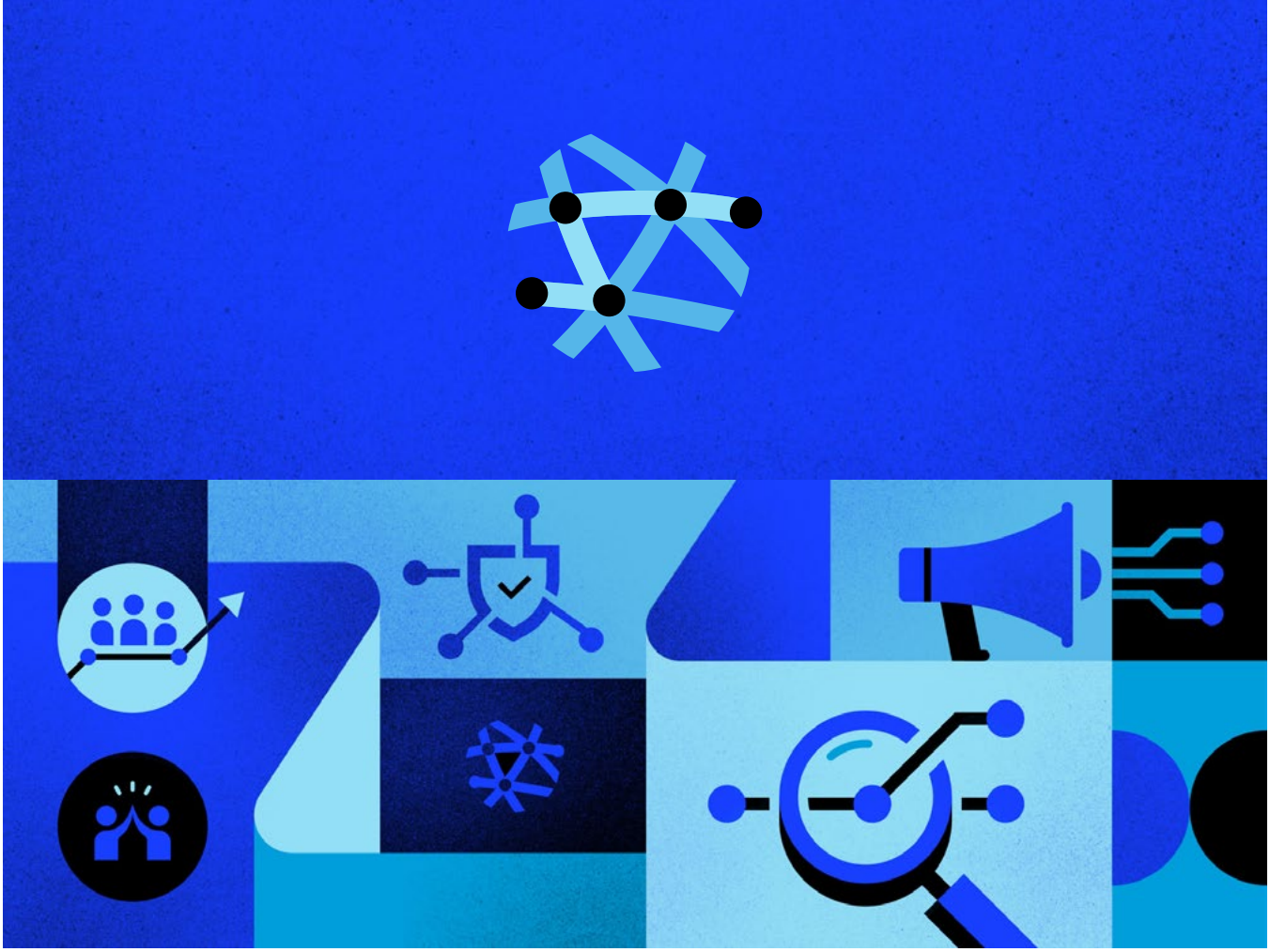


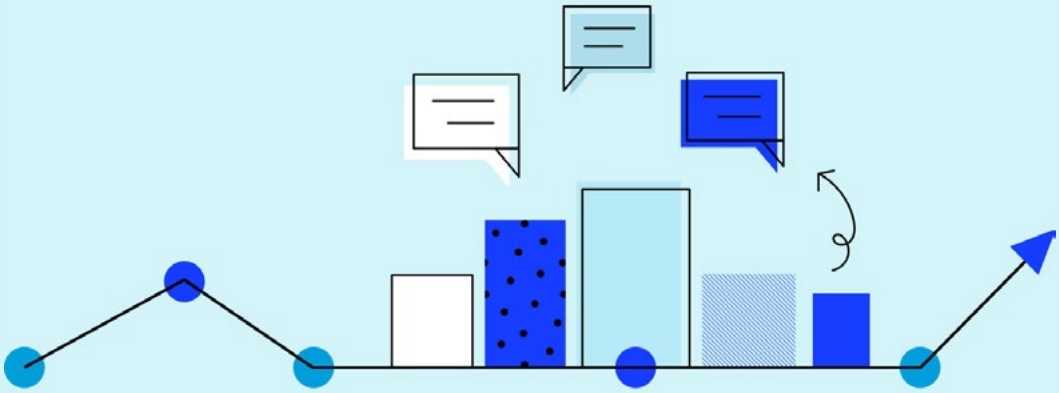
United Nations Global Principles For Information Integrity

Recommendations for Multi-stakeholder Action

United Nations

Table of Contents

The Information Ecosystem in the Digital Era



Technological advances have in a few short decades revolutionized communications, connecting individuals and communities on a previously unthinkable scale and presenting unparalleled opportunities for the diffusion of knowledge, cultural enrichment and sustainable development. They have in many ways raised ambitions for the integrity of the information ecosystem—where freedom of expression is fully enjoyed and where accurate, reliable information, free from discrimination and hate, is available to all in an open, inclusive, safe and secure information environment.

While these advances have enabled the mass dissemination of information, they have also facilitated the spread of misinformation, disinformation and hate speech by many kinds of actors at historically unprecedented volume, velocity and virality, risking the integrity of the information ecosystem. Such risks encompass a range of current, emergent and future threats amid rapid breakthroughs in artificial intelligence technologies.

This erosion of the integrity of information spaces can undermine people's ability to exercise human rights and can hamper efforts to achieve peace, prosperity and a livable future on our planet. In this way, the task of strengthening information integrity presents one of the most urgent challenges of our time.

Information integrity entails a pluralistic information space that champions human rights, peaceful societies and a sustainable future. It holds within it the promise of a digital age that fosters trust, knowledge and individual choice for all.

Promoting information integrity involves empowering people to exercise their right to seek, receive and impart information and ideas of all kinds and to hold opinions without interference. In an increasingly complex digital information environment, this means enabling individuals to navigate information spaces safely with privacy and freedom.

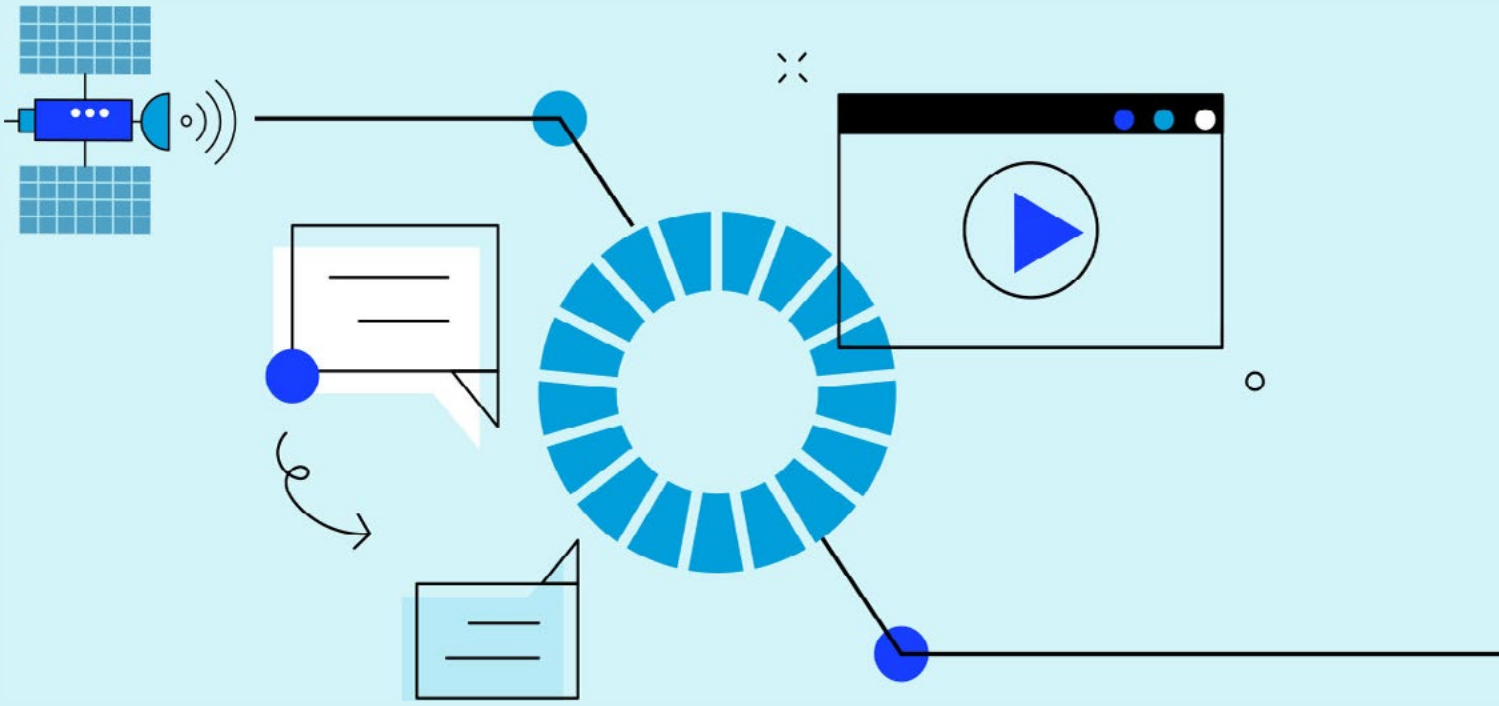
Information Integrity and the Sustainable Development Goals

Efforts to strengthen information integrity are crucial to preserve and further advance the Sustainable Development Goals. Erosion of the integrity of the information ecosystem can worsen existing vulnerabilities in achieving the Goals, in particular for countries in the global South.

Groups in situations of vulnerability and marginalization are disproportionately affected. The participation of greater numbers of women in the global workforce, for instance, is vital to achieve the Goals. However, in addition to discriminatory laws and policies that exist in many countries, gender-based hate speech, disinformation and violence are used to systematically subjugate women by silencing and pushing them out of the public sphere. This can have devastating and long-lasting consequences for women's participation, suppressing women's voices and fuelling self-censorship, causing professional and reputational damage and jeopardizing hard-won progress in gender equality.

Efforts to exploit the information space to undermine climate action further highlight the urgency of the challenge. Coordinated disinformation campaigns, often driven by commercial interests, seek to deny or sow doubt about the scientifically agreed basis for human-induced climate change, its causes or impacts, in order to delay or derail action to meet climate goals. Public figures—activists, scientists and broadcasters—have become targets of hate speech, threats and harassment for their efforts to provide information about and address the climate crisis.

Across the spectrum of the Goals, from good health and zero hunger to peace, justice, education and reducing inequalities, measures to strengthen information integrity will boost efforts to achieve a sustainable future and leave no one behind.



Towards United Nations Global Principles for Information Integrity

The United Nations conducted wide-ranging and diverse consultations on information integrity across all regions with its Member States, civil society, including youth-led organizations, media, academia and private sector representatives. Stakeholders spoke through country-level discussions, virtual sessions, bilateral meetings and via a globally disseminated public online form.

These consultations highlighted a demand for unifying recommendations that are applicable across all geographies and contexts and that address the requirements of all individuals, in particular attending to the needs of groups in situations of vulnerability and marginalization.

In response, the United Nations Global Principles for Information Integrity offer a holistic framework to guide multi-stakeholder action for a healthier information ecosystem. This framework consists of five principles for strengthening information integrity, each of which include recommendations for key stakeholder groups.

The principles are: societal trust and resilience; independent, free and pluralistic media; transparency and research; public empowerment; and healthy incentives. They all share at their core an unwavering commitment to human rights.

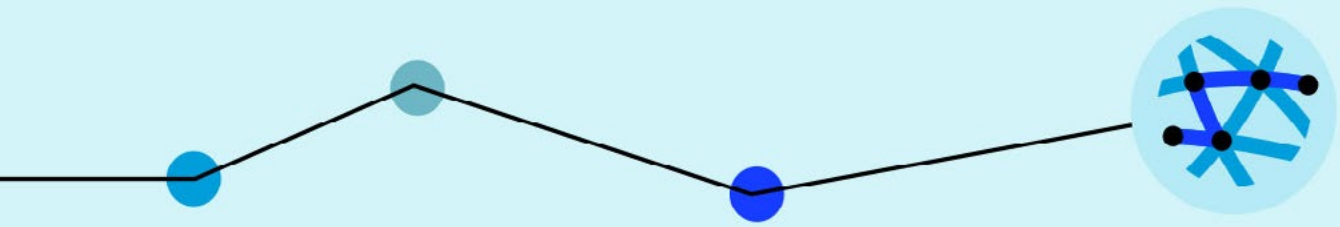
The Global Principles acknowledge and build on the extensive efforts and progress already made by States, civil society, the private sector and other stakeholders. They provide a unified point of departure for protecting and promoting information integrity in all walks of life, and in all languages and contexts, recognizing the global solidarity and breadth of responses required at an unprecedented scale, speed and intensity.

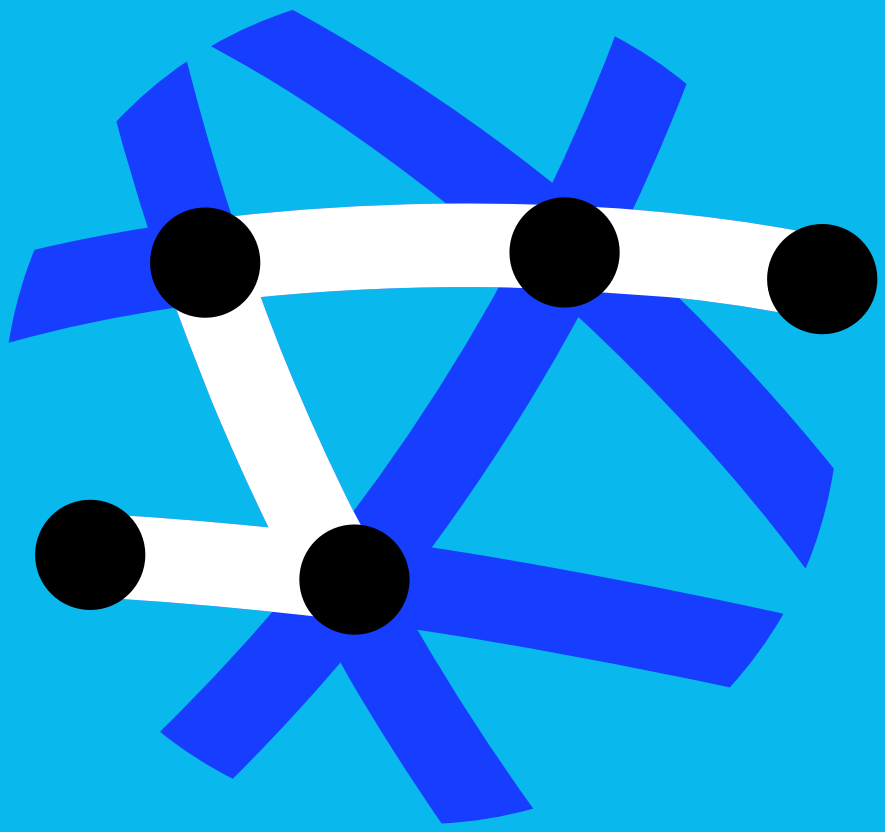
The Global Principles present an occasion for individuals, public and private entities, including the United Nations system, Governments, media, civil society organizations and for-profit corporations across the technology, advertising and public relations sectors, to align with the rights and freedoms enshrined in international law and form broad coalitions for information integrity.

The Global Principles build on the ideas proposed in Our Common Agenda and in the United Nations Secretary-General's policy brief 8: information integrity on digital platforms.

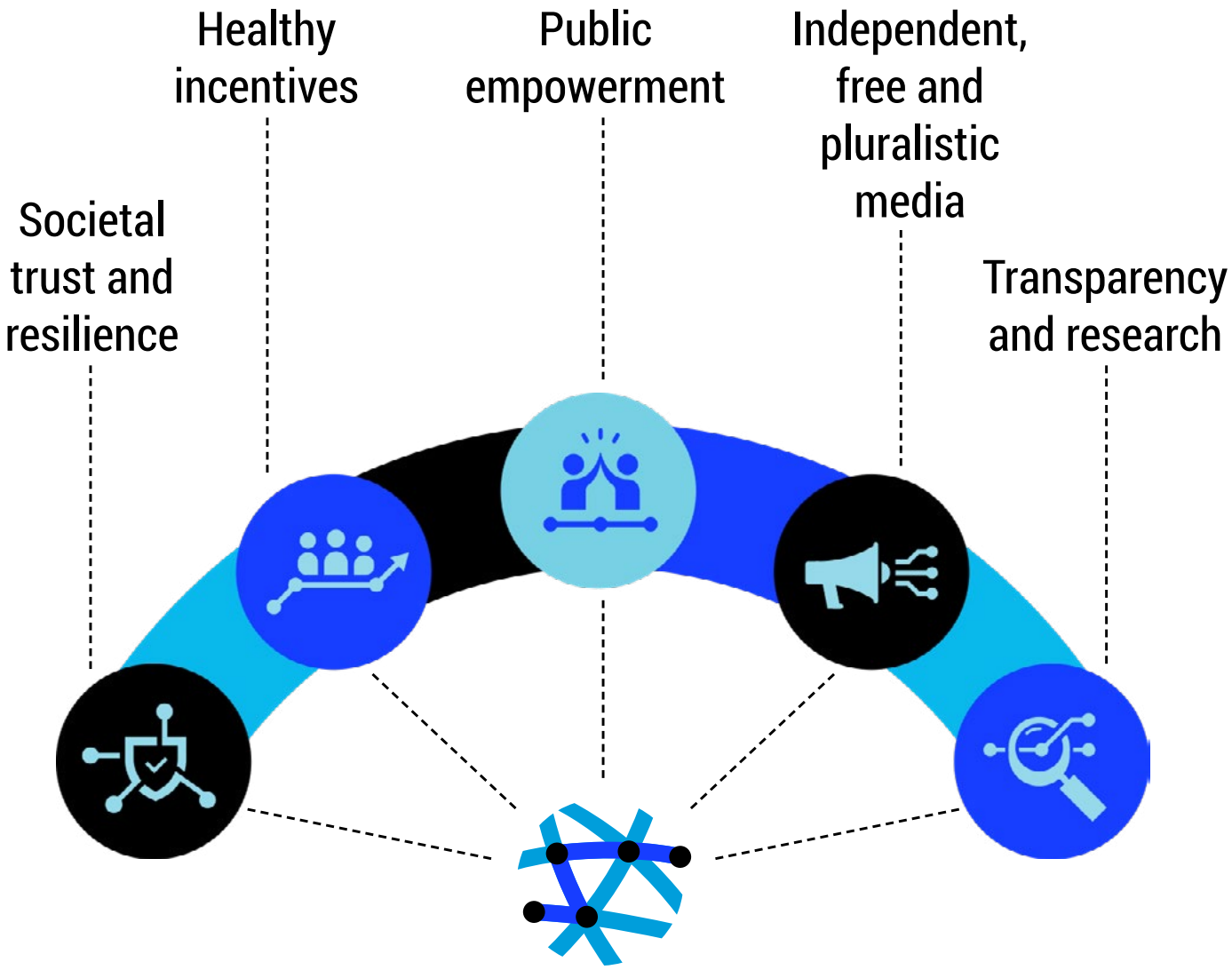
In addition to being grounded in international law, including international human rights law, the Global Principles complement the relevant United Nations Guiding Principles on Business and Human Rights, the UNESCO Guidelines for the Governance of Digital Platforms, the United Nations Plan of Action on the Safety of Journalists and the Issue of Impunity, the UNESCO Recommendation on the Ethics of Artificial Intelligence and the United Nations Strategy and Plan of Action on Hate Speech. The Global Principles offer a resource for United Nations Member States in their considerations towards A Pact for the Future and the Global Digital Compact.

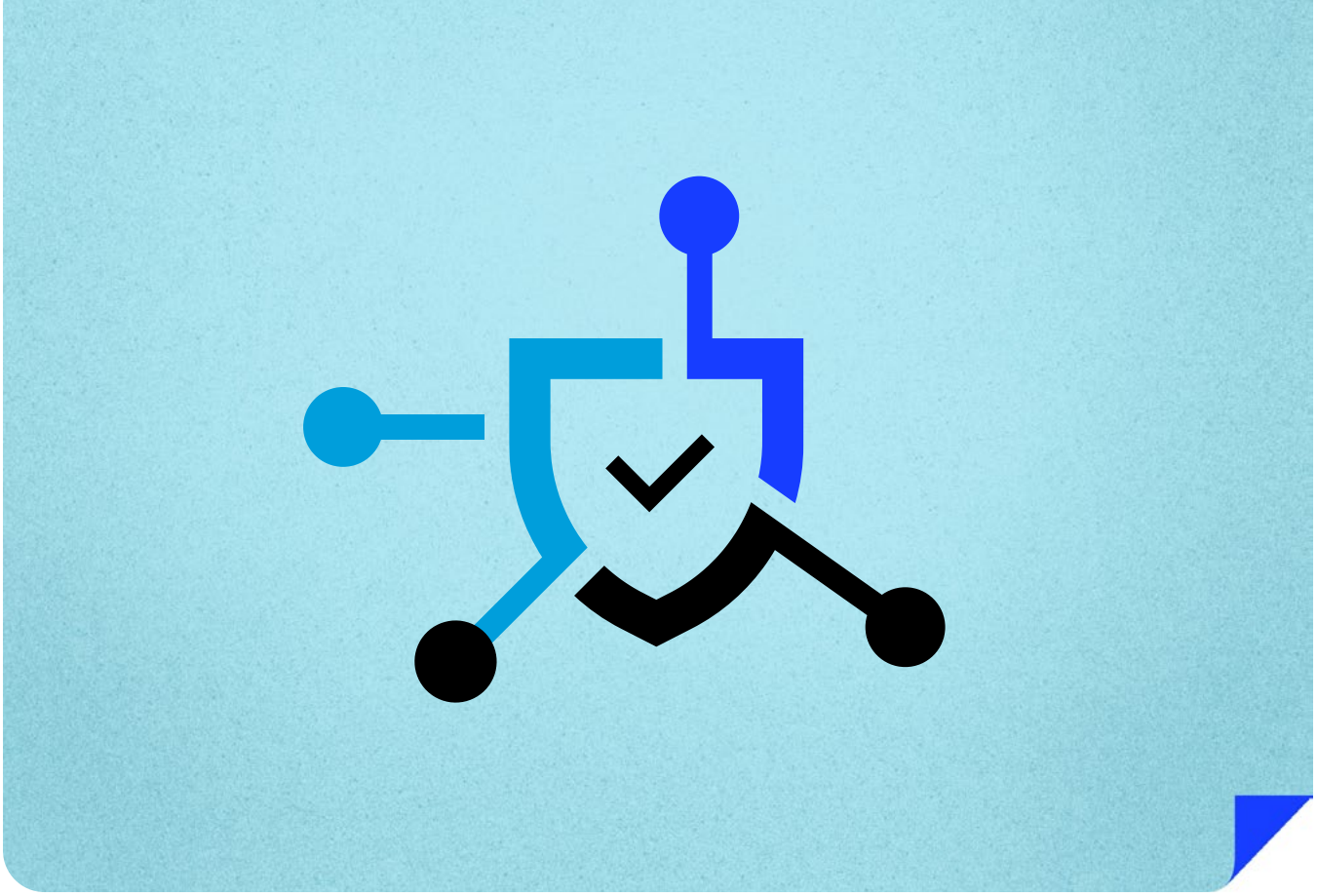
In this way, the Global Principles further reflect the unwavering commitment of the United Nations to strengthening information integrity and are intended to guide the work of the Organization into the future.





5 UN GLOBAL PRINCIPLES FOR INFORMATION INTEGRITY





Societal Trust and Resilience

Trust and resilience throughout societies are key components of information integrity. Trust, in this context, refers to the confidence that people have in the sources and reliability of the information that they access, including official sources and information, and in the mechanisms that allow information to flow throughout the ecosystem. Resilience refers to the ability of societies to handle disruptions or manipulative actions within the information ecosystem.

Trust and resilience are vulnerable to actions driven by State and non-State actors who seek to exploit the information ecosystem for strategic, political or financial gain. These actions, at times widely coordinated, can result in a range of harms and jeopardize people's ability to critically assess science and facts.

Large technology companies hold significant power in the information ecosystem and exercise inordinate influence over the manner in which stakeholders, including other businesses, advertisers, news media and individual users, interact with and access information. Advances in artificial intelligence (AI) technologies, such as generative AI, have introduced the means to create risks to information spaces at scale and with minimal costs. AI-generated or -mediated content, purporting to be real or original, can be highly believable, emotionally resonant and hard to detect and can spread rapidly across algorithm-driven platforms and media outlets. This has the potential to exponentially create, accelerate and deepen trust deficits.

Addressing risks to information integrity demands robust, forward-looking and innovative digital trust and safety practices, enforced consistently across languages and contexts. These practices should reflect the insights of groups in situations of vulnerability and marginalization that are disproportionately exposed to potential harm.

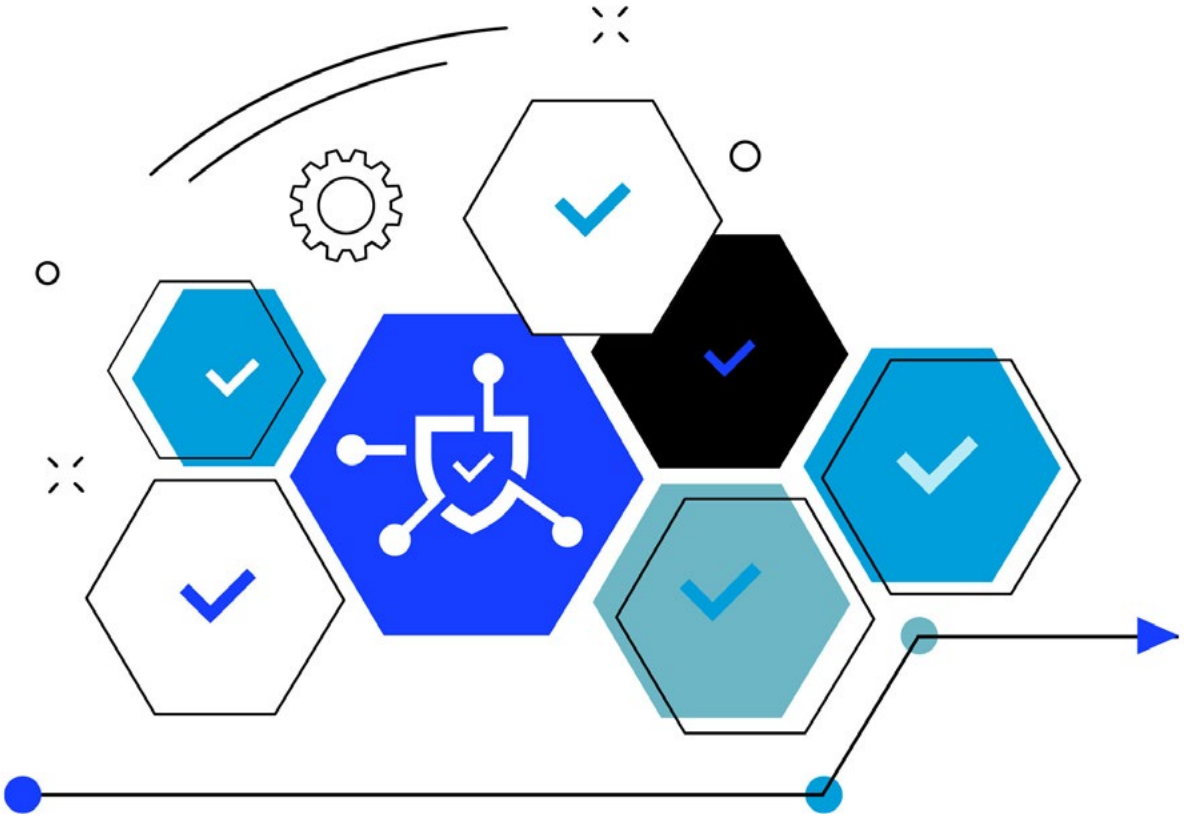
Women, older persons, children, youth, persons with disabilities, Indigenous Peoples, refugees and stateless people, LGBTIQ+ people and ethnic or religious minority groups need to be particularly considered.

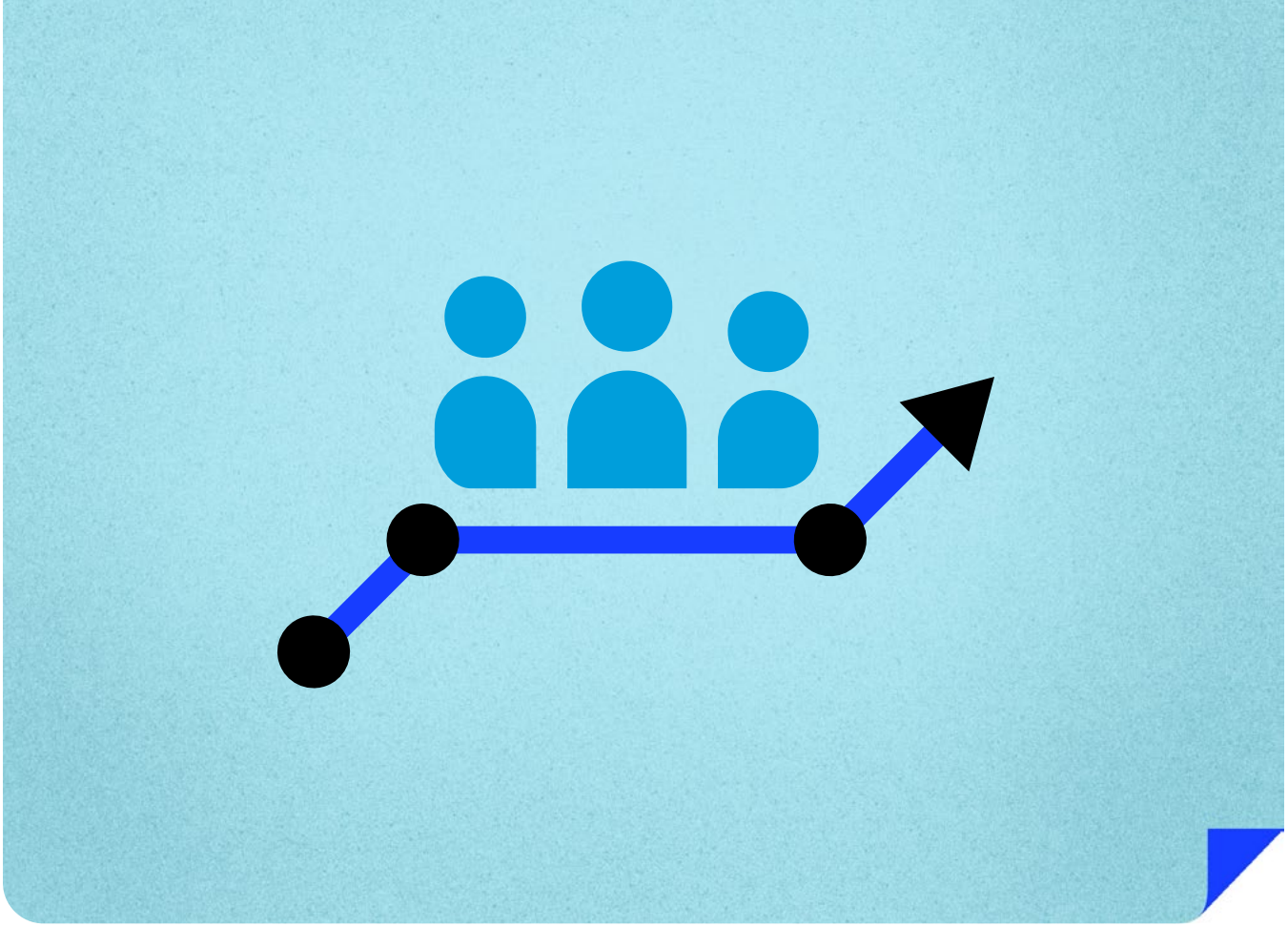
Many young people and children spend a significant portion of their lives online and obtain a vast range of information from digital channels. They already often bear the brunt of risks to information spaces and will be most directly affected by emerging technologies and media trends.

People are generally more resilient and better equipped to pre-empt and navigate such risks when they have access to a diverse range of information sources and feel included, equal, socioeconomically secure and politically empowered. When that is not the case, these risks can often find more fertile ground to proliferate. Responses should therefore acknowledge underlying societal needs to boost long-term resilience.

All stakeholders committed to acting in the public interest can strive to adapt to the realities of a constantly evolving communications landscape by harnessing information spaces for common benefit. This is particularly critical at pivotal societal moments such as elections, natural hazards and human-made crises, when risks to information spaces are pronounced, can deepen social polarization, undermine people's ability to participate in public life, and, in extreme cases, be used to incite violence.

Activists, journalists, humanitarians and United Nations personnel, including peacekeepers, election workers, scientists, medical professionals and others, can become targets, with potentially dire consequences. Online harassment and other insidious tactics can result in the silencing of voices and shrinking of civic spaces. Concerted efforts to safeguard such individuals are paramount.





Healthy Incentives

Creating healthy incentives involves addressing the critical implications for information ecosystem integrity resulting from current business models, which depend on targeted advertising and other forms of content monetization as the dominant means of revenue generation.

These models have provided unprecedented growth opportunities for businesses of all sizes, foremost the technology companies that own and operate digital platforms, and have given rise to a creator economy powered by and benefiting countless people. These models have also enabled financial incentives and opportunities for purveyors of disinformation and hate who exploit the attention economy in which technology companies track user behaviour to collect data, feeding algorithms that prioritize engagement in a bid to maximize potential revenue for advertisers and creators. Messaging designed to polarize and produce strong emotions is often that which generates the most engagement, with the result that algorithms have led to rewarding and amplifying harmful content.

Actors exploiting these business models include information manipulators and mainstream public relations firms contracted by States, political figures and private sector entities to provide orchestrated manipulation campaigns, at times transnationally.

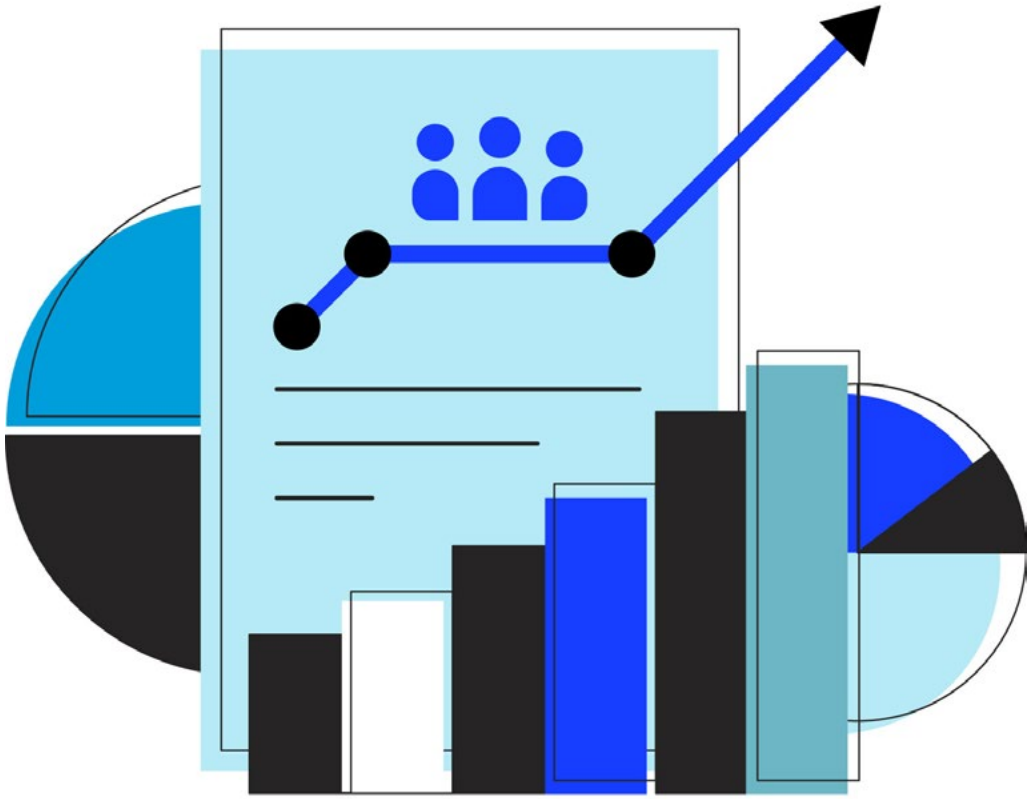
The technology sector has designed digital advertising processes to be complex and opaque with minimal human oversight. This is advantageous to many actors in the advertising technology (ad tech) supply chain, with large technology companies profiting most of all.

Such opaque design can lead to advertising budgets inadvertently funding individuals, entities or ideas that advertisers might not have intended to support, which can constitute a material risk for brands. These advertisement placements can also negatively impact advertising campaign effectiveness and brand safety.

The handful of companies who dominate ad tech are at the same time responsible for implementing advertising standards on the platforms that they own, where enforcement of such standards can be patchy and inconsistent.

Such erosion of information ecosystem integrity highlights the need for a fundamental shift in incentive structures. This can happen through business models guided by human rights and that do not rely on algorithm-driven targeted programmatic advertising that is based on behavioural tracking and personal data.

Advertisers can benefit the information ecosystem in a way that both strengthens information integrity and makes good business sense. While technology companies are unlikely to readily abandon current business models, healthier incentives can be achieved through greater transparency for advertisers into advertising processes and the adherence to human-rights responsible advertising policies by advert deliverers. By gaining more control of a transparent supply chain, advertisers can also see a better return on their investment.



Public Empowerment

Empowerment for individuals navigating the information ecosystem demands that people have control over their online experience, can make informed decisions as to the media that they choose to consume and can express themselves freely. Public empowerment requires consistent access to diverse and reliable sources of information.

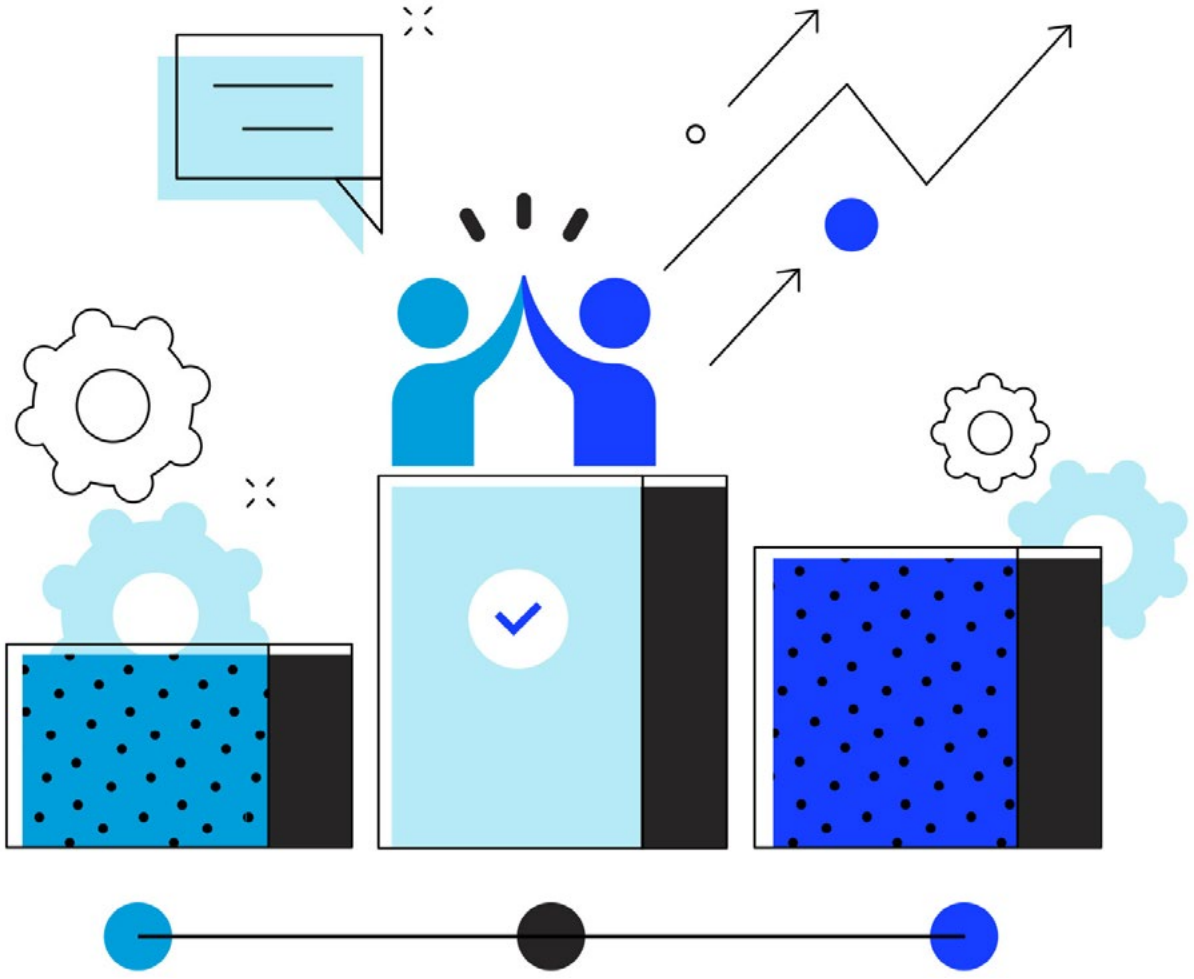
Digital spaces have in many ways served as catalysts for inclusive participation in public life, connecting people across geographical boundaries with shared aspirations for progress. When harnessed for good, these spaces can help empower individuals and give agency to those who are often excluded and marginalized.

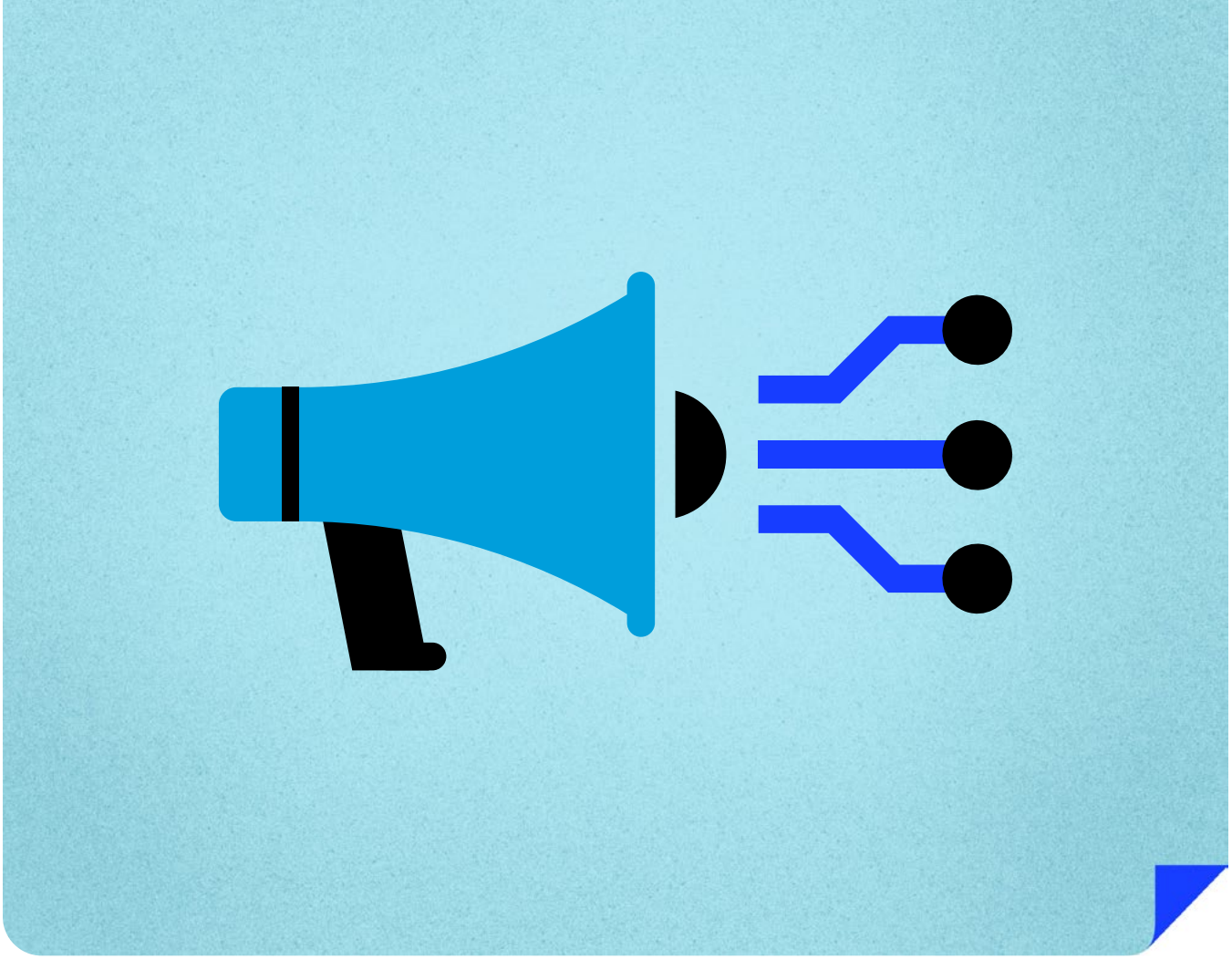
At the same time, digital technologies can hinder genuine empowerment. Individuals frequently have little control over how their personal data are used, or over algorithmic content personalized by large technology companies, and face obstacles from information providers to understanding and accessing the criteria and mechanisms used by them in prioritizing and promoting specific types of content.

Technology companies should empower users to provide input and feedback on all aspects of trust and safety, privacy policy and data use, recognizing user privacy rights. They should enhance user control and choice, including interoperability with a range of services from different providers.

Media, information and digital literacy training initiatives should focus attention on empowering all, in particular focusing on the specific challenges faced by women, older persons, children, youth, persons with disabilities and groups in situations of vulnerability and marginalization.

While Internet connectivity is growing, a third of the world remains offline. Even for those online, inadequate access can hinder their ability to fully harness the Internet's resources, which leaves them vulnerable to risks in information spaces. As barriers to connectivity rapidly fall, initiatives need to be put in place to empower new Internet users and equip those lacking access with the digital literacy skills necessary for safe and productive online experiences.





Independent, Free and Pluralistic Media

Information integrity is achievable only with an independent, free and pluralistic media.

A free press underpins the rule of law and serves as a cornerstone of democratic societies, enabling an informed civic discourse, holding power to account and protecting human rights. The press can be considered free wherever journalists and media workers—including women and those in vulnerable and marginalized situations—are consistently at liberty to report and operate safely and openly, and all individuals have consistent access to pluralistic and reliable sources of news.

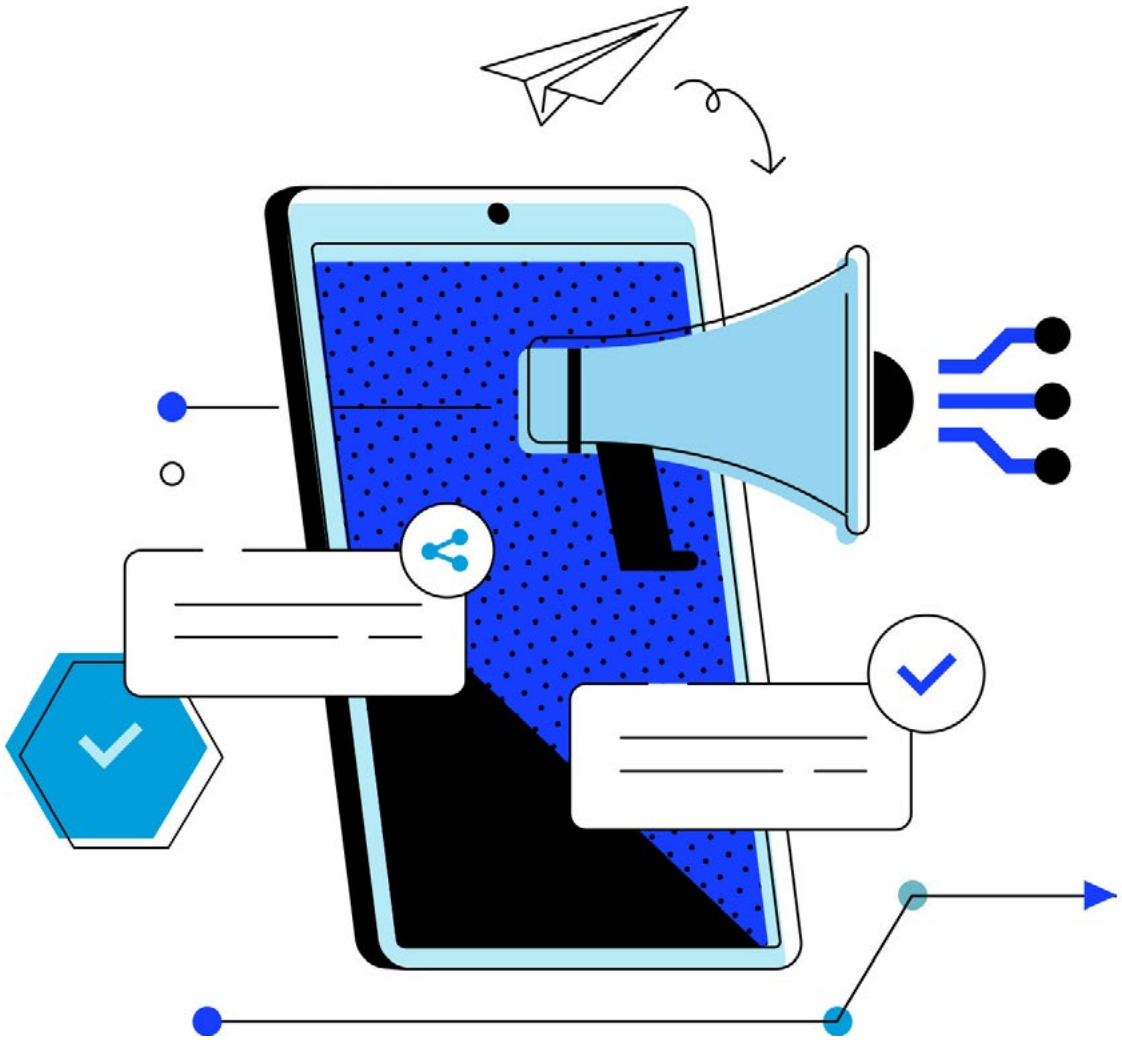
The media have a particular role and responsibility in providing reliable and accurate information and mitigating risks to information spaces. However, press freedom faces significant and sustained threats around the world despite the right to freedom of expression, including free, uncensored and unhindered press or other media. Media workers face online and offline harassment, threats and

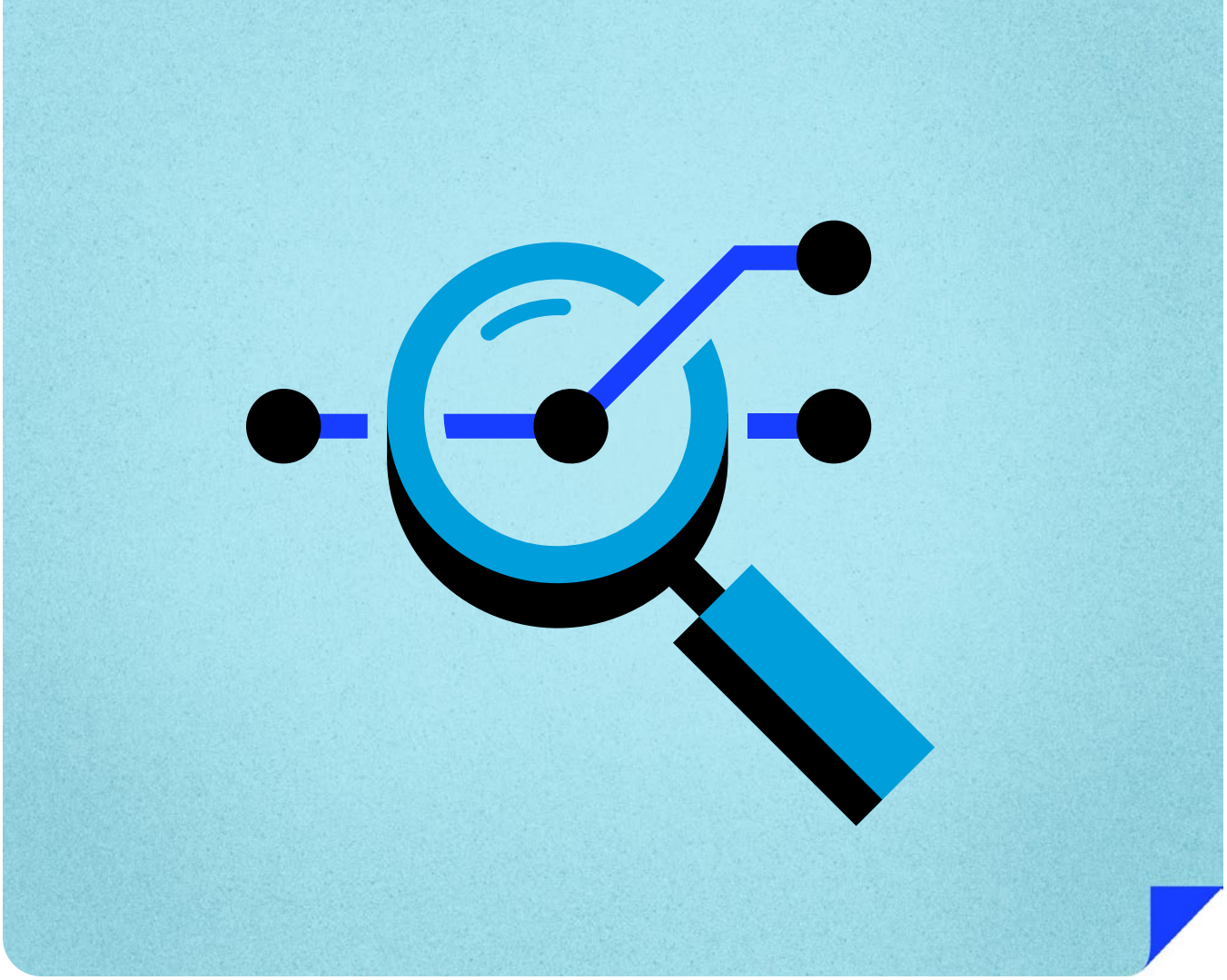
violence, leading at times to self-censorship and heightening professional risk.

At the same time, the news industry has suffered from the migration of advertising revenue to the digital space, which is dominated by large technology companies. These factors have allowed corporate interests to further tighten their grip on media outlets, threatening media diversity and undermining local and public interest journalism. Where editorial standards are not robustly upheld, media outlets can drive and amplify risks to information integrity, which can cross-pollinate between online and offline spaces.

Robust and urgent responses are needed to support public interest news organizations, journalists and media workers, acknowledging contexts with limited media infrastructure where citizen journalists provide a vital service for local constituents. Such responses can include robust and sustained media development assistance, utilizing local implementers.

States and technology companies wield considerable influence in shaping information flows and policy and should strengthen efforts to ensure press freedom and the immutable safety of journalists.





Transparency and Research

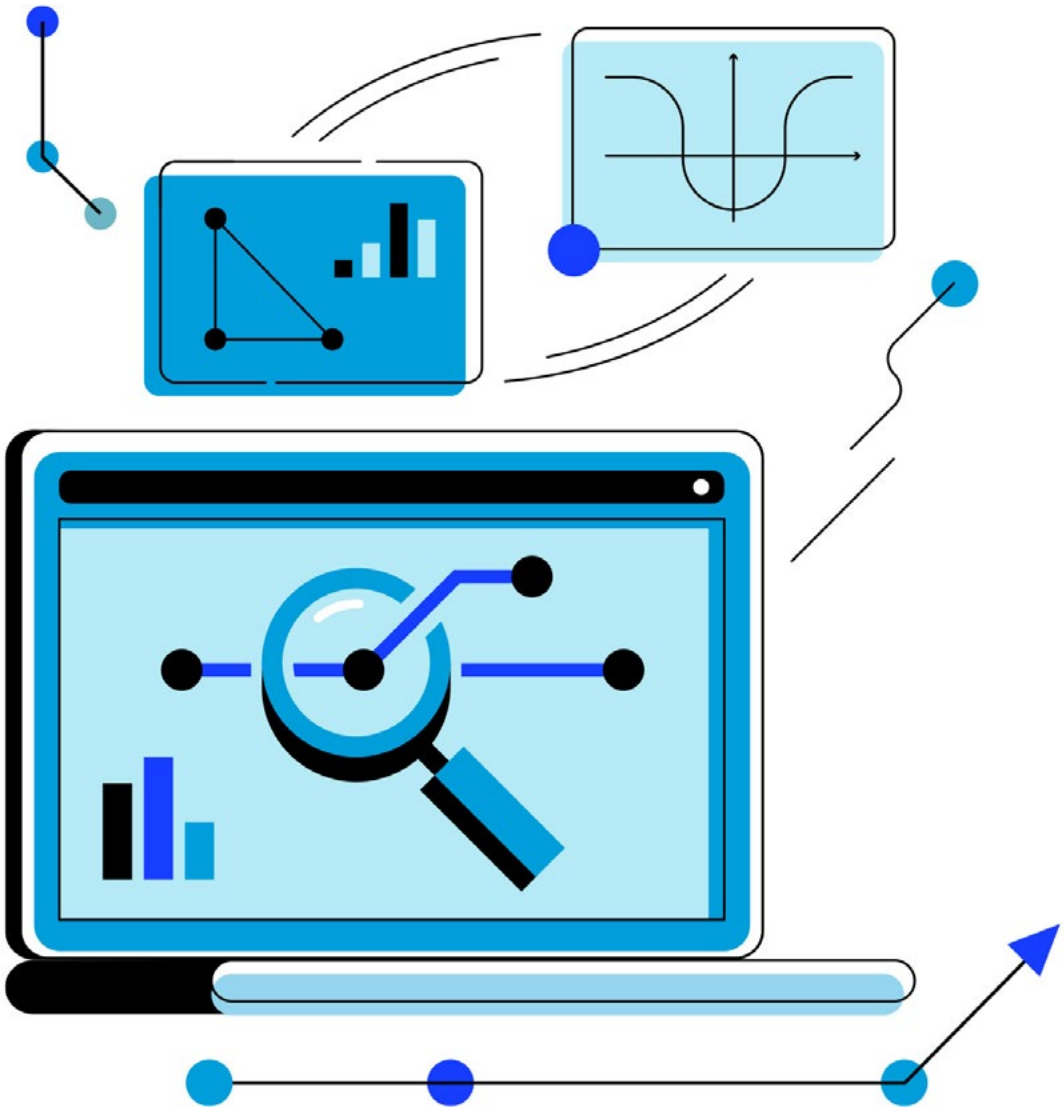
Increased transparency by technology companies and other information providers can enable better understanding of how information is spread, how personal data are used and how risks to information integrity are addressed.

However, power imbalances create barriers to transparency. A handful of technology companies enjoy access to an unprecedented volume of data and, along with some media owners, have significant control in the information ecosystem, sometimes in close relationship with States, political and economic actors.

Furthermore, regulatory choices around transparency made in a small number of countries where the majority of technology companies are headquartered deeply affect the rest of the world. These imbalances often restrict public interest research and can hamper efforts to ensure equity and to meet the needs of underserved and underresearched contexts and communities.

The deployment of AI technologies, the full impact of which remains unknown, adds further challenges in researching and understanding the information ecosystem.

Establishing a more nuanced global understanding of information environments and enhancing targeted, evidence-based actions for the promotion of information integrity will require expanding the availability, quality and usability of data and insights. Ensuring privacy-preserving data access for a diverse range of researchers will strengthen collective efforts to fill research gaps and inequalities. Academics, journalists and civil society must be protected and supported in carrying out their vital work free from fear or harassment.



Calls to Action

The aim of the following recommendations is to operationalize the five principles into actionable steps for stakeholders across the information ecosystem. Intended as a holistic blueprint, these recommendations range from the legal obligations of States to the responsibilities of the technology sector to best practices for media and civil society.

RECOMMENDATIONS FOR **STAKEHOLDERS**

- TECHNOLOGY COMPANIES
- ARTIFICIAL INTELLIGENCE (AI) ACTORS
- ADVERTISERS AND OTHER PRIVATE SECTOR ACTORS
- NEWS MEDIA
- RESEARCHERS AND CIVIL SOCIETY ORGANIZATIONS
- STATES AND POLITICAL ACTORS
- UNITED NATIONS

Technology Companies

Large technology companies, many headquartered in locations where technology regulation is limited, wield immense power. They profit from vast troves of data harvested on user behaviour, allowing them to shape transnational information flows and control digital experiences on a global scale.

To redress this power imbalance, a framework is needed that prioritizes both transparency and independent oversight. Users deserve control over their data and online experiences, with clear avenues for complaint and redress. Accountability mechanisms are needed to address the responsibility of technology companies for the consequences of the design and use of their products and services on human rights and social cohesion, including in crisis and conflict situations.

This will require a critical and transparent assessment of platform architecture to identify features that erode information integrity and undermine human rights. Strategies to prevent and mitigate such erosion should be implemented while safeguarding freedom of expression and access to information.

Disinformation and hate should not generate maximum exposure and massive profits. New commercially viable business models that do not rely on targeted programmatic advertising could foster innovation, increase user empowerment and serve the public interest. This multifaceted approach can create a more balanced information ecosystem that respects user rights and fosters a trustworthy online environment.

Recommendations

a. Integrate safety and privacy from design to delivery. Embed robust safety and privacy policies into the full life cycle of all products and services, including every phase of design, development, delivery and decommission, applying policies consistently to both human and AI-generated media. Cooperate with independent, third-party organizations to conduct and make public ongoing human rights risk assessments related to all products and services to proactively minimize societal risks and mitigate potential harms, including in advance of and around pivotal societal moments. Take measures to protect and empower groups in situations of vulnerability and marginalization, members of civil society and others often targeted online; and to address gender-based and other forms of violence which occur through or are amplified by the use of technology. Innovate to address emerging challenges including the potential prevalence of risks to information ecosystem integrity resulting from AI technologies. Ensure diversity and inclusion in staffing at all stages of product development, and in trust and safety teams. Establish procedures for internal information-sharing to ensure that risk and policy assessments are shared and collectively understood at all levels and functionalities of the company, including leadership. Ensure consistent enforcement of all trust and safety policies.

b. Re-evaluate business models. Assess whether and how platform architecture contributes to the erosion of information ecosystem integrity and undermines human rights, and take proportionate mitigation and remediation measures while respecting freedom of expression. Scope innovative, commercially viable business models that do not rely on targeted programmatic advertising and that serve the public interest.

c. Protect children. Establish and enforce measures to protect and uphold the rights of children, such as age verification and parental controls. Implement policies and practices to prevent and counter child sexual exploitation and abuse which occurs through or is amplified by the use of technology. Establish and publicize special reporting and complaints mechanisms for children.

d. Allocate resources. Allocate sufficient and sustained dedicated in-house trust and safety resources and expertise that are proportionate to risk levels. Designate sufficient resources to address sociocultural linguistic contexts and languages of operation and the differentiated needs of groups in situations of vulnerability and marginalization, in particular in contexts experiencing conflict or facing unstable conditions.

e. Ensure consistent content moderation. Cooperate with independent, third-party organizations to develop content moderation processes in line with international human rights standards and ensure that such policy is enforced consistently and non-arbitrarily across areas of operation. Allocate sufficient resources for human and automated content moderation and curation, applied consistently across all languages and contexts of operation. Take measures to address content that violates platform community standards and undermines human rights, such as limiting algorithmic amplification, labelling and demonetization. Make publicly available disaggregated data on the implementation of content moderation policies and on resources allocated for content moderation across languages and contexts of operation.

f. Uphold labour standards. Provide working conditions that are aligned with international labour and human rights law and prioritize initiatives that ensure the welfare, safety and quality training of all workers, including content moderators, involved in trust and safety efforts.

g. Establish independent oversight. Commission regular external human rights independent audits, which cover terms of service and community standards; trust and safety and advertising policies; risk management; the impacts of advertising and recommender systems across language and operational contexts; content moderation; complaints and appeals processes; transparency mechanisms; and data access for researchers. Assess the impact of products and services on groups in situations of vulnerability and marginalization, on gender equality and on children's rights. Make the results of these audits public, accessible and understandable for all users.

h. Develop industry standards. Partner with civil society and other stakeholders to co-develop industry accountability frameworks with clearly defined roles and responsibilities, committing to audited public reporting and independent oversight and to robust standards for privacy, transparency, risk management and trust and safety. Make specific provisions for the needs of groups in situations of vulnerability and marginalization and in fragile contexts, establishing effective ways to measure and address risks to human rights. Ensure cooperation between platforms and services, recognizing that risks can spread across various information spaces, each with unique design flaws and policy gaps that can be exploited.

i. Elevate crisis response. Working with stakeholders operating in high-risk areas, establish early warning and escalation processes with accelerated and timely response rates in contexts of crisis and conflict. Establish mechanisms to enable prominent, timely access to reliable, accurate information that serves the public interest.

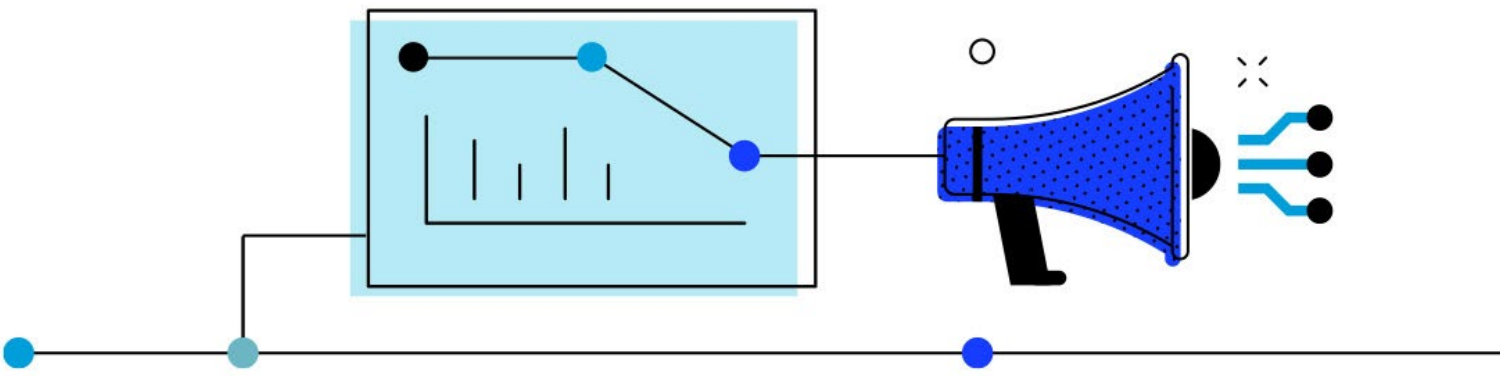
j. Support political processes. Undertake and make publicly accessible human rights risk assessments of all products and services in advance of and throughout elections and other political processes. Enforce all related policies to uphold information integrity, taking measures to address disinformation, harassment and violence against women and other groups commonly targeted in public life, including political candidates.

k. Collaborate with stakeholders. Proactively engage with a diverse range of stakeholders, including States, academia, civil society, children, youth-led organizations and the technical community, to gain deeper understanding of risks to the integrity of the information ecosystem and augment and calibrate trust and safety mechanisms accordingly.

l. Establish robust complaint mechanisms. Ensure transparent, safe, secure and accessible user and non-user complaint, reporting, appeals and redress mechanisms in a timely manner, including special processes for those in situations of vulnerability and marginalization. Establish and enforce procedures to prevent misuse of the reporting and complaints mechanisms, such as through coordinated inauthentic behaviour.

m. Communicate clear policies. Make terms and conditions, policies, community standards and enforcement procedures easily accessible, consistent and understandable, including for children. Make clear all policies, guidelines and rules concerning news and political content.

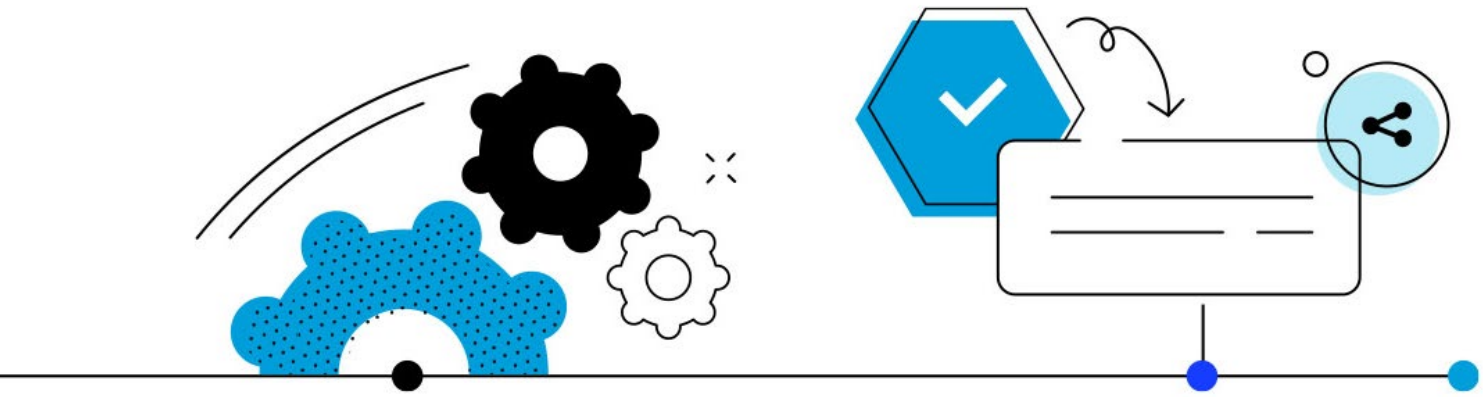
n. Enforce advertising policies. Establish, publicize and enforce clear and robust policies on advertising and the monetization of content. Review existing publisher and advertising tech partnerships on an ongoing basis to assess whether such policies are upheld by partners in the ad tech supply chain. Publicly report annually on the effectiveness of policy enforcement and any other actions taken.



o. Demonstrate advertising transparency. Clearly mark all adverts, making information on the advertiser, the parameters used for targeting and any use of AI-generated or -mediated content transparent to users. Maintain full, accessible, up-to-date and searchable advertising libraries with information on the source or purchaser, how much was spent and the target audience. Give detailed data to advertisers and researchers on exactly where adverts have appeared in any given timescale, and the accuracy and effectiveness of controls and services around advertising placements and brand adjacency. Undertake transparent reporting regarding revenue sources and sharing arrangements with advertisers and content creators. Clearly label all political advertising, including to indicate content that has been AI-generated or -mediated, and provide easily accessible information on why recipients are being targeted, who paid for the adverts and how much.

p. Support media safety and diversity. Create an enabling environment for the distribution of pluralistic news content, allowing consumers to access a range of media sources. Support independent, free and pluralistic media, especially local and citizen journalism conducted in diverse languages and contexts, while respecting editorial independence. Take all measures to uphold the rights of journalists and media workers online. Make explicit, transparent provisions to help safeguard journalists and media workers against harassment, abuse and threats of violence, reflecting the risks faced by journalists, especially during pivotal societal moments such as elections, natural hazards and human-made crises. Update trust and safety policies and practices specifically to mitigate and address the targeting of women journalists.

q. Provide data access. Provide researchers, including academics across disciplines, journalists, civil society and international organizations, access to the data that they need to better understand information integrity, inform policy and best practice and improve accountability, while respecting user privacy and intellectual property. Such data should be disaggregated to allow for effective study of information ecosystem integrity, including societal risks, impacts on differentiated communities and populations, the implications of the use of AI technologies, potential impacts on the achievement of the Sustainable Development Goals and the effectiveness of risk



mitigation measures. It should include information on: algorithm-driven recommender systems, including explanations of how algorithms are trained to rank, recommend, distribute and flag content; accounts removed, banned or demoted; and resource allocation for trust and safety across languages and contexts. Facilitate data delivery for researchers at minimal cost in accessible, machine-readable formats.

r. Ensure disclosure. Make public State requests for content removal or placement. Disclose all collaborations with fact-checking organizations, including funding or other support provided; and funding provided to political bodies and candidates.

s. Offer control and choice. Offer user-friendly tools, functions and features that ensure informed consent and empower people to easily control their own online experience, including through interoperability with other services, allowing greater choice and providing informed consent over the content they see and how and where their data are used.

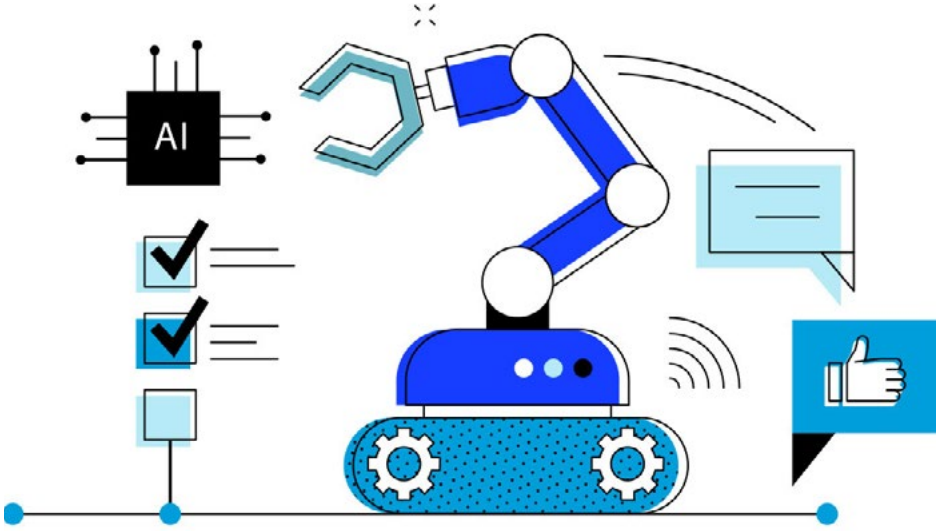
t. Label AI content. Clearly label AI-generated or -mediated content, investing in and developing solutions at the organizational level, to ensure that users can easily identify such content and to strengthen rather than undermine user trust in information ecosystem integrity more broadly. This includes information in the metadata that identifies such content as AI-generated or -mediated.

u. Ensure privacy. Ensure that the collection, use, sharing, sale and storage of data respects the privacy of users and that users can easily access information on how their personal data are harnessed, including for algorithmic decisions, and on how their personal data are shared with and obtained from other entities.

v. Foster digital literacy. Support media and information literacy drives to boost digital skills, including to improve public understanding of the function, effects and implications of algorithms. Dedicate literacy and capacity-building resources for all languages and areas of operation, especially fragile contexts. Provide safety-related training materials to children and youth. Enable and make publicly available independent external evaluations of the effectiveness of literacy initiatives.

Artificial Intelligence (AI) Actors

Government, private and public sector actors involved in at least one stage of the artificial intelligence system life cycle



As artificial intelligence (AI) technologies rapidly evolve, these capabilities are poised to reshape our world. From automating everyday tasks to aiding scientific discovery, the potential benefits are vast. However, alongside this progress lies a critical need to ensure safe, secure and trustworthy AI design, development, deployment and decommission.

Bias and lack of diversity in training data can lead AI systems to generate misleading information and perpetuate unfairness. This ability to generate realistic content can be misused at scale to create risks to the integrity of the information ecosystem.

Emerging risks can be mitigated by prioritizing transparency and fairness in the life cycle of AI technologies. A collaborative effort across government, technology companies and academic and research institutions is needed to ensure that AI is designed, developed, deployed and decommissioned safely and responsibly across its life cycle. By working together, these stakeholders can ensure that AI technologies benefit society and human well-being.

Recommendations

a. Ensure safe, secure and trustworthy AI. Take measures to ensure the safe, secure and trustworthy design, development, deployment, use and decommission of AI technologies. Address and publicly communicate the implications of any innovations or advancements in the field that may present risks to the integrity of the information ecosystem, including malicious uses of AI technologies, overreliance on AI technology without human oversight and any related potential for further erosion of trust across geographies and societal contexts. Train AI on reliable, inclusive information sources on issues critical to public well-being and take measures to mitigate bias stemming from training data, including on gender and racial bias. Partner with a diverse range of stakeholders in carrying out human rights risk assessments to proactively minimize societal risks and mitigate potential harms, including to women, children, youth and other groups in situations of vulnerability and marginalization.

b. Commission independent audits. Commit to providing access and legal and technical safe harbour to institutional and individual researchers to conduct independent audits of AI models, with appropriate safeguards, such as compliance with company vulnerability disclosure policies. Ensure public accessibility of the results of independent audits, data about risks related to AI systems—such as the potential for harmful discrimination and "hallucinations", namely, content that appears factual but is completely made up—and steps taken to prevent, mitigate and address potential harms.

c. Respect intellectual property. Respect intellectual property rights, ensuring fair compensation for use of intellectual property, including original journalism, used in training AI tools.

d. Display data provenance. Develop and implement solutions and policies on provenance, through visible and invisible forms, such as authenticity certification, watermarking and labelling. Undertake multi-stakeholder efforts towards the standardization of user-friendly labelling.

e. Support literacy. Invest at the organizational level in the development and deployment of literacy initiatives to enhance public understanding of how AI models function and the implications for information consumers globally, with a focus on risks to information integrity.

f. Enable user feedback. Provide users the ability to alert or report inaccurate or misleading provenance information, while protecting user privacy.

Advertisers

Advertisers can exert singular influence on the integrity of the information ecosystem by helping to cut off financial incentives for those seeking to profit from disinformation and hate. In doing so, advertisers can better protect their brands and address material risk, boosting their bottom line while conducting business in line with their corporate values.

Recommendations

a. Establish human rights-responsible advertising. Establish safeguards to ensure that advertising does not drive risks to information spaces and upholds human rights, including the rights of children. Avoid discriminatory targeting practices based on sensitive data and perceived user traits. Advertise with media outlets and platforms that bolster information integrity, including public interest journalism, through methods such as inclusion and exclusion lists, ad verification tools and manual vetting. Require ad tech companies to publish criteria that a website or channel must adhere to before they are able to monetize.

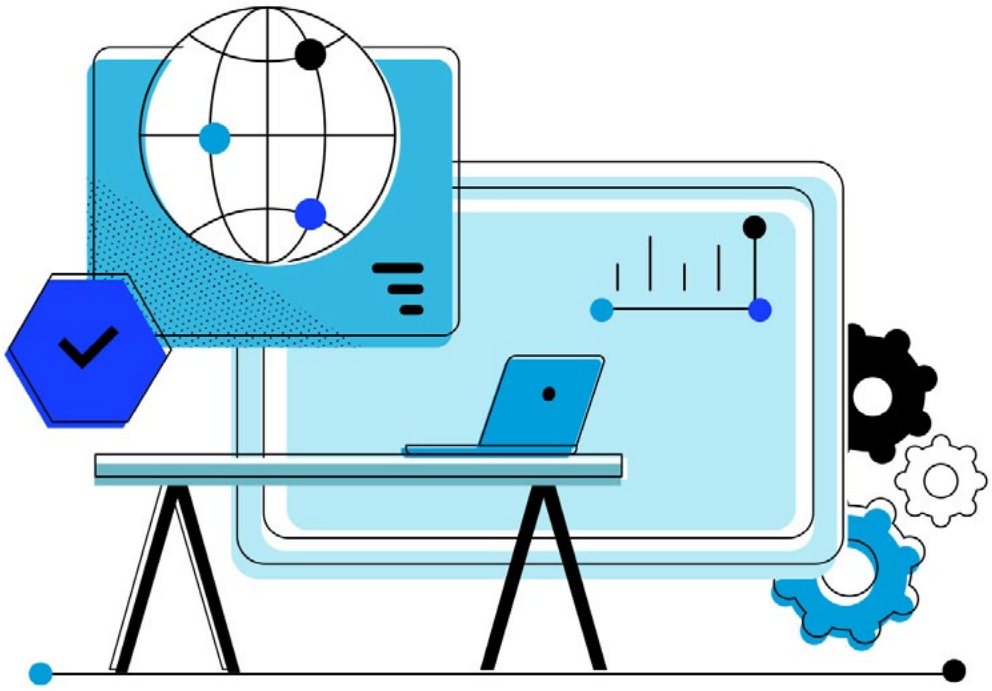
b. Harness industry standards. Make use of industry standards to develop clear policies to minimize risks to information integrity and help ensure brand safety.

c. Form coalitions. Collaborate across industry and with civil society to share best practices and lessons learned about information integrity in a timely manner, including assessing the impacts of advertising, and systematic mitigation of risks and potential harms stemming from advertising and content monetization.

d. Require data. Establish a full and detailed overview of advert adjacency on an ongoing basis, requiring granular data showing where adverts have appeared and conducting suitability reviews before advert placement. Carry out thorough audits of advertising campaigns.

e. Obligate transparency. Require ad tech companies to adopt transparency standards that enable end-to-end validation of the advertising tech supply chain, and share full advertising campaign data with clients and researchers including placement and blocking data at the log level.

f. Undertake audits. Require ad tech companies to carry out independent third-party audits and vetting of ad exchange supply partners.



Other Private Sector Actors

The actions of a broader range of private sector entities not directly involved in the technology sector can impact information spaces, serving to both weaken and support information integrity. Businesses have a responsibility to respect human rights, including the right to freedom of expression and to information, and can form collaborative partnerships with other stakeholders to help achieve a healthier information ecosystem.

Recommendations

a. Uphold integrity. Uphold human rights, including the right to freedom of expression and opinion and refrain from wilfully spreading or sponsoring risks to the integrity of the information ecosystem for financial or any other strategic goal.

b. Invest in literacy. Invest in media and information literacy for personnel at the organizational level, partnering with and calling on the expertise of relevant civil society actors.

News Media

Independent, free and pluralistic media serve a critical role in informing the public on matters of public interest, fostering civic engagement and promoting accountability of those in power.

Direct and indirect threats to media independence, freedom and diversity, and the decline of local and public interest journalism, however, can undermine these vital functions. Where professional standards are not rigorously maintained, news media can erode information integrity. Through ethical reporting and editorial practices and a commitment to transparency, bolstered by quality training and working conditions, journalists offer an indispensable service and can help restore balance in the face of risks to the integrity of the information ecosystem.

Recommendations

a. Cover information integrity. Invest in capacity-building for data-driven and investigative journalism to proactively cover and inform the public about risks to the integrity of the information ecosystem. Employ robust editorial processes and standards, including in the sourcing of information, to help maintain and secure trust among media consumers. Establish fact-checking mechanisms as a reference for the public.

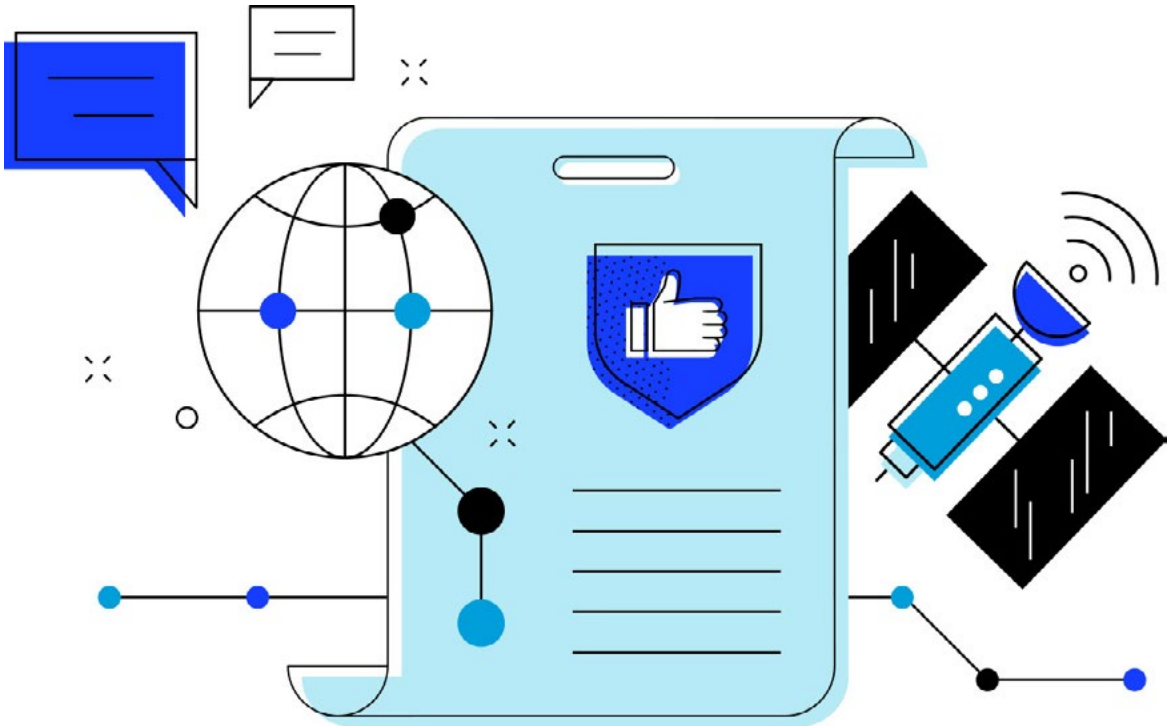
b. Provide crisis response. Commit to providing cost-free and timely information to the public during emergency and crisis situations when risks to information ecosystem integrity may be heightened.

c. Maintain professional and ethical standards. Commit and adhere to globally recognized norms and practices of professional and ethical journalism produced in the public interest, emphasizing impartiality and editorial independence, and actively adopt self-regulatory accountability mechanisms. Provide periodic, quality training to advance ethical, accurate and impartial reporting, and to update skills for promoting innovation and adaptability to changes in the communications landscape, including by adopting a “solutions” or “constructive” journalism approach. Disclose funding sources, ownership structure and financial incentives so that individuals can be better informed about the news that they select and consume.

d. Use AI ethically. Establish robust policies for ethical use of AI technologies, including clearly labelling AI-generated or -mediated material when publishing or broadcasting. This includes information in the metadata that identifies such content as AI-generated or -mediated.

e. Establish transparent, human rights-responsible advertising. Take steps to ensure that advertising does not drive risks to information spaces. Clearly distinguish between news, opinion and sponsored content, and ensure transparency as to the funding of opinion pieces and potential conflicts of interest. Clearly mark all paid and AI-generated or -mediated advertising and advertorial content. Provide transparent reporting on advertising revenue sources and clear, accessible advertising policies and practices.

f. Respect labour standards. Establish working conditions aligned with international labour and human rights law and prioritize initiatives that help ensure the welfare and safety of journalists, including in digital spaces, and pay specific attention to discrimination, abuse, harassment and threats of violence against women journalists and media workers.



Researchers and Civil Society

Researchers and civil society organizations are pivotal to efforts to understand and address the multifaceted impacts of risks to the integrity of the information ecosystem. Their efforts can help expose risks to information spaces, strengthen the evidence base for advocacy and foster resilience, in particular for groups in situations of vulnerability and marginalization. Collaborative partnerships and knowledge exchange are essential for bridging the gap between research insights and effective solutions.

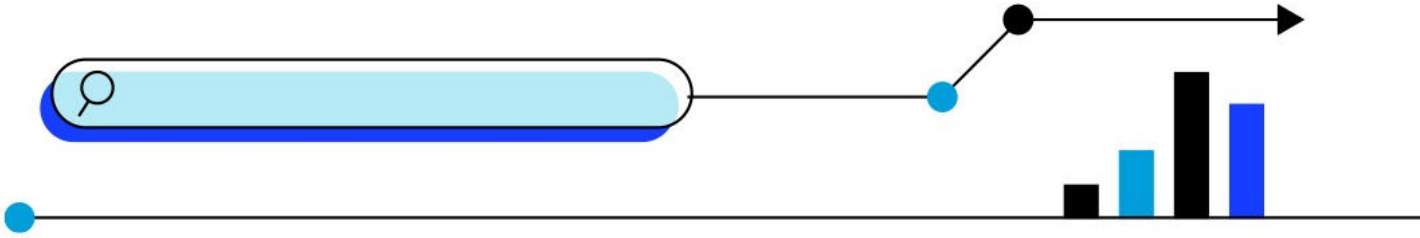
Recommendations

a. Collaborate. Partner with stakeholders across geographies and contexts to share effective and ethical approaches to strengthening the integrity of the information ecosystem.

b. Uphold integrity and ethical standards. Uphold human rights and refrain from wilfully spreading or sponsoring risks to information integrity. Conduct all research in an ethical, transparent and privacy-conscious manner.

c. Promote open access. Undertake open access initiatives, making research freely available, promoting collaboration across disciplines.

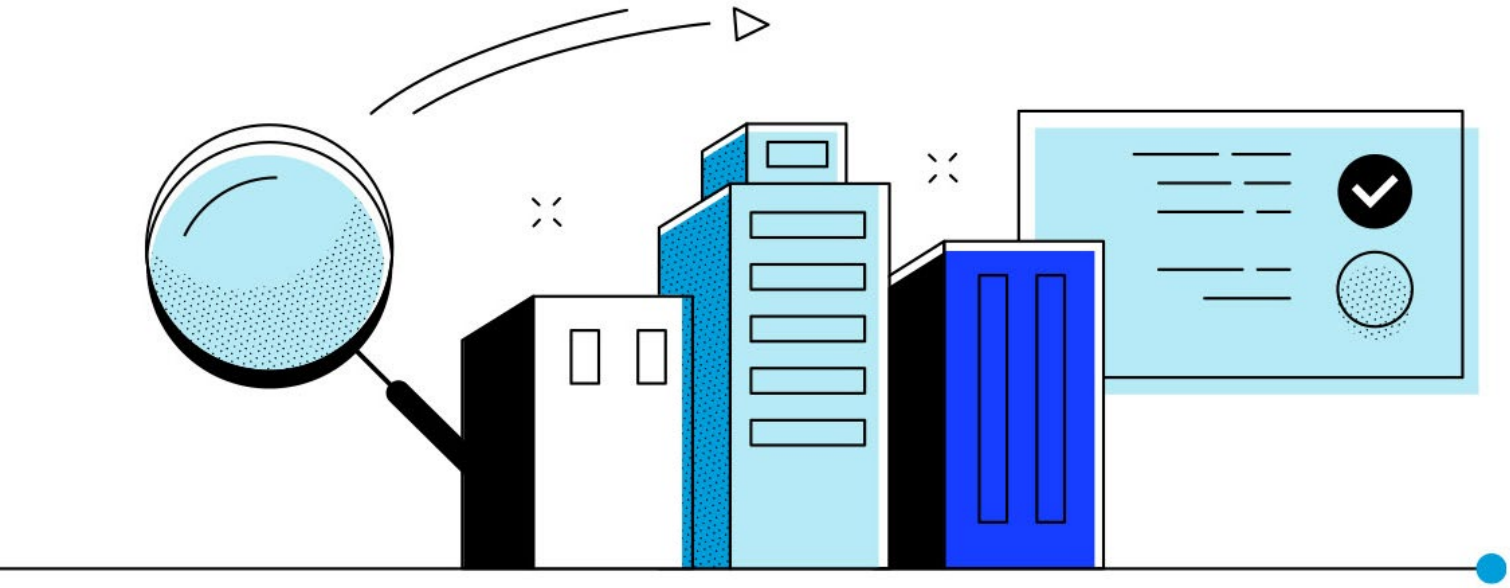
d. Strengthen inclusive research. Explore multidisciplinary research on the information ecosystem across geographies, languages and thematic areas, including the potential impact of risks to information integrity on the Sustainable Development Goals, with particular focus placed on under-studied, vulnerable and marginalized contexts and communities. Develop rigorous methodologies for measuring such risks and related harms.



RECOMMENDATIONS FOR **FACT-CHECKING ORGANIZATIONS** AND **NETWORKS**

a. Maintain professional standards. Commit to professionalism and ethics and adhere to standards of independence, non-partisanship and transparency evident in organizational composition and governance, funding sources, ownership and working practices.

b. Disclose funding. Take measures to publicly disclose funding sources and any collaboration with stakeholders such as technology companies, media outlets and civil society organizations.



States

States shoulder an indispensable responsibility for fortifying the Global Principles for Information Integrity. This begins with State obligations to respect, protect and promote human rights, in particular the right to freedom of expression, including the right to seek, receive and impart information.

States play a central role in shaping information spaces owing to their legal and regulatory authority, control over public resources and ability to build domestic and international coalitions, among other factors. As part of their human rights obligations, States must protect against human rights abuses within their territory and/or jurisdiction by businesses, taking appropriate steps to prevent, investigate, punish and redress such abuse through effective policies, legislation, regulations and adjudication.

States have different technical and financial capacities when engaging with the information ecosystem. Gaps in infrastructure and in access to technology and financial resources have contributed to a digital divide. At the same time, many large technology companies, while achieving near global market penetration and dominance, are based in a small number of countries in the global North.

To ensure that all States can contribute to and benefit from the information ecosystem, urgent and sustained initiatives are needed to increase the ability of States to expand digital connectivity, proactively head off the potential emergence of an "AI divide" and strengthen their capacities to adequately address risks in information spaces, while respecting human rights. Ultimately, these efforts will strengthen information integrity, promote human rights and contribute to achieving the Sustainable Development Goals.

Recommendations

a. Respect, protect and promote human rights. Respect, protect and promote human rights, in particular the right to freedom of expression and opinion, including the right to information, based on international human rights standards and norms. Ensure that regulation or other measures implemented to address the various elements of information integrity comply with applicable international law, including international human rights law, with the full participation of civil society, forming part of a broader effort to strengthen human rights and build trust. Ensure that restrictions on the right to freedom of expression are exceptional, and, where restrictions are imposed, they must comply with the requirements under international human rights law, i.e. be provided by law, and be necessary for the protection of the rights or reputations of others, or of national security, public order, or public health or morals, and comply with the

principles of proportionality. Ensure that restrictions do not serve in practice to stifle freedom of expression. Adopt and effectively enforce protections for personal data privacy that are consistent with international law, including international human rights law.

b. Safeguard integrity. Abstain from conducting or sponsoring information operations, domestically or transnationally, that wilfully spread disinformation or harness hate speech. Refrain from any form of Internet shutdowns or throttling. Uphold and implement relevant United Nations Security Council resolutions, including those relating to the protection of United Nations peace operations from risks to information ecosystem integrity impacting mandate implementation.

c. Protect populations. Reaffirm and redouble efforts to ensure in law and in practice the protection and empowerment of groups in situations of vulnerability and marginalization which are often targeted in online and offline information spaces, such as women or LGBTIQ+ individuals or minority ethnic or religious groups, while addressing the particular needs and rights of children. Comply with the obligation under international human rights law to prohibit by law propaganda for war or advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.

d. Provide access to information. Provide timely access to publicly held information, without discrimination—including for news media—in all languages and formats that are understandable and usable for all, while promoting access for underserved communities. Ensure access to reliable and accurate information in crisis situations. Adopt ethical trustworthy communications practices to proactively engage communities and build trust in public institutions.

e. Ensure media freedom. Ensure, protect and promote a free, viable, independent and pluralistic media environment, taking robust measures to safeguard journalists, media workers and fact checkers, with particular attention to women and members of groups in vulnerable and marginalized situations, from all forms of discrimination, abuse, harassment and threats of violence. Respect and protect in law and in policy the rights of digital content creators and citizen journalists.

f. Protect researchers and civil society. Protect academics and civil society from intimidation, harassment or retaliatory action, respecting academic freedom.

g. Provide transparency. Provide full transparency regarding requirements and data requests placed on technology companies and media organizations. Take measures to address non-transparent and deceptive lobbying tactics and conflicts of interest between technology companies and policymakers that undermine information integrity, such as unethical hiring practices and financial incentives.

h. Strengthen global solidarity, capacity-building and development assistance. Engage in collaborations and partnerships between countries to support capacity-building to strengthen information integrity and increase resilience to risks to information spaces, in particular in developing countries. Allocate financial resources, with full transparency, for training and capacity-building in digital, information and media literacy and awareness programmes, including in AI technologies, in all languages. Support developing countries in nationally led efforts to build societal resilience to risks to the integrity of the information ecosystem, undertake robust media and information literacy training and bolster public interest media, including through dedicated and adequate development assistance. Support the work of public institutions, including libraries, in improving access to literacy training and resources.

i. Promote political participation. Protect the access of all electoral stakeholders to accurate and timely information throughout electoral processes. Take measures to promote inclusive political participation and leadership and to uphold the rights of women in public life, including protection from all forms of discrimination, abuse, harassment and threats of violence.

j. Prioritize inclusive, public-interest research. Prioritize, invest in and support independent research that abides by ethical standards and review across disciplines related to information integrity, including in light of the emergent and the as-yet unknown capabilities and impacts of AI technologies. Support research conducted across geographies, languages and thematic areas, including the potential impact of risks to information ecosystem integrity on the Sustainable Development Goals, focusing particularly on underserved, underresearched and at-risk contexts and communities. Promote and publicize open access to research findings to enable the equitable sharing of information within and between countries.

k. Foster literacy. Foster a critical and informed public discourse through targeted media and information literacy drives, seamlessly integrating digital skills into formal and informal education curricula from an early age. Actively improve public understanding and awareness, including among children, of online rights, how digital information environments work and how personal data are used, taking into account specific social, cultural and linguistic needs of people of all

ages and backgrounds. Prioritize the literacy needs of individuals and groups in vulnerable and marginalized situations, including women, children, youth, older persons, persons with disabilities and the billions about to come online. Undertake literacy efforts around specific problems related to AI technologies and continually update literacy efforts to reflect new and emerging technologies and challenges.

l. Empower children, parents, guardians and educators. Provide sustained resources for children, parents, guardians and educators on safe and responsible digital behaviour, on navigating online media and on understanding children's rights to freedom of expression and information. Involve all parties in developing media and digital literacy guidelines and initiatives for safer online experiences, while harnessing the digital fluency of youth.

RECOMMENDATIONS FOR ALL **POLITICAL ACTORS**

Individuals, groups and entities involved in and influencing political processes

a. Maintain election integrity. Refrain from and publicly denounce efforts to undermine information integrity, including on voter eligibility, polling, ballot counting and results.

b. Protect inclusion. Publicly denounce and take measures to address abuse and harassment targeting candidates and public officials, especially women and members of groups in vulnerable and marginalized situations.

c. Offer transparency. Maintain transparency in communications, including the sources of funding for advertisements and the use of data driven targeting techniques.

The United Nations

The Global Principles for Information Integrity apply to the United Nations and its international civil servants. By adhering to the Global Principles, the Organization sets a compelling example for responsible stewardship of information integrity within the global community.

Scaling up its work to strengthen the integrity of the information ecosystem will contribute to advancing the Organization's mission of securing peace, fostering sustainable development and promoting and protecting human rights for all.

The United Nations Will

a. Scale up efforts. Intensify efforts to strengthen information integrity, including through context-specific research, monitoring, risk assessment, community engagement and coalition building across diverse contexts and languages. Integrate information integrity into programmes and operations to enhance prevention, mitigation and response and identify emerging opportunities and challenges.

b. Support capacity-building initiatives. Assist with capacity-building in States by offering skill development initiatives, including training for young people, to help strengthen information integrity, with particular attention to the needs of developing countries.

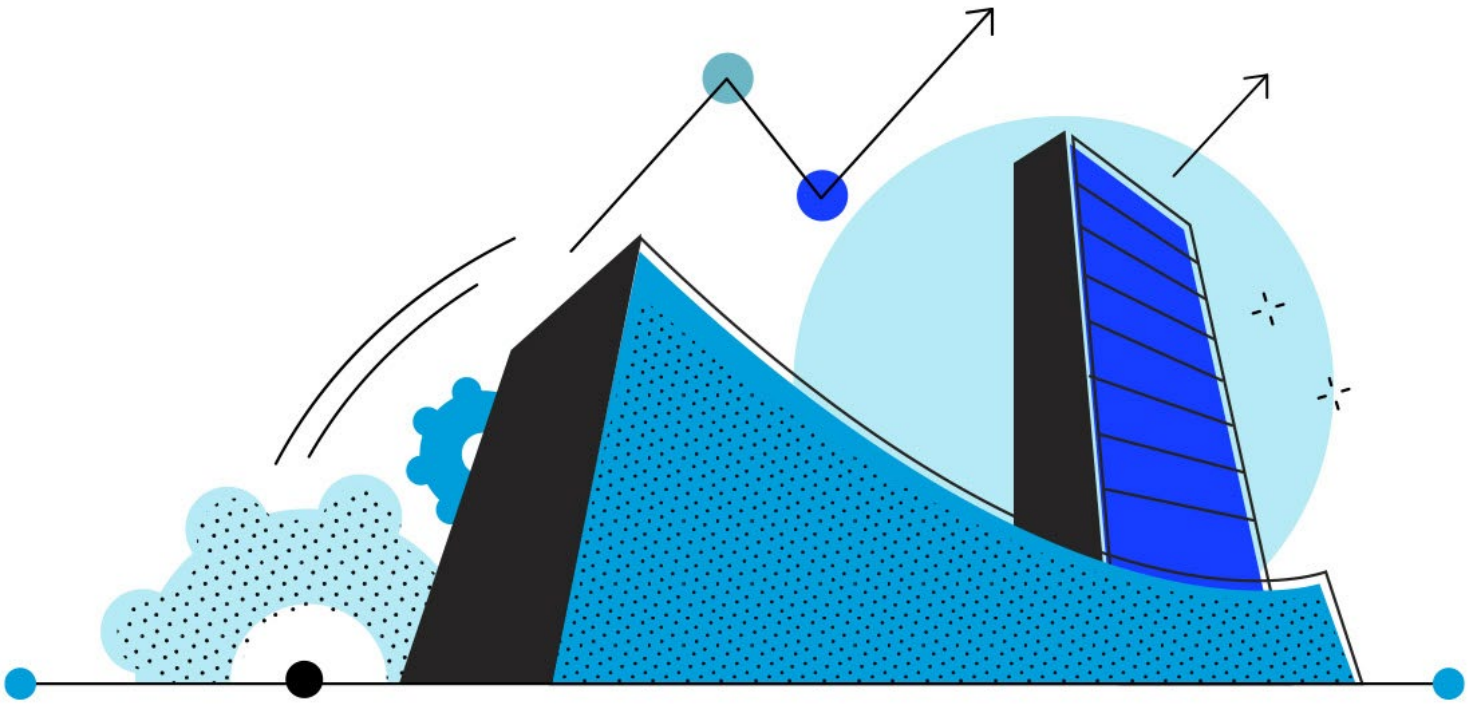
c. Undertake advocacy. Promote and advocate for the Global Principles at the global level and across countries and communities, with particular attention to underserved contexts and groups in situations of vulnerability and marginalization. Actively contribute to social cohesion and strengthen resilience of communities to risks to information integrity, supporting efforts to realize the Sustainable Development Goals.

d. Increase dedicated capacity. Establish a central unit in the United Nations Secretariat to develop innovative and nuanced approaches to addressing risks to the integrity of the information ecosystem affecting United Nations mandate delivery and substantive priorities, coordinating with other capacities and serving the whole United Nations system as required.

e. Develop agile communications strategies. Harness innovative, evidence-based, agile and tailored communication strategies, utilizing digital and offline information spaces for the common good and to better meet the needs of all the people that the United Nations serves.

f. Provide multilingual resources. Establish a multilingual online information integrity resource hub with shared research, guidance and best practices applicable to diverse contexts to support initiatives at the global, regional and national levels.

g. Support multi-stakeholder action plans. Support regional and national multi-stakeholder action plans and coalitions, making use of existing mechanisms and calling on the Organization's expertise and experience in international capacity-building and coordination.



Next Steps

The urgency of strengthening information integrity cannot be overstated in the face of escalating risks to the integrity of the information ecosystem and the emergence of readily available advancements in AI technologies. The Global Principles offer a holistic and unified framework for action to protect and promote information integrity as the world navigates the complexities of the digital age and looks to find multilateral solutions at the Summit of the Future.

To this end, stakeholders are urged to:

- Publicly **commit to, adopt and actively publicize** the United Nations Global Principles for Information Integrity as a framework for immediate action.
- Harness the Global Principles to form and actively participate in **broad cross-sector coalitions on information integrity**, convening diverse expertise and approaches, including for capacity-building, from civil society, academia, media, government and the international private sector, and ensuring full and meaningful youth engagement, such as through dedicated youth advisory groups.
- Collaborate to develop **multi-stakeholder action plans** at the regional, national and local levels, engaging communities to support and learn from grass-roots initiatives and ensuring full and meaningful youth engagement.

By embracing the United Nations Global Principles for Information Integrity, stakeholders from all sectors can demonstrate solidarity and collaboratively forge a path towards a reinvigorated information ecosystem that fosters trust, knowledge and individual choice for all.

Appendix

RESOURCES

1. United Nations Secretary-General's "Our Common Agenda policy brief 8: information integrity on digital platforms" (2023)
https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf

2. UNESCO Guidelines for the Governance of Digital Platforms (2023)
https://www.unesco.org/en/internet-trust/guidelines

3. Report of the Secretary-General, "Countering disinformation for the promotion and protection of human rights and fundamental freedoms", 2022 (A/77/287)
https://www.ohchr.org/sites/default/files/2022-03/NV-disinformation.pdf

4. UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)
https://unesdoc.unesco.org/ark:/48223/pf0000381137

5. United Nations Strategy and Plan of Action on Hate Speech (2019)
www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf

6. Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (2012)
https://www.ohchr.org/en/documents/outcome-documents/rabat-plan-action

7. Guiding Principles on Business and Human Rights (2011)
https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf