From Principles to Practice

STRENGTHENING INFORMATION INTEGRITY

Issue Brief 1
September 2025

United Nations

United Nations Department of Global Communications

EXECUTIVE SUMMARY

Strengthening information integrity has become a critical policy priority worldwide as governments, organizations and communities grapple with the impacts of evolving information risks—including disinformation, hate speech and the erosion of independent, pluralistic media—and the complex influence of major technology platforms, which brings both opportunities and uncertainties. No single stakeholder can address these challenges alone as they transcend sectors and borders in the global information ecosystem.[1]

While the urgency is clear, significant gaps persist in understanding the full scope of these challenges and translating concepts and discussions into actionable solutions. Drawing from a growing body of policy and practice across diverse contexts in support of information integrity, this issue brief, as part of a new series, contributes to efforts to move from principles to practice.

Part I: Why information integrity matters now

- Introduction
- Foundational concept

Part II: Understanding the challenges

- Risks to information integrity
- Cross-cutting challenges

Part III: Solution frameworks

- Theory of change for information integrity
- Multi-stakeholder coalitions
- Human rights guardrails
- International initiatives

Part IV: Operational model

- Research, Risk assessment and Response (3R)
- Conclusion

The accompanying **Guide** in the Annex provides practical guidance on the 3R approach, including detailed risk classification and response measures to support prevention, protection, mitigation and recovery efforts for the integrity of information environments.

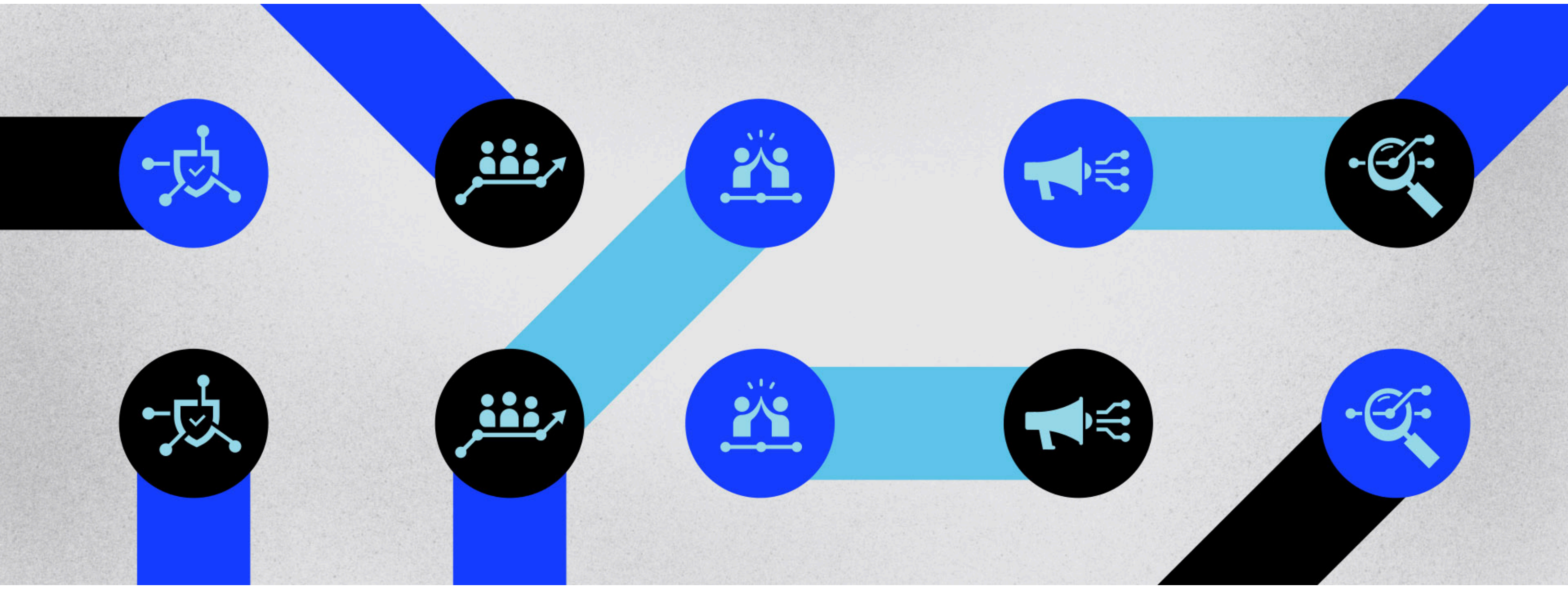
Introduction to the series

This series of issue briefs builds on the foundation of the *UN Global Principles for Information Integrity* (*UN Global Principles*) and related normative frameworks,[2] including the Pact for the Future and the Global Digital Compact.[3]

Each issue brief will focus on a specific, timely topic, drawing from research findings and expert insights. The series aims to open pathways for action, moving from principles to practice.

About the authors

The information integrity team in United Nations Global Communications works at the intersection of policy, research and communications to uphold and strengthen information integrity, a key foundation for United Nations thematic priorities and operational mandates. The team's work contributes to emerging normative frameworks for information integrity, actionable initiatives, coalition-building, strategic communications and tools to help policymakers and practitioners find effective solutions to address evolving risks across information landscapes.



PART I: WHY INFORMATION INTEGRITY MATTERS NOW

INTRODUCTION

The first ever *UN Global Risk Report*, released in July 2025, identified information risks such as misinformation and disinformation as a top global vulnerability—a serious, already unfolding, threat the international community remains insufficiently prepared to address. The report called for the United Nations to strengthen its response capacity to information risks.[4]

Information integrity cuts across virtually every aspect of international and national affairs, from democratic governance and human rights promotion and protection to public health responses, climate action and conflict prevention. Given this broad impact, establishing a coherent framework for understanding and effectively addressing information integrity challenges is essential.

The cross-sectoral and systemic nature of information risks requires meaningful participation from diverse stakeholders to respond to the risks. The *UN Global Principles* call for the inclusion and participation of governments, civil society, academia, technology companies, media, the United Nations system and others.[5] To support multistakeholder cooperation, this first issue brief outlines key foundational concepts and emerging normative frameworks for information integrity. The brief sets out a theory of change to guide practical efforts and introduces elements of an operational model centered around research, risk assessment and comprehensive response, strengthening the integrity of the information ecosystem.

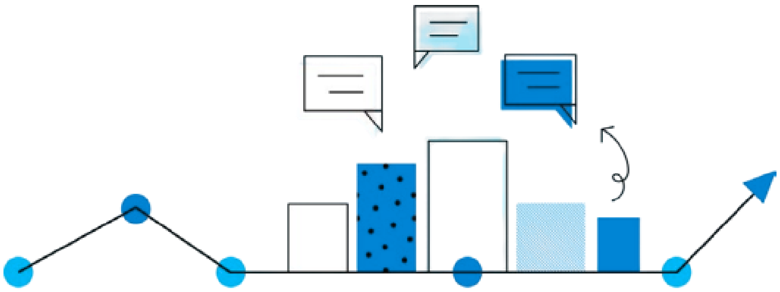
FOUNDATIONAL CONCEPT

The concept of information integrity, as laid out in the *UN Global Principles*, represents an important step towards building a rights-based global information ecosystem that benefits all people. This concept integrates the social, cultural, technological, political and economic dimensions of information, calling for coordinated stakeholder action across geographies to strengthen the information ecosystem. Information integrity has a direct bearing on democratic processes, the rule of law, societal cohesion, sustainable development and individual empowerment.

Information Integrity

Information integrity refers to an information ecosystem in which reliable and accurate information is available to all, enabling people to engage meaningfully in public life, make informed decisions and exercise their rights. This ecosystem is shaped by the actions of a diverse range of actors, including governments, technology companies, media, civil society and individuals.

Strengthening information integrity means protecting the right to freedom of expression and access to information, ensuring inclusive access to a range of information sources and enabling people to navigate information spaces safely, with privacy and freedom. It involves building resilient societies that foster trust, knowledge and public empowerment. Challenges to information integrity encompasses a spectrum of risks, such as disinformation, hate speech, restrictions on press freedom and the malicious use of technologies.



PART II: UNDERSTANDING THE CHALLENGES

RISKS TO INFORMATION INTEGRITY

In this brief, risks to information integrity (or **information risks**) refer to actions, conditions or factors that undermine the integrity of information environments and, in doing so, degrade public access to evidence-based information, decision-making, or societal trust and cohesion. These risks are often enabled and exacerbated by underlying socioeconomic and political factors.

Information risks contribute to broader systemic risks that cut across spatial and sectoral boundaries.[6] For example, in practical terms, information risks can include the declining trust in expertise and evidence-based information sources, erosion of shared understanding or knowledge, information pollution (i.e. oversupply of misleading or false information), lack of access to accurate, reliable information, lack of digital and media literacy, restrictions on freedom of expression, inadequate guardrails and accountability deficits.

Information risks can also be categorized in terms of how they are enabled:

- **Technology-enabled risks**, such as the malicious use of digital media and artifical intelligence (AI) technologies to manipulate, exploit or harm; inadequate trust and safety measures in technology design, development and deployment.

- **Access and distribution risks**, such as the decline of independent journalism (e.g. news deserts) and press freedom, algorithmic biases limiting information availability and diversity; digital divides.

- **Content-based risks**, such as the spread and amplification of disinformation campaigns, hate speech, harassment targeting individuals or groups.

Understanding risks in these different categories allows practitioners and policymakers to address information integrity challenges more effectively, as different risk types require distinct intervention strategies ranging from immediate response to long-term systemic solutions. Risk analysis must consider the scale and scope of risks while considering factors like emerging AI technologies.

It must also consider the motivations and tactics of **adversarial actors**— individuals, groups or organizations, including State and non-State actors, private companies, extremist groups and criminal organizations, that seek to undermine information integrity for financial or strategic gain.[7] These actors often operate across multiple domains simultaneously, exploiting vulnerabilities in digital and offline information spaces. Understanding their methods is essential for those working to protect organizational mandates and build resilience to such threats.

CROSS-CUTTING CHALLENGES

Strengthening information integrity requires addressing interconnected challenges that create both obstacles and opportunities for building a more resilient global information ecosystem. Below is a short overview of some cross-cutting challenges, which will be explored in greater detail in subsequent issue briefs.

- Sectorial silos
- Cross-domain information manipulation
- Systematic targeting of information defenders
- Research limitations and methodological bias
- Perception gaps
- Artifical intelligence and emerging technologies

Sectoral silos

Information risks cut across thematic and geographic boundaries, yet research and responses typically remain compartmentalized in distinct sectors and geographic contexts, such as elections, public health, climate and conflict. More effective responses must apply lessons about addressing information risks and tackling adversarial behaviour across these areas rather than treating each in isolation.

Cross-domain information manipulation

Information risks transcend the artificial boundary between "online" and "offline" environments. As trust in digital platforms erodes, particularly amid uncertainties around emerging AI technologies, offline spaces may gain new significance for individuals and communities seeking reliable information. Meanwhile, adversarial actors—State or non-State—exploit both domains in tandem to shape public perceptions and influence policy outcomes.

Information risks become especially dangerous when adversarial actors use periods of heightened vulnerability, such as elections, economic instability or natural disasters. At such moments, elevated stress and uncertainty make individuals more susceptible to information manipulation.[8] This is compounded when access to reliable information is limited in different spaces and information voids are quickly filled with misleading content and data voids are exploited and misrepresented.[9]

For example, adversarial actors deliberately spread conspiracies and false claims about climate change in the context of extreme weather events, preconditioning communities to dismiss scientific evidence and official guidance.[10] Online conspiracy theories with unclear origins gain legitimacy when repeated offline by trusted local figures, media or pseudo-experts. This distorts local decision-making about safety measures, resource allocation and community preparedness.[11]

Systematic targeting of information defenders

Adversarial actors strategically harass and seek to neutralize credible voices, including individuals working to maintain information integrity. Prime targets include researchers, journalists, fact-checkers, civil society activists and academic institutions who employ rigorous approaches and adhere to ethical standards. Harassment campaigns aim to undermine credibility, restrict funding and silence contributions through sustained intimidation. These campaigns—frequently gendered and sexualized—create lasting deterrent effects, systematically eroding research capacity and advocacy precisely when most needed.[12]

Research limitations and methodological bias

Research into information risks faces significant constraints. Data access and availability varies widely across commercial digital platforms, pushing researchers to over-rely on accessible or convenient sources while under examining others. This can include dependence on algorithmically personalized individual feeds on digital platforms rather than research-grade datasets. Meanwhile, vast areas of the global information ecosystem remain underresearched, with most related efforts concentrated in English-language contexts.

Researchers may also face limitations in designing and applying appropriate analytic techniques, and may not fully account for potential methodological and cognitive biases. These factors can contribute to assessments that may not fully reflect the complete picture, resulting in skewed interpretation and suboptimal responses.

Perception gaps

Adversarial actors target information spaces monitored by researchers, journalists and decision-makers to create misperceptions around narrative prevalence or public sentiment, at times triggering disproportionate reactions or attention. These actors exploit existing perception gaps in public opinion, amplifying their influence through manufactured consensus.

For example, research from a globally representative survey reveals a significant perception gap regarding climate action. The survey found that **69% of the global population expressed willingness to contribute 1% of their personal income to fight climate change and yet systematically underestimated their fellow citizens' willingness to act by an average of 26 percentage points**.[13]

This "pluralistic ignorance"[14] creates a vulnerability to adversarial tactics like "astroturfing," which has been used to target political leaders and policy makers with artificial grassroots opposition to climate action policies.[15] Media outlets and political leaders may interpret and respond to this manufactured activity as genuine, thereby creating a feedback loop validating the original misperception. This dynamic presents research as well as policy challenges.

Artificial intelligence and emerging technologies

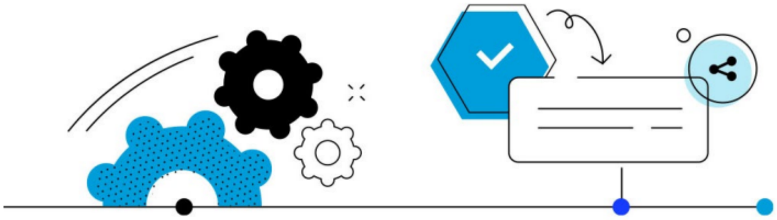
AI technologies are fundamentally transforming how people access information, effectively making societies involuntary participants in a large-scale information experiment with far-reaching consequences.[16]

AI tools are proliferating in the public domain but cannot uniformly be relied on as sources of accurate information. Despite a lack of transparency as to how AI tools work and the sources of information used, people are increasingly using this convenient yet flawed data without requisite AI literacy to assess veracity or reliability. This problematic dynamic is compounded by the misuse of AI in facilitating the creation and dissemination of false and hateful information at scale by a broad range of actors for financial or strategic gain. Collectively, these factors are contributing to the erosion of trust in any information source and in the information ecosystem more broadly.

AI technologies also carry risks for information pluralism and deepening of the digital divide. AI tools are trained on limited data (such as languages, subject matter) and undermine the economic viability of media and other industries, with journalism and other reliable data scraped and summarized without permission or compensation.

Furthermore, there is a lack of transparency regarding the resources required for AI deployment, its environmental impacts, and the specific purposes for which government and private sectors are deploying AI technologies.

With massive investments being made in AI, the benefits for humanity remain unclear. Addressing AI's impact on information integrity requires a systemic approach across multiple domains, including international regulatory frameworks, AI literacy, support for independent media, research into emerging risks, development of technical solutions and other safeguards, and agile communications strategies.



PART III: SOLUTION FRAMEWORKS

THEORY OF CHANGE

Emerging normative frameworks for information integrity are being built on interdisciplinary collaboration and recognize that healthy information environments are essential for human rights, democratic resilience, sustainable development, and peace and security. Information integrity also supports critical sectors including business and finance, scientific advancement, technological innovation, public health, education and the creative industries.

Rapid transformations in the global information ecosystem have intensified risks and vulnerabilities.[17] No single stakeholder can address these challenges alone—effective responses require multi-stakeholder collaboration and action. While some stakeholders such as States and major technology companies hold greater power, resources and responsibilities, others contribute vital perspectives and lived experiences that inform solutions.

Given these complexities, it has become increasingly important to articulate a coherent theory of change that reflects the broad global efforts to strengthen information integrity and can build momentum to guide joint action across stakeholder groups.

Figure 1

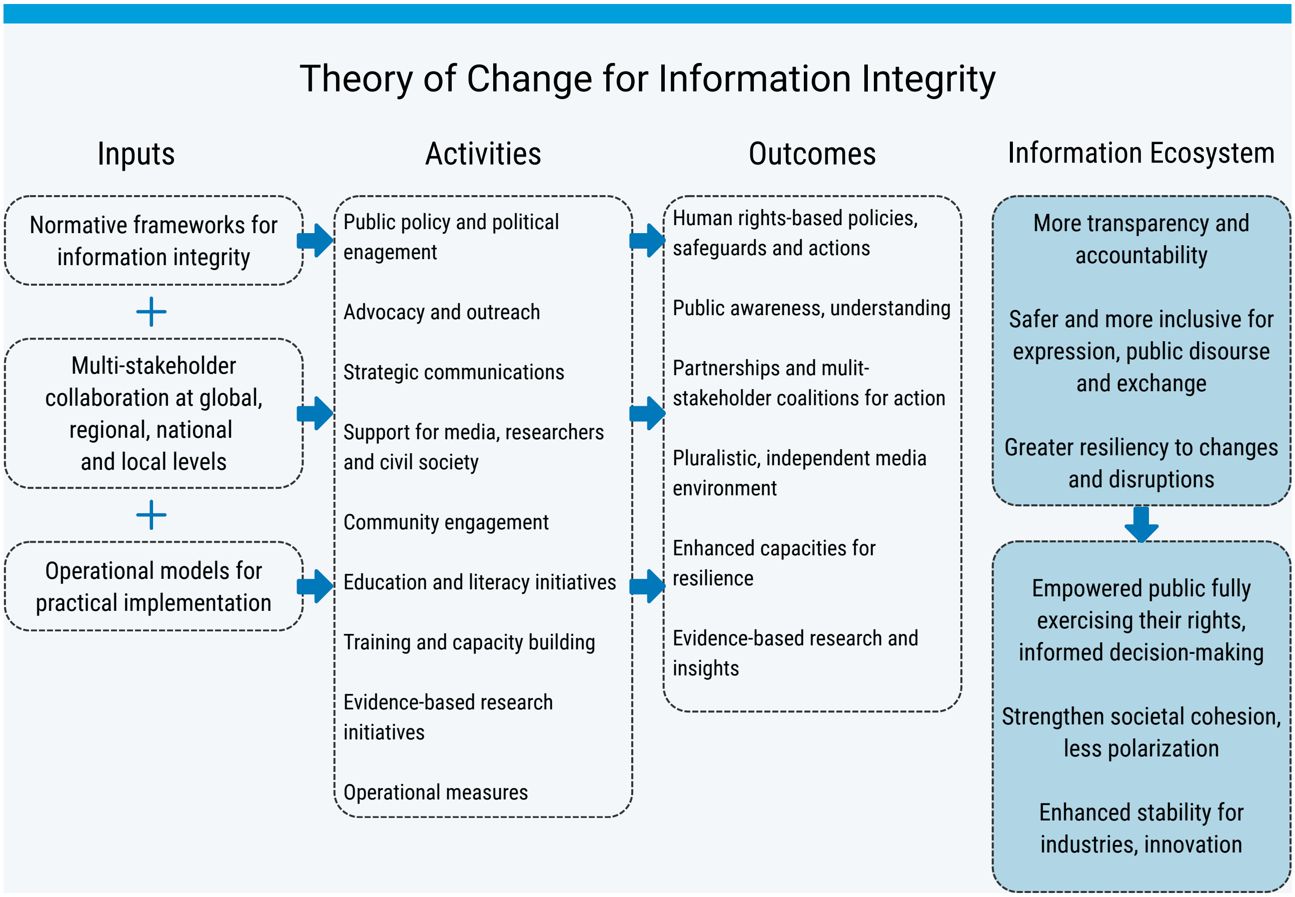

Theory of Change for Information Integrity
Inputs
Activities
Outcomes
Information Ecosystem
Normative frameworks for information integrity
+
Multi-stakeholder collaboration at global, regional, national and local levels
+
Operational models for practical implementation
Public policy and political enagement
Advocacy and outreach
Strategic communications
Support for media, researchers and civil society
Community engagement
Education and literacy initiatives
Training and capacity building
Evidence-based research initiatives
Operational measures
Human rights-based policies, safeguards and actions
Public awareness, understanding
Partnerships and mulit-stakeholder coalitions for action
Pluralistic, independent media environment
Enhanced capacities for resilience
Evidence-based research and insights
More transparency and accountability
Safer and more inclusive for expression, public disourse and exchange
Greater resiliency to changes and disruptions
Empowered public fully exercising their rights, informed decision-making
Strengthen societal cohesion, less polarization
Enhanced stability for industries, innovation

Figure 1 presents a theory of change for strengthening information integrity by which the global information ecosystem can be made more transparent and accountable, more resilient to disruptions and threats, safer and more inclusive.

This is achieved through collaboration between a range of stakeholders—States, private sector companies, research and academic institutions, media, civil society and others. By adopting a shared, rights-based approach grounded in international law, stakeholders can undertake a diversity of activities from community engagement to evidence-based research in order to facilitate the strengthening of the global information ecosystem.

This theory of change seeks to empower individuals and better equip them to make informed decisions and fully exercise their rights. It supports increased societal cohesion and reduced polarization, and contributes to more stable economies for consumers and industries.

MULTI-STAKEHOLDER COALITIONS

Complex challenges to information integrity require diverse expertise and stakeholder participation to find and enable wide-ranging entry points for change. In resource-limited contexts, multi-stakeholder coalitions enable more strategic and targeted efforts.

To succeed, coalitions must be built on shared understanding, trust and values, while aligning on concrete actions:

- Build trust through open communication, shared leadership and mutual accountability.
- Engage governments, media, civil society, academia, private sector and affected communities to broaden reach and impact.
- Collaborate on research to generate actionable insights and scalable tools.
- Maintain agility to respond rapidly to changing threats, especially during crises or high-risk moments.
- Pool funding, capacity building and technical support for efficiency and scale.
- Use campaigns and calls to action to galvanize political will and sustain public attention.
- Address risks that transcend geographic and sectoral boundaries through sustained, adaptable collaboration.

HUMAN RIGHTS-BASED GUARDRAILS

It is vital that measures to strengthen information integrity respect, protect and promote human rights--particularly freedom of expression, in line with international human rights law and with the full participation of civil society.[18] Well-designed, human-rights based guardrails enhance freedom of expression rather than restrict it and protect those who feel unsafe in information spaces, helping to give voice to those who might otherwise be silenced. Guardrails facilitate inclusive access to information, encourage innovation and help foster public trust in fast-emerging AI technologies.

Guardrails can encompass:

- Human rights-based regulatory and legal frameworks.
- Trust and safety around digital platforms.
- Technology standards and accountability.
- Support for independent media and academic research and institutions.
- Meaningful civil society participation, including through multi-stakeholder coalitions.
- Individual empowerment measures, including privacy protections and literacy programmes.

INTERNATIONAL INITIATIVES

A constellation of local and international initiatives have emerged to strengthen information integrity and safeguard fundamental rights in response to mounting information risks. Civil society actors and governments are leveraging international initiatives as advocacy tools and policy inputs to help ensure that regulatory efforts and other actions are rooted in human rights and undertaken in an inclusive manner. A sample of international initiatives are highlighted below.

United Nations Global Principles for Information Integrity: Launched by the Secretary-General in June 2024, the Principles provide a foundational framework for multi-stakeholder action built around five core principles with accompanying calls to action:

Societal trust and resilience involves building resilient communities that can withstand risks to the integrity of the information ecosystem.

Healthy incentives focuses on innovating business models and engaging advertisers to demand transparency on digital advertising processes.

Public empowerment ensures everyone has the tools and digital and media literacy to engage safely and confidently online and gain better control of their personal data.

Independent, free and pluralistic media supports a diverse range of trustworthy media voices, providing accurate and reliable information, free from undue influence or censorship.

Transparency and research promotes openness about how information systems work and supports evidence-based policies.

The Global Digital Compact (GDC): Adopted as part of the Pact for the Future in September 2024, the GDC establishes commitments for UN Member States on information integrity and digital trust and safety, creating new mechanisms to address information risks.[19] The *UN Global Principles* offer a key resource for UN Member States in meeting these commitments.

G20 Leadership: The G20 has emerged as a critical driver of international momentum on information integrity. Under the Presidency of Brazil in 2023-2024, members formally recognized that information integrity is central to economic prosperity and democratic governance through the *Digital Economy Working Group Maceio Ministerial Declaration*.[20] The focus is on information integrity as essential for trust in the digital economy, in public institutions, as well as in governance and democratic processes. This recognition creates powerful incentives for cooperation and coordinated action across developed and developing economies, representing some two-thirds of the world population.

The Global Initiative for Information Integrity on Climate Change

Effective multistakeholder coalition building

The Global Initiative for Information Integrity on Climate Change[21] demonstrates how strategic coalition-building can integrate information integrity into critical global processes. Co-chaired by the Government of Brazil, the UN and UNESCO, with participation from several countries, civil society organizations and academic institutions, the Initiative was launched at the November 2024 G20 Leaders' Summit in Rio de Janeiro.
Three strategic pillars:

- *Strengthening research* into climate information integrity globally, with particular focus on expanding understanding beyond the few countries where much research investment has been concentrated.
- *Supporting evidence-based action and solutions*, including strategic communications, advocacy and investigative journalism.
- *Integrating information integrity into the Conference of the Parties (COP) process* and international climate governance.

Breaking new ground

For the first time ever, information integrity has been included in the Action Agenda of the COP—the top decision-making body of the United Nations Framework Convention on Climate Change—reflecting increasing recognition that urgent and ambitious climate action is not possible without addressing climate disinformation and related tactics of delaying climate action.

Generating momentum

The Initiative has prompted significant engagement. Its Global Fund to support networked, in-depth research and strategic action for climate information integrity received an impressive number of submissions, while the "Mutirão" (Call to Action) launched in July 2025 seeks to identify existing concrete solutions and good practices for climate action. The civil society response has been particularly strong, notably through efforts such as the Climate Information Integrity Summit held in Brazil in March 2025.

Evidence-based approach

The Initiative is gathering evidence from around the world to inform strategic action and develop shared tools and practices. This comprehensive approach recognizes that tactics to undermine climate information operate across borders and contexts and require coordinated international responses.

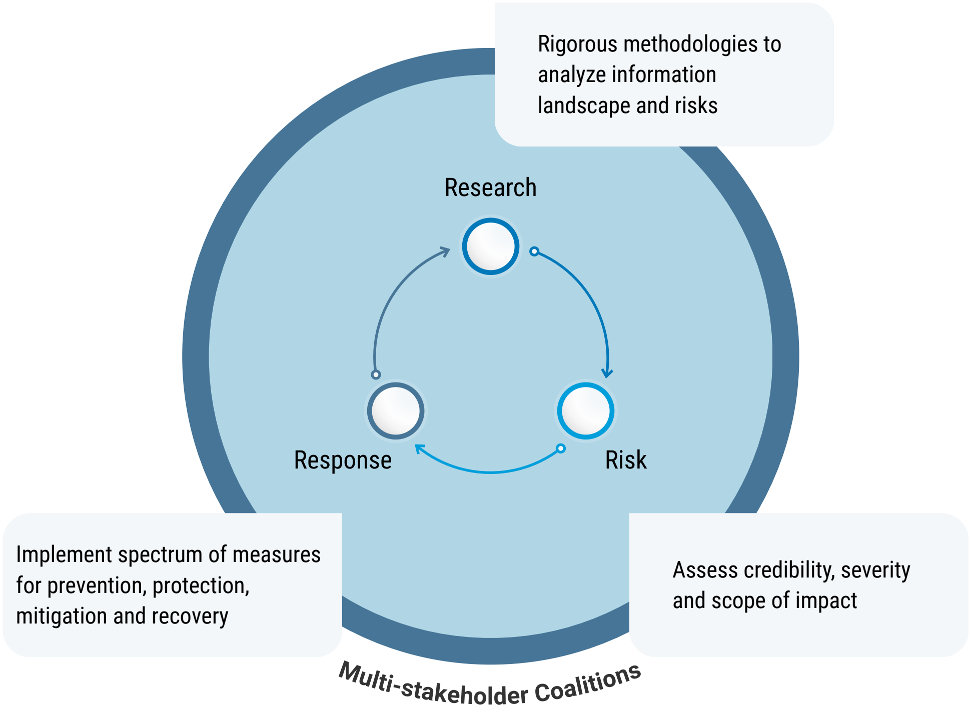
PART IV: OPERATIONAL MODEL

3R APPROACH

While there is no one-size-fits-all solution, State and non-State organizations need viable operational models to strengthen information integrity across a range of contexts and upskill their capacities to meet contemporary challenges. The 3R approach offers a practical starting point centered around **Research, Risk assessment and Response**.

3R assumes a holistic ecosystem view, recognizing information environments as complex and continually evolving, with interlinked risks, vulnerabilities and solutions. 3R aims to develop evidence-based interventions, grounded in research and human rights norms through multi-stakeholder collaboration, engaging across sectors, disciplines and geographies.

Figure 2



Rigorous methodologies to analyze information landscape and risks

Research

Risk

Response

Implement spectrum of measures for prevention, protection, mitigation and recovery

Assess credibility, severity and scope of impact

Multi-stakeholder Coalitions

The success of the 3R approach is facilitated by a number of **enabling conditions**. Organizational leadership commitment to information integrity and evidence-based decision-making are an essential foundation for sustaining 3R work and prioritizing resources. The core functions should be carried out by skilled personnel with expertise in public and internal communications, research (including open-source research), data analysis and policy, supported by access to relevant datasets and technology. The following section provides an overview of the 3R approach, with a more detailed guidance included in the Annex.

Research

Information integrity research improves understanding of the information ecosystem and provides the basis for evidence-based, context-specific interventions. While there are many uncertainties and ambiguities across information landscapes, establishing rigorous methodologies helps to overcome perception and cognitive biases, identify emerging risks, uncover policy and practice gaps, and assess potential solutions.

Resource constrained organizations can prioritize these efforts by leveraging existing expertise, partnerships, publicly available data and open-source research techniques.[22] Research efforts should address critical questions, such as: *What risks are present? Who or what are the drivers of the risks? Which tactics make the risks effective? Which audiences are targeted and what are the impacts?*

Core research activities adapted to fit organizational contexts can utilize the following:

- **Desk review** synthesizes existing reports and studies to establish baseline understanding of information risks and evidence-based measures, which can include work from international organizations, governments, think tanks, media and academic institutions, civil society organizations and others.

- **Situational analysis** (or assessment) maps the information landscape of the operating environment, including political, social, economic, cultural and technological factors. Mapping efforts should aim to identify challenges and needs of the population, stakeholders and their interests, possible and known adversarial actors, as well as areas of resilience and opportunities. It can include information on the regulatory environment, internet use and coverage,[23] media and digital literacy of the population,[24] available public opinion or survey research, degree of press freedom and access to information, and multidisciplinary studies from a range of reliable sources.[25]

- **Campaign analysis** examines influence operations or disinformation campaigns targeting specific individuals, communities, institutions or States, with a focus on adversarial tactics and false narratives and claims. Established frameworks such as ABCDE (Actors, Behaviours, Content, Degree, Effect)[26] and DISARM (Disinformation Analysis and Risk Management),[27] which have been harnessed by a growing number of governments and organizations around the world, provide structured approaches to aid analysis.

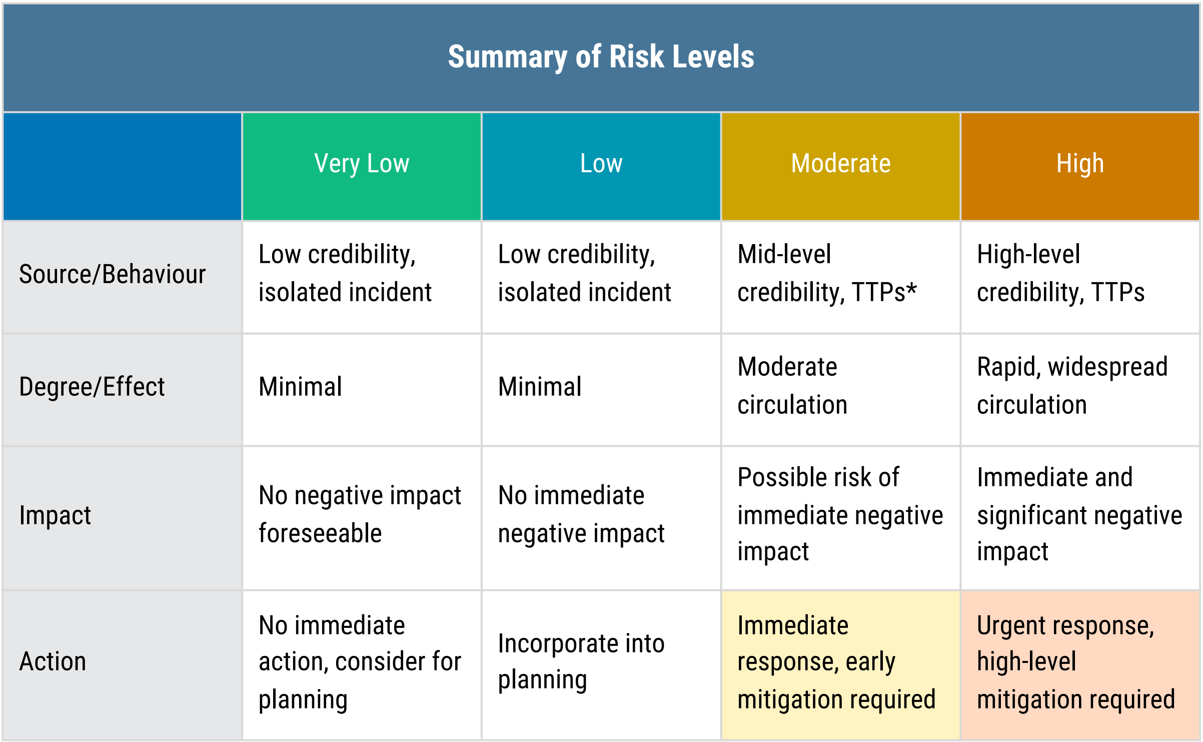
Risk assessment

Assessment of information risks bridges research and action by establishing clear criteria to prioritize responses according to severity, credibility and potential scope of harm. As risks such as disinformation campaigns can spread rapidly and cause immediate harm to targeted populations or organizations, real-time risk assessment is an important tool for effective decision-making.

The risk assessment process evaluates factors such as source authenticity, behavioural patterns, narrative content, distribution levels and impacts on target audiences across broad social, political and human rights domains, as well as within specific operational environments. Standardizing risk classifications from very low, low, moderate and high can help determine the urgency of resource allocation and response, and improve information sharing and coordination with partners.

Figure 3 below provides an example of how risk levels can be applied to possible influence operations or disinformation campaigns for real-time assessment.[28]

Figure 3



Summary of Risk Levels

	Very Low	Low	Moderate	High
Source/Behaviour	Low credibility, isolated incident	Low credibility, isolated incident	Mid-level credibility, TTPs*	High-level credibility, TTPs
Degree/Effect	Minimal	Minimal	Moderate circulation	Rapid, widespread circulation
Impact	No negative impact foreseeable	No immediate negative impact	Possible risk of immediate negative impact	Immediate and significant negative impact
Action	No immediate action, consider for planning	Incorporate into planning	Immediate response, early mitigation required	Urgent response, high-level mitigation required

See Annex for more detailed risk table.

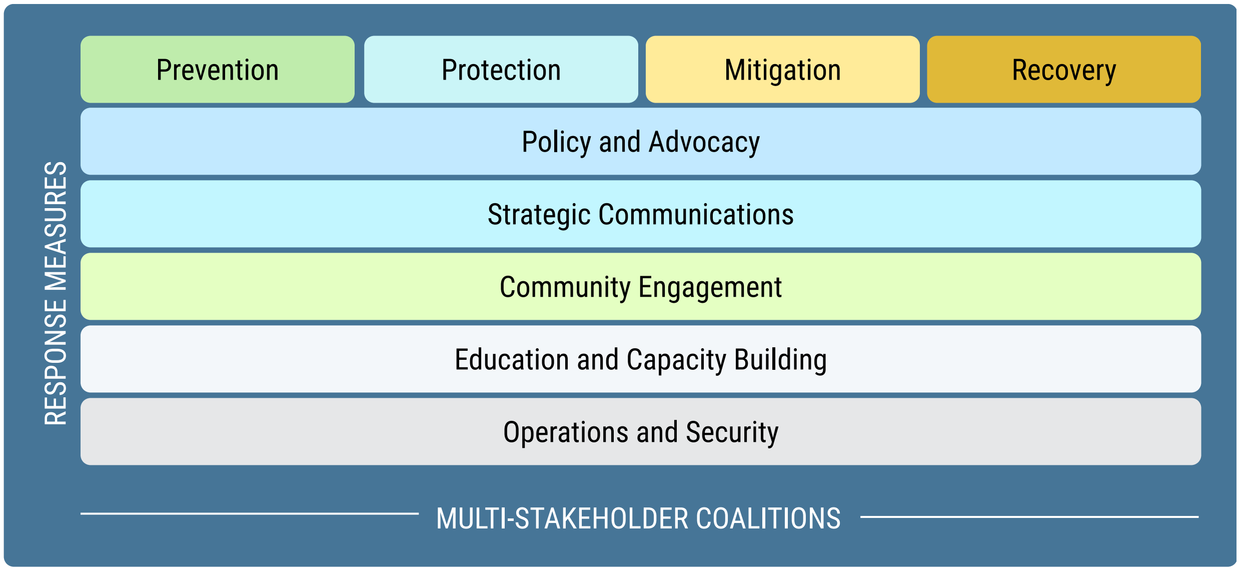
**TTPs refers to tactics, techniques & procedures.*

Response: Prevention, Protection, Mitigation, Recovery

Response covers a spectrum of measures, which can be organized across four core objectives: prevention, protection, mitigation and recovery.

1. **Prevention**: Build long-term resilience to prevent information risks from undermining societal cohesion, human rights, peace and security, and sustainable development.
2. **Protection**: Put in place targeted safeguards in anticipation of high-risk moments and vulnerabilities.
3. **Mitigation**: Contain risk escalation and reduce impacts in real-time.
4. **Recovery**: Restore disrupted capabilities while rebuilding trust and resilience.

Figure 4



See Annex for more detailed response tables.

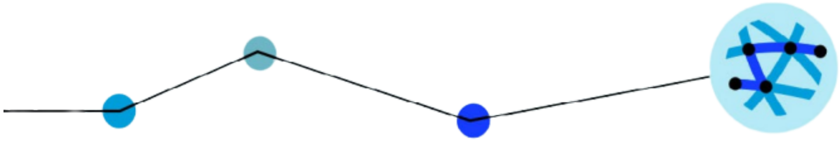
These objectives support planning, coordination and efficient use of resources. Organizations should contextualize response efforts within their specific mandates, capacity constraints and risk environments. All efforts must uphold ethical standards and human rights, particularly freedom of expression. Response actions should in no way themselves undermine information integrity.

A more detailed guide in the Annex offers a non-exhaustive catalogue of possible response measures for each objective, covering: policy and advocacy; strategic communications; community engagement; education and capacity building; and operations and security.

CONCLUSION

Information integrity challenges require urgent, coordinated efforts that move beyond fragmented, reactive approaches towards proactive, systematic capacity-building and resilience. Drawing from a growing body of policy and practice across diverse contexts, this issue brief lays out key concepts and challenges and offers solution frameworks and practical implementation through the 3R model.

The current groundswell of concern and support for a healthier, safer global information ecosystem presents an opportunity for transformative action. **Moving from principles to practice** requires several critical shifts towards a rights-based information ecosystem, diverse coalitions, rigorous research for evidence-based interventions and proactive holistic solutions. This means engaging with and supporting the growing community of information integrity practitioners and scholars through ongoing dialogue and sharing insights and innovations, thereby reinforcing information integrity as an indispensable foundation for the stability, health and progress of people and the planet.



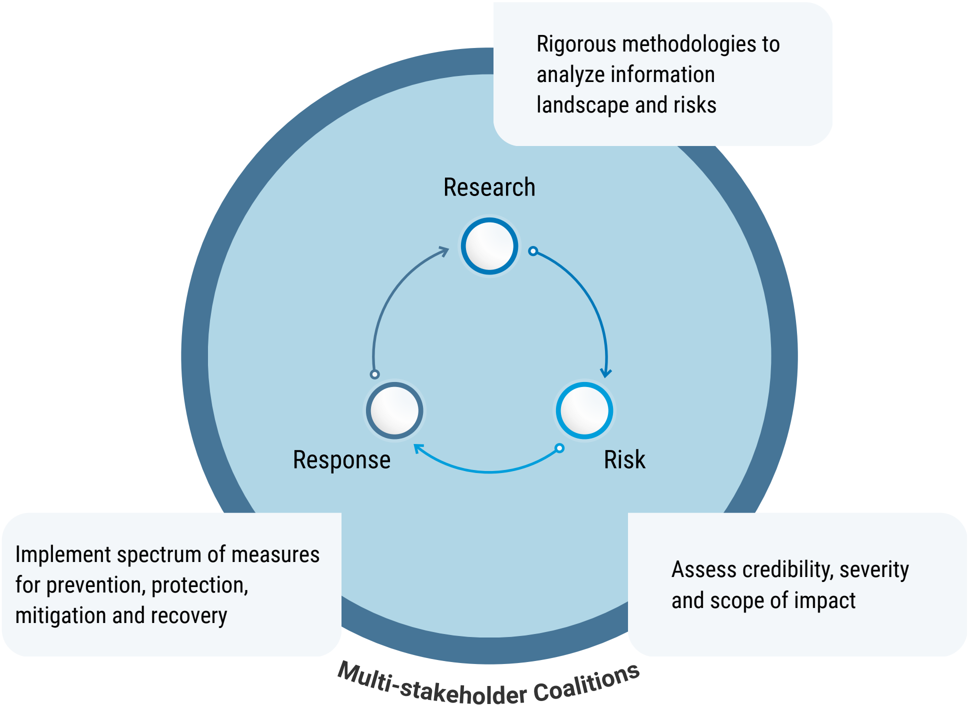
ANNEX: 3R GUIDE

Research, Risk assessment and Response

The 3R Approach—Research, Risk assessment and Response—was developed as an adaptable operational model for strengthening information integrity across organizational contexts and upskilling capacities to address evolving information risks. The following guide provides an overview of key components of the 3R approach to support resource constrained organizations and teams.[29]

Getting started

3R assumes a holistic ecosystem view, recognizing information environments as complex and continually evolving, with interlinked risks, vulnerabilities and solutions. 3R aims to develop evidence-based interventions, grounded in research and human rights norms through multi-stakeholder collaboration, engaging across sectors, disciplines and geographies.



Rigorous methodologies to analyze information landscape and risks

Assess credibility, severity and scope of impact

Implement spectrum of measures for prevention, protection, mitigation and recovery

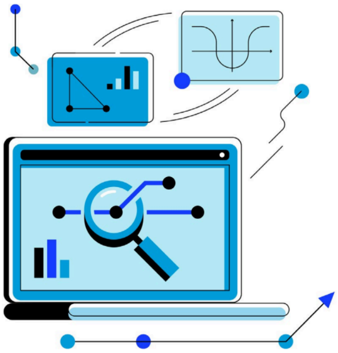
Information integrity research improves understanding of the information ecosystem and enables evidence-based, context-specific interventions. **Risk assessment** bridges research and action by establishing clear criteria to prioritize responses according to severity, credibility and potential scope of harm. **Response** encompasses a spectrum of measures with four key objectives in mind:

1. **Prevention**: Build long-term resilience to prevent information risks from undermining societal cohesion, human rights, peace and security, and sustainable development.
2. **Protection**: Put in place targeted safeguards in anticipation of high-risk moments and vulnerabilities.
3. **Mitigation**: Contain risk escalation and reduce impacts in real-time.
4. **Recovery**: Restore disrupted capabilities while rebuilding trust and resilience.

Research

Information integrity research inherently involves examining opaque and deliberately manipulated information environments, often with limited data and resources. This challenging landscape makes it essential to establish rigorous methodologies that help overcome perception and other cognitive biases to accurately identify information risks, uncover critical gaps and assess potential solutions.

To address methodological and resource challenges, research efforts can leverage existing expertise, partnerships, publicly available data and open-source analytical techniques. Core research activities can include desk review and situational analysis for understanding of the broader context and challenges, while campaign analysis can be used to examine influence operations or disinformation campaigns.[30]



Research efforts should aim to address critical questions to support evidence-based interventions, such as: *What risks are present? Who or what are the drivers of the risks? Which tactics make the risks effective? Which audiences are targeted and what are the impacts?*

Sample of research questions	
Information landscape	What types of information risks are present in the environment?
Actors / Sources	Who or what are the drivers? What is their level of influence?
Behaviours	What makes them effective? What are the tactics and techniques being used? Is this an isolated incident or a pattern of behaviour?
Content / Narratives	What are the top-level false or misleading narratives and claims?
Degree	Which audiences are targeted or affected? Which platforms or information spaces?
Effect	What are the impacts? Are they short or longer term?
Risk and Response	What is the level of risk? *See risk table.* Which evidence-based actions or mitigation measures would address the risks?

When conducting situational analysis or campaign analysis, established frameworks such as ABCDE (Actors Behaviours, Content, Degree, Effect)[31] and DISARM (Disinformation Analysis and Risk Management),[32] provide structured approaches to examining, documenting and developing findings on information risks.

Adversarial actors and behaviours

Practitioners must distinguish between unintentional or low-impact incidents and deliberate, higher-impact campaigns orchestrated by adversarial actors, i.e. individuals, groups or organizations including State and non-State actors, private companies, extremist groups and criminal organizations that seek to undermine information integrity for financial or strategic gain.

Understanding adversarial methods and motivations is essential for developing effective, proportionate response measures, building systemic resilience, and assessing risk and potential harm. These may include:

- **Coordinated inauthentic behaviour**[33] - manipulative communication via fake accounts, bots or paid operatives and other tactics to manipulate or deceive the public.
- **Information environment manipulation** - restriction of access to pluralistic information sources, often targeting a particular group or individual; algorithmic manipulation; creating or exploiting information deserts/voids; shadow banning by technology companies; removing opposing views in public discourse; information pollution (crowding out access to accurate, reliable information with false information).
- **Content manipulation and false narrative strategies** - development and deployment of disinformation and hate speech narratives; falsified or manufactured "evidence"; false authority or manufactured expertise; manufactured credibility; pseudo-science/expertise; information laundering where false information is deliberately cycled through multiple platforms or networks to appear credible.[34]
- **Psychological, cognitive or emotional manipulation** - a broad range of tactics, including fearmongering (manufacturing enemies or threats); manufacturing outrage or moral panic grievance framing; clickbait or ragebait;[35] grooming or preconditioning.[36]

Risk assessment

Information risks, such as disinformation campaigns, can spread rapidly and cause immediate negative impacts on targeted populations or organizations. Real-time risk assessment can provide clear criteria for effective decision-making on appropriate response measures based on:

- Source authenticity, transparency, credibility and influence.
- Source behaviour, current and past.
- False or misleading narratives and other claims.
- Calls to action, such as harassment or violence.
- Distribution levels.
- Effects on target audiences and potential impact areas (e.g. social/political, operating capacities, financial, rule of law, public health, safety and security, and human rights).

The following **risk characteristics** table provides an example of how to determine very low to high levels of risk and actions:

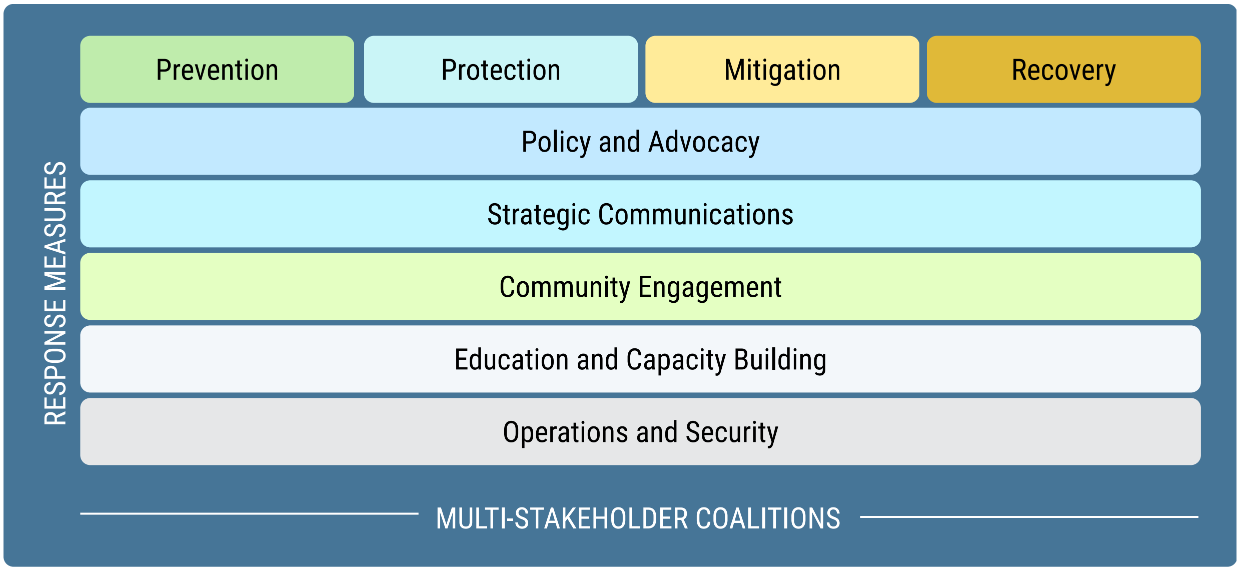
Risk Level	Source/behaviour	Degree/Effect	Impact	Action Required
Very Low	• Low-credibility, low-influence; isolated to specific platform/space • No history or patterns of adversarial behaviour • No clear connections to other adversarial actors • No evidence of coordination or amplification	• No or minimal circulation • Nominal engagement/shares/views/reactions with diminishing trends (<1 day) • No cross-platform dissemination • No significant public or media attention	• No evidence of harm or activity disruption • No call to action	• No immediate intervention required • Routine monitoring sufficient • Consider incorporating insights into response measures for prevention and protection objectives
Low	• Low-credibility, low-influence; isolated to specific platform/space • Possible history or pattern of adversarial behaviour but without broader, influential connections • Minimal coordination or amplification signs • Sporadic, not sustained activity	• Minimal circulation but isolated to small group • Slight upward engagement trend within group (<week) but no broader amplification • Limited cross-platform dissemination within group • May provoke emotional reactions within group • May reinforce existing false narratives or social divisions	• No evidence of harm or activity disruption • No call to action	• Routine monitoring required • Incorporate insights into response measures for prevention and protection objectives, including routine updating of content banks and proactive strategic communications, coordination and information sharing with stakeholders

Continued on next page

Risk Level	Source/behaviour	Degree/Effect	Impact	Action Required
Moderate	• Mid-level influencer(s) or groups/networks • Emerging coordination among actors/groups • Clear malicious intent • Use of recognizable adversarial tactics and techniques, e.g. coordinated inauthentic behaviour, to provoke, manipulate or disrupt • May promote violent rhetoric or harassment (directly or indirectly)	• Moderate circulation attracting attention beyond originating source(s), possibly amplified by influential voices/accounts • Cross-platform dissemination • Increased trend in engagement and exposure • Early signs of public/news media interest • Signs of escalating audience response, e.g. strong emotional reaction, confusion or provocation	• Possible disruption to activities • Direct or indirect calls to action • Increased disinformation scenarios, public confusion • Emerging misrepresentation in news media and relevant information platforms • Stakeholder inquiries, public attention	• Immediate action required • Incorporate insights into response measures for protection and mitigation, including rapid strategic communications and stakeholder engagement, coordinated mitigation with partners • Resource allocation for active mitigation
High	• High-level influencer(s) or groups/networks, possibly with significant resources and decision-making capacities to affect communities/ populations • Coordination across multiple platforms, geographies, languages • Clear malicious intent • Use of recognizable adversarial tactics and techniques, e.g. coordinated inauthentic behaviour, to provoke, manipulate or disrupt • May promote violent rhetoric or harassment (directly or indirectly)	• Rapid spread achieving high attention levels, possibly dominating public discourse • Cross-platform/multilingual dissemination • Mainstream news media involvement • Broad, diverse audience attention with rapid escalation • Strong emotional responses from key audiences • Promotes violent rhetoric, harassment or incitement	• Immediate or potentially significant disruption to activities • Direct or indirect calls to action • Exploitation of volatile or sensitive contexts • Increased disinformation scenarios, public confusion • Misrepresentation in news media and relevant information platforms • Stakeholder inquiries, public attention	• Urgent, coordinated protection and mitigation required • Crisis communication protocols activated • High-level engagement with stakeholders/partners/ media • Additional resources allocated • Additional protection measures may be needed • Consideration of recovery measures may be needed

Response: prevention, protection, mitigation, recovery

Response covers a spectrum of measures across four core objectives: **prevention, protection, mitigation and recovery**. These objectives support planning, coordination and efficient use of resources, recognizing the complexity of overlapping actions across timeframes and functions.



Organizations should contextualize responses within their specific mandates, capacity constraints and risk environments. All efforts must uphold privacy and human rights, particularly freedom of expression. Response actions should in no way themselves undermine information integrity.

A comprehensive mapping exercise can serve as a strategic first step to identify the most appropriate and feasible actions for each objective area. Timeframes can range from immediate (first 48 hours), short term (1 week to 3 months) to longer term (beyond 3 months).

The guidance below offers a non-exhaustive catalogue of possible response measures, covering: policy and advocacy; strategic communications; community engagement; education and capacity building; and operations and security.

Prevent: Building long-term systemic resilience through proactive measures

Success indicators: stronger partnerships, improved preparedness, enhanced organizational and public resilience

Focus Area	Response Measure
Policy & Advocacy	• Long-term multi-stakeholder coalitions with community leaders, civil society organizations, journalists and media, educators, researchers, private sector, advertisers, creators and other public figures • Advocacy for: human rights-based regulatory measures and normative/policy frameworks, emphasizing freedom of expression and access to information; media and digital literacy education at all levels; healthy incentives to address consequences of business models of digital technology; sustainable business models for public interest media
Strategic Communications	• Mapping of audiences and media landscape, utilizing learnings from prior information risk incidents (research) • Rapid message approvals and crisis communications protocols • Content libraries with accurate, jargon-free messaging on key topics • Messaging strategies for anticipated high-risk moments to fill information voids and address data voids • Narratives emphasizing shared values/common ground and transparency and openness, countering adversarial "us vs. them" tactics • Relevant platforms for community-specific strategies, expansion into relevant and niche information spaces tailored to audience interests, even if not traditionally directly related to topic of messaging • Ongoing relationships with journalists and media for information sharing; collaborate on trainings and capacity building • Collaboration with creative communities for sustained positive messaging and opportunities to be reflected in cultural spaces and trends
Community Engagement	• Listening and feedback sessions, taking on board legitimate community concerns • Engagement with local experts, researchers, and trusted community voices to co-develop solutions and support public discourse • Community-driven information sharing and content development • Tailored solutions for improving access to accurate information
Education & Capacity Building	• Media, digital and AI literacy into educational programmes at all levels • Training materials addressing AI technologies, privacy, and online safety issues • Targeted educational initiatives tailored to the specific needs of vulnerable and marginalized groups • Public awareness campaigns on media, digital and AI literacy
Operations & Security	• Conduct periodic organizational vulnerability assessments and preparedness planning • Staff training and capacity building programmes on evolving information environments and risks, including personal digital safety and hygiene and use of AI tools • Establish internal protocols with IT and safety/security personnel

Protect: Targeted safeguard measures

Success indicators: Maintained safety and security of targeted groups, operational capacities (e.g. humanitarian delivery) and progress on specific issues (e.g. climate information integrity)

Focus Area	Response Measure
Policy & Advocacy	• Protective measures for journalists and other targeted groups during high-risk periods • Policies for staff and individual protection and safety due to information risks • Monitoring, documenting and reporting violations of platform policies and community standards • Support for educational and research institutions and experts, especially when under attack
Strategic Communications	• Agile internal protocols for rapid information sharing and coordination • Internal risk assessment to determine appropriate action or non-action, avoiding unforced errors • Quality control and rapid review processes to ensure aligned messaging with established organizational positions • Continued organizational presence in relevant information spaces to address information voids • Engagement of subject matter experts to ensure availability of reliable information in a wide range of information spaces and platforms • Rapid protocols with journalists and media for accurate, accessible information during crisis situations and pivotal societal moments
Community Engagement	• Work with communities to identify vulnerabilities, especially marginalized and other at-risk groups • Protective strategies and protocols with community stakeholders
Education & Capacity Building	• Protective educational campaigns and pre-bunking content with communities for high-risk moments • Training for journalists, media workers and partner organizations
Operations & Security	• Communications protocols for high-risk moments or crises • Digital security measures and conduct periodic red teaming exercises • Safety measures for high-profile and vulnerable personnel • Mental health protocols for staff most exposed to information risks

Mitigate: rapid response and risk reduction

Success indicators: Faster response times, effective messaging, minimized impact

Focus Area	Response Measure
Policy & Advocacy	• Activation of existing partnerships and networks • Assessment of tech platform policy implementation, including during crises • Monitoring and assessment of legal protections and regulatory responses • Support and advocacy for protective measures for journalists and other targeted groups during high-risk periods • Monitoring and reporting of platform policy and community standards violations
Strategic Communications	• Scaled communications capacity to address escalation of risk • Crisis management and staff protection measures • Real-time rapid analysis and risk assessment to inform decision-making and adapt messaging • Internal protocols and pre-planned messaging and content developed as preparedness measure, maintaining messaging coherence and discipline • Speed, high frequency and repetition of accurate, jargon-free information to counter information risks and address voids • Tailored approaches for different platforms and appropriate tone/content for specific communities • Immediate clarifications to address false or misleading claims (directly or indirectly) • Proactive media engagement and real-time briefings to provide journalists with factual information and expert sources
Community Engagement	• Outreach to affected communities to address inquiries, concerns and context-specific needs • Accurate information through local channels and community networks established as a preparedness measure
Education & Capacity Building	• Support for affected youth, educators and institutions; include in stakeholder engagement • Pre-prepared information and resources provided for immediate use in educational settings • Coordination with schools, universities and educational organizations on crisis communications
Operations & Security	• Staff protection during escalation of threats • Enhanced security and other technical measures, such as cybersecurity, physical security

Recover and restore impaired or disrupted capabilities while rebuilding trust and resilience and conducting a systematic review

Success indicators: Organizational learning, strengthened systems, renewal of capacity and trust

Focus Area	Response Measure
Policy & Advocacy	• Review of implementation and efficacy or regulatory responses and platform policies and develop corresponding recommendations • Develop measures to strengthen regulatory and normative frameworks and organizational resilience • Reassess and strengthen coalition and advocacy partnerships • Support affected staff and partner groups • Advocate for improved trust and safety measures of digital platforms and other affected information spaces
Strategic Communications	• Updated communication strategies based on evaluation and learnings • Improvements in messaging and correcting false information (debunking) • Transparency, including acknowledgment of any errors or gaps, to rebuild credibility • Renewal of media and partnership engagement, supporting pluralism in media ecosystem, especially in settings with limited media diversity and infrastructure • Continued expansion of presence across relevant information spaces
Community Engagement	• Rebuilding of community trust through sustained, transparent engagement and collaborate on trust-building and trauma-informed support • Inclusion of local leaders and communities in stakeholder and partnership reviews to strengthen resilience; documenting learnings from community-level interventions, public dialogues
Education & Capacity Building	• Evaluation learnings reflected in updates for educational materials and curricula • Implementation of new and revised training and capacity building
Operations & Security	• Provide necessary resources for affected staff and individuals • Systematic reflection exercises and operational assessments • Strengthened organizational systems based on incident experience • Updated protocols and procedures for improved future response

Considerations for effective implementation:

- Uphold human rights and adhere to ethical and professional standards.
- Collaborate across teams and with a range of stakeholders.
- Be transparent. Maintain open communication, information sharing.
- Keep organized records of research, analysis and reporting of findings.
- Ground actions in evidence-based information.
- Prioritize affected community needs and organizational mandates.
- Adapt strategies based on context and feedback.
- Conduct regular review and case studies for continuous learning and improvement.

ENDNOTES

1.This issue brief series uses different terms to describe the global information ecosystem. Information ecosystem is used to describe the complex, interconnected global system where actors, technologies and information inter-relate. Information environments or landscapes are used interchangeably, focusing more on contextualized functional or operational settings. Information space is used for more detailed, and more technically focused contexts, such as specific digital platforms or spaces dedicated to specific communities.
2.See United Nations Global Principles for Information Integrity for a more detailed exploration of the normative landscape.
3.The Pact for the Future and its annexes—the Global Digital Compact and the Declaration on Future Generations—were adopted at the Summit of the Future on 22 September 2024.
4.See UN Global Risk Report (2024).
5.See United Nations Global Principles for Information Integrity (2024).
6.See UN Disaster Risk Reduction briefing note for more on systemic risks (2022). A key attribute of systemic risk is that it can transgress spatial and sectoral boundaries with other systems, sectors and geographical regions, thus leading to cascading effects.
7.The accompanying Guide in the Annex provides more information on adversarial behaviour.
8.For more information on psychological aspects of information manipulation, see First Draft (2020), The Psychology of Misinformation: Why We're Vulnerable.
9.The term information void is used to refer to a situation in which accurate, reliable information is not readily available in a specific information space or context. Data void refers to a situation in which clear evidence is not yet available, for instance due to a lack of advanced research on an issue.
10.Reuters Institute (0 Feb 2025), Watching chaos through a screen: How social media is changing the way we follow extreme weather events.
11.See International Panel on the Information Environment (2025), Information Integrity about Climate Science: A Systematic Review; Climate Action Against Disinformation (2024), Extreme Weather, Extreme Content: How Big Tech Enables Climate Disinformation in a World on the Brink.
12.See UNESCO (2024), Press and planet in danger: safety of environmental journalists; trends, challenges and recommendations; UNESCO (202), The Chilling: global trends in online violence against women journalists.
13.See Andre, P., Boneva, T., Chopra, F. et al. Globally representative evidence on the actual and perceived support for climate action. Nat. Clim. Chang. 4, 253–259 (2024). https://doi.org/0.038/s4558-024-0925-3.
14.See United Nations (2024), Governing AI for Humanity: Final Report.
15.See Climate Action Against Disinformation (September 2023), Climate Mis-/Disinformation Backgrounder.
16.See United Nations (2024), Governing AI for Humanity: Final Report.
17.See UN Global Risk Report (2024).
18.See UN Global Principles and Article 9, International Covenant on Civil and Political Rights.
19.See Global Digital Compact (2024).
20.For more information, see Digital Economy Working Group Maceio Ministerial Declaration (2024).
21.For more information, see websites Global Initiative for Information Integrity on Climate Change and UNESCO: Global Initiative for Information Integrity on Climate Change.
22.For more information on a model to understand the information ecosystem, see Internews Information Ecosystem model (204).
23.Open-source research techniques refers to research or investigative methods used to examine publicly available information, which can be collected, retained and stored without special authorization. While using open-source techniques, ethical, privacy and safety considerations must be made.
24.See ITU for more information on internet use statistics (2024).
25.See UNESCO for more information on data and statistics on freedom of expression, media and digital literacy and other communications and information areas.
26.See Pammant, J. (2020). The EU's role in fighting disinformation: Crafting a disinformation framework. Carnegie Endowment for International Peace.
27.See DISARM Foundation for more information on the DISARM framework used for documenting influence operations.
28.See Guide in Annex for a more detailed version of this risk characteristics table.
29.This guide will not address specific resource requirements as they will differ depending on the organizational context.

30.A brief description of these research activities can be found in Issue Brief.
31.See Pammant, J. (2020). The EU's role in fighting disinformation: Crafting a disinformation framework. Carnegie Endowment for International Peace.
32.See DISARM Foundation for more information on the DISARM framework used for documenting influence operations.
33.See EU Disinfo Lab for more information on coordinated inauthentic behaviour.
34.See Center for International Media Assistance (209), Information Laundering and Globalized Media — Part I: The Problem.
35.See rage-bait definition, Merriam-Webster.com dictionary.
36. See Union of Concerned Scientists (8 May 208), The Disinformation Playbook: How Business Interests Deceive, Misinform, and Buy Influence at the Expense of Public Health and Safety.