human rights and, beyond that, can play an active role in advancing human rights. However, the process of collectively building norms for *how* governments and service providers can apply international human rights frameworks to assess and address the risks related to the digital ecosystem is still ongoing. The Global Principles on Digital Safety are intended to establish the framework for that work, and are designed to support governments and online service providers in advancing digital safety through a multistakeholder approach.

Digital safety is both a desired outcome and an evolving discipline. Fundamentally, digital safety is about preventing and reducing harm, including through moderating illegal or harmful content or conduct, driving responsible platform design and governance, or designing tools to empower users to tailor their online experiences. Harm can be highly local or context-specific: unique risks may arise in different countries or regions or for different communities. It is important to acknowledge that digital safety requires a complex range of deliberations, balancing legal, policy, ethical, social and technological considerations. Digital safety decisions must be rooted in international human rights frameworks.

Existing human rights laws and principles recognize that fundamental human rights are indivisible; restrictions on these rights are acceptable only when certain conditions are met. Any action to create a safer digital ecosystem should take a rights-respecting approach and ensure decisions are grounded in the principles of necessity, proportionality and legality.

These Global Principles on Digital Safety build on a range of efforts to crystallize rights-based and multistakeholder approaches to digital safety. They have been developed in the context of significant regulatory developments and informed by a broad range of stakeholder consultations.[1] Digital safety challenges are exacerbated by the fluidity and volume of activity across a broad, interconnected digital ecosystem. Harmful behaviour is seldom confined to a single platform, and fully understanding the spectrum of harms and effectively mitigating them requires diverse viewpoints in decision-making.

The principles are content- and technology-neutral. They are intended to be sufficiently flexible to adapt to the rapidly evolving nature of technology and the threat landscape and to ensure that actions to address harms are proportionate to the context in which they occur. The principles apply to decision-making at all times: from product or policy inception, through business-as-usual activity and in times of crisis.

1. While a range of more topic-specific approaches have been developed, no single set of principles provides guidance across the full suite of decisions that may be required to advance digital safety. See Appendix: Digital safety resources for a more complete list, but extant principles include: the Voluntary Principles to Counter Online Child Sexual Abuse and Exploitation; the Christchurch Call: To Eliminate Terrorist & Violent Extremist Content Online; the Santa Clara Principles; the Australian eSafety Commissioner Safety by Design Principles; the Digital Trust and Safety Partnership's Best Practices; and more.

1 The role of civil society

Civil society is critical to advancing digital safety: genuinely multistakeholder approaches are not possible without active participation and perspectives from civil society groups to understand the impacts of digital ecosystems on the communities they serve. Moreover, many civil society or non-governmental organizations provide related services, ranging from victim and survivor support to advocacy for free expression, to hotlines, to the provision of safety education or human rights impact assessment. Civil society groups also help bridge the gulf between public and private institutions and individuals, including by amplifying the voices of the most vulnerable, such as survivors of child sexual abuse or children. Civil society can therefore play a vital role in ensuring the commitments for governments and the private sector are fulfilled and by holding these sectors to account.

2 Joint principles

Supporters of the principles should:

- Collaborate with diverse stakeholders to build a safe, trusted and inclusive online environment, enabling every person to enjoy their rights in the digital environment.
- Seek insights and diverse perspectives from civil society to inform policy-making, understand emerging harms and support inclusive and informed decision-making on digital safety.
- Support innovative and evidence-based multistakeholder solutions to assess, address and advance digital safety and prevent harm.
- Advance transparency about approaches to, and the outcomes of, efforts to advance digital safety to improve the collective response.
- Recognize the particular importance of helping vulnerable and marginalized groups to realize their rights in the digital world, including the importance of defending children's safety and privacy online.

3 Additional principles for governments

Supportive governments should consider the following principles:

- Seeking to prioritize human rights-based, evidence-based and data-driven approaches to digital safety policy-making, including by:
  - Undertaking multistakeholder consultations to identify problems and possible solutions at the outset of a policy process
  - Ensuring that credible, evidence-based and equitable research informs policy
  - Carrying out impact assessments to understand their citizens' needs and the potential consequences of legislation, including for vulnerable and marginalized groups and those who may be disproportionately harmed online
  - Accurately defining and reflecting the roles and responsibilities that entities across all layers of the internet infrastructure have in relation to digital safety
  - Maintaining the space for innovation and experimentation to achieve safety outcomes
- Distinguishing between illegal content and content that is lawful but may be harmful, and differentiating any regulatory measures that apply to each category accordingly, including by:
  - Appropriately delineating between the roles and responsibilities of governments and government entities versus online service providers
  - Providing industry with the flexibility to develop systems and processes appropriate to their services, while remaining within the bounds of the law and international human rights frameworks
- Embracing human rights-based approaches to digital safety by:
  - Encouraging online service providers to undertake human rights due diligence as well as to understand their risk profile with respect to clearly defined harms

  - Empowering providers to take appropriate and proportional mitigation measures

- Ensuring law and policy respect and protect the rights of all users, including by driving domestic regulatory coherence and incentivizing collaboration among domestic regulators (including among privacy, safety and security authorities).

- Seeking to evaluate the impact of new policy or regulatory measures, including the effectiveness of measures to reduce harm, understanding any unintended consequences and embracing continuous improvement and innovation. Such evaluations should be transparent and predictable, and incorporate a wide range of multistakeholder feedback.

- Supporting a free, open and interoperable internet, including by engaging internationally to:

  - Protect and promote the global free flow of information, ensuring that the economic and social benefits of the internet and related digital technologies continue to flourish and support the UN Sustainable Development Goals

  - Engage in multistakeholder and multilateral contexts to support the application of human rights online and to advance shared understandings of online harm issues

  - Encourage and recognize multistakeholder collaboration to build shared frameworks, systems and protocols that are harmonized across borders while respecting national sovereignty

- Supporting appropriate and effective legal processes that enable the investigation of, and a justice system response to, illegal content or conduct online.

- Supporting victims and survivors of abuse or harm, including facilitating access to justice and resources tailored to the needs of vulnerable groups (e.g. children, women, LGBTQI+, journalists or Indigenous communities) and ensure their perspectives and needs inform policy-making.

- Seeking to invest in harm prevention and education, including taking action to foster inclusive societies, increase digital safety and literacy, enhance media literacy skills, and educate communities on digital citizenship.

4 Additional principles for online service providers

Supportive online service providers should consider the following principles:

- Committing to respecting human rights responsibilities, as set out in the UN Guiding Principles on Business and Human Rights through a clear statement or policy and regular due diligence and disclosure.

- Establishing the necessary infrastructure and frameworks to embed digital safety throughout the business.

- Investing in and embedding a multidisciplinary approach to safety by design throughout the business life cycle of products and services, including empowering users by providing tools, methods and resources to tailor their experiences, help them safeguard themselves and report harm.

- Embracing innovative, evidence- and risk-based approaches to digital safety; for instance, through undertaking risk assessments and implementing tailored and targeted policy, technical and operational harm-mitigation measures that incorporate holistic human rights considerations and draw on best practices that recognize the differences between unique services.

- Providing clarity and transparency about a service's approach to digital safety, including by making public, and providing transparency on, a service's community standards or other terms of service, the processes and systems in place to mitigate against abuse and data on the outcomes.

- Establishing a rights-based rationale for actions taken, including:

  - Developing and maintaining content policies in line with international human rights law and tailored to the nature and impact of the harm, alongside other considerations

  - Consistently enforcing those policies with actions tailored to the nature and gravity of the harm as well as the nature of the service

  - Seeking to understand important contextual differences between countries, regions and cultures

  - Providing clear pathways for users to complain or appeal against moderation or enforcement decisions where appropriate and ensuring such requests receive a timely response

– Ensuring that technologies and tools used to advance digital safety uphold human rights, including the rights to equality and non-discrimination, privacy and freedom of expression. For example, this includes tailoring any application of safety technologies to the specific service by considering factors such as: the nature and scale of the risk; any potential biases; the public or private nature of any communications; user expectations of the service; the accuracy of the tools; and the scale of any necessary human intervention.

– Mitigating the potential risk of adverse impacts on staff and other personnel tasked with mitigating abuse.

– Collaborating with other online service providers to share best practices and support a safer online ecosystem.

5 Way forward

Supporters of the principles should endeavour to:

– Make decisions and take actions aligned with the principles.

– Raise awareness of these principles across the online ecosystem, including through active promotion, targeted outreach and the encouragement of multistakeholder adoption.

– Build a multistakeholder community among supporters of the principles to facilitate dialogue between supporters of the principles and existing multistakeholder safety efforts, to share learning and to advance rights-respecting approaches to online harms.

– Share best practices in developing inclusive processes to facilitate multistakeholder input, including on designing processes to seek perspectives from children and victims or survivors of online abuse.

Appendix: Digital safety resources

Included here is a list of existing principles and frameworks (not exhaustive).

1. Australian eSafety Commissioner – Safety by Design Principles
2. Christchurch Call: To Eliminate Terrorist & Violent Extremist Content Online
3. Council of Europe Convention on the Protection of Children Against Sexual Exploitation and Sexual Abuse
4. Convention on the Rights of the Child
5. Glion Human Rights Dialogue 2020 – Human Rights in the Digital Age: Making Digital Technology Work for Human Rights
6. Global Network Initiative (GNI) Principles
7. Guiding Principles on Business and Human Rights (UNGP)
8. International Covenant on Civil and Political Rights (ICCPR)
9. International Covenant on Economic, Social and Cultural Rights (ICESCR)
10. Internet Governance Forum (IGF)
11. Oasis Consortium – User Safety Standards for Our Digital Future
12. OECD – Transparency Reporting on TVEC Online
13. OSCE – Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual
14. Paris Call – Paris Call for Trust and Security in Cyberspace
15. Tech Coalition – TRUST: Voluntary Framework for Industry Transparency
16. The Santa Clara Principles
17. The XRSI Privacy Framework
18. Trust and Safety Best Practices – Digital Trust and Safety Partnership
19. Trusted Cloud Principles
20. UN Committee on the Rights of the Child General Comment #25 on Children's Rights in Relation to the Digital Environment
21. Universal Declaration of Human Rights (UDHR)
22. Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse

Contributors

Lead Authors

Maria Cristina Capelo
Head of Safety Policy, Meta Platforms

Roxanne Carter
Senior Manager, Government Affairs and Public Policy, Google

Jeffrey Collins
Director, AWS Trust and Safety, Amazon

Iain Drennan
Executive Director, WeProtect Global Alliance

Courtney Gregoire
Chief Digital Safety Officer, Microsoft

Lea Kaspar
Executive Director, Global Partners Digital

Collin Kurre
Technology Policy Principal, Office of Communications , (Ofcom)

Liz Thomas
Director of Public Policy, Digital Safety, Microsoft

World Economic Forum

Minos Bantourakis
Project Lead, Global Coalition for Digital Safety

Cathy Li
Head, Shaping the Future of Media, Entertainment and Sport; Member of the Executive Committee

Acknowledgements

Josianne Galea Baron
Children's Rights and Business Specialist
United Nations Children's Fund (UNICEF)

Gustavo Silveira Borges
Professor of Human Rights and Social Media,
University of Extreme South of Santa Catarina
(UNESC)

Brent Carey
Chief Executive Officer, Netsafe

Daniel Child
Industry Affairs and Engagement Manager,
Office of the eSafety Commissioner

Francisco Brito Cruz
Executive Director, InternetLab

Louis-Victor de Franssu
Chief Executive Officer, Tremau

Miriam Estrin
Senior Policy Manager, Google

Theodoros Evgeniou
Professor of Decision Sciences and Technology
Management, INSEAD

Sonia Facchini
Director of Relationships and Policy, Internet
Commission

Steven Feldstein
Senior Fellow, Carnegie Endowment for
International Peace

Inbal Goldberger
Vice-President Trust and Safety, ActiveFence

Susie Hargreaves
Chief Executive Officer, Internet Watch Foundation

Sasha Havlicek
Chief Executive Officer, Institute for Strategic
Dialogue

Peggy Hicks
Director, Thematic Engagement, Special
Procedures, and Right to Development Division
Office of the High Commissioner for Human Rights
(OHCHR)

Afrooz Kaviani Johnson
Child Protection Specialist, United Nations
Children's Fund (UNICEF)

Adnan Laeeq
Global Head of Digital and Innovation, Plan
International

Farah Lalani
Project Lead, Global Coalition for Digital Safety
(2019–2022), World Economic Forum

Kat Lo
Content Moderation Lead and Research, Meedan

Namrata Maheshwari
Asia Pacific Policy Counsel, Access Now

Victoria Nash
Director, Oxford Internet Institute

Susan Ness
Distinguished Fellow, Annenberg Public Policy
Center of the University of Pennsylvania

Elina Noor
Director, Political-Security Affairs, Asia Society Policy Institute

Ioanna Noula
Senior Leader – Operations, Internet Commission

Nina Jane Patel
Co-Founder, Vice-President of Metaverse Research Kabuni

Jason Pielemeier
Executive Director, Global Network Initiative

Andrew Puddephatt
Chair, Internet Watch Foundation

Ram Puvinathan
Strategy Lead, PUBLIC

Courtney Radsch
Fellow, UCLA Institute for Technology, Law & Policy

Var Shankar
Director of Policy, Responsible AI Institute

Brett Solomon
Executive Director, Access Now

Nikki Soo
Subject Expert, Harmful Content Public Policy, Europe, TikTok Information Technologies

Ian Stevenson
Chief Executive Officer, Cyacomb

John Tanagho
Executive Director, IJM's Center to End Online Sexual Exploitation of Children

Deepak Tewari
Chief Executive Officer, Privately

Bertie Vidgen
Chief Executive Officer and Co-Founder, Rewire Online

Alicia Wanless
Director, Partnership for Countering Influence Operations, Carnegie Endowment for International Peace

Our thanks to Ruby, Liberty and Joy (pseudonyms), International Justice Mission IJM, Philippines Survivor Network

Editing and design

Laurence Denmark
Designer, Studio Miko

Alison Moore
Editor, Astra Content

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org