



World Disasters Report

2026

A photograph of an IFRC worker in a red jacket with a white patch on the back, standing in a camp of tents. The worker is seen from behind, and the patch on the jacket features a red cross and a red crescent symbol above the letters 'IFRC'. In the background, a young girl is standing in a gravelly area, looking down. The overall scene is a disaster relief camp.

**Truth, Trust and
Humanitarian
Action in the
Age of Harmful
Information**

The International Federation of Red Cross and Red Crescent Societies (IFRC) is the world's largest humanitarian network, with 191 National Red Cross and Red Crescent Societies and nearly 17 million volunteers. Our volunteers are present in communities before, during and after a crisis or disaster. We work in the most hard-to-reach and complex settings in the world, saving lives and promoting human dignity. We support communities to become stronger and more resilient places where people can live safe and healthy lives, and have opportunities to thrive.

Cover photo

Palestine Red Crescent Society volunteers go door to door in Bourj el-Barajneh camp, Lebanon, to listen to concerns, share practical information and guide families to UNRWA clinics, helping children catch up on missed vaccinations. This people-centred approach reduces barriers to access while upholding dignity, trust and continuity of care in a challenging setting. September 2025. Lama Chidiac, IFRC.

Inside photos

Egyptian Red Crescent • James Tamba-Martha, Sierra Leone Red Cross Society • Lama Chidiac, IFRC • Aki Kolehmainen, Finnish Red Cross • Lama Chidiac, IFRC • Syrian Arab Red Crescent • Lama Chidiac, IFRC • IFRC • Syrian Red Crescent • Lama Chidiac, IFRC • IFRC • Paul Wu, IFRC • IFRC and Ecuadorean Red Cross • Pakistan Red Crescent • Turkish Red Crescent • Aki Kolehmainen, Finnish Red Cross • IFRC and Ecuadorean Red Cross • Aki Kolehmainen, Finnish Red Cross • Turkish Red Crescent • Aki Kolehmainen, Finnish Red Cross • Lama Chidiac, IFRC • The President of the IFRC at the Red Cross and Red Crescent Statutory Meetings, Geneva, Switzerland, 2024, IFRC • Aki Kolehmainen, Finnish Red Cross • Italian Red Cross • Pakistan Red Crescent • Aki Kolehmainen, Finnish Red Cross • Syrian Arab Red Crescent • Paul Wu, IFRC • Finnish Red Cross • Turkish Red Crescent • Aki Kolehmainen, Finnish Red Cross • James Tamba-Martha, Sierra Leone Red Cross Society • Aki Kolehmainen, Finnish Red Cross • Aki Kolehmainen, Finnish Red Cross

© International Federation of Red Cross and Red Crescent Societies, Geneva, 2026

Any part of this publication may be cited, copied, translated into other languages or adapted to meet local needs without prior permission from the International Federation of Red Cross and Red Crescent Societies, provided that the source is clearly stated.

Contact us

Requests for commercial reproduction should be directed to the IFRC secretariat:

Address

Chemin des Crêts 17, Petit-Saconnex, 1209 Geneva, Switzerland

Postal address

P.O. Box 303, 1211 Geneva 19, Switzerland

T: +41 (0) 22 730 42 22 | F: +41 (0) 22 730 42 00

E: secretariat@ifrc.org | W: ifrc.org

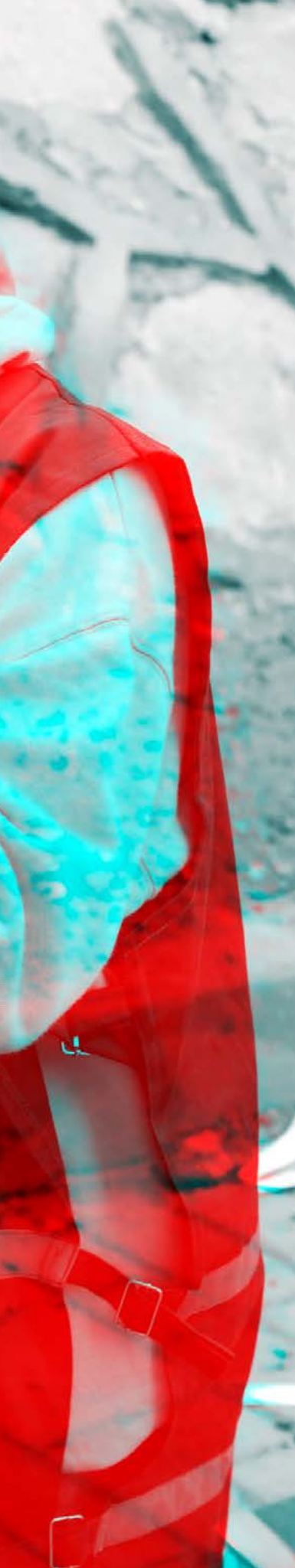
World Disasters Report

2026

**Truth, Trust and
Humanitarian
Action in the
Age of Harmful
Information**



Contents



Acknowledgements	9
Foreword	11
Acronyms	15
Introduction	19
Endnotes	24

Chapter 1	Crisis, chaos and confusion: Understanding harmful information in humanitarian contexts	29
	Introduction	31
	1.1 Defining harmful information	34
	1.2 The evolving information ecosystem	34
	1.3 Who are the threat actors?	37
	1.4 Information integrity in crisis situations	38
	1.5 From broadcast to two-way engagement	41
	1.6 Who is most vulnerable to harmful information – and why?	44
	1.7 What is the impact of harmful information?	46
	1.8 Artificial intelligence and harmful information	50
	1.9 Narratives shape perception, trust and action	53
	1.10 Controlling accessibility and content moderation	55
	1.11 Analysis and monitoring of harmful information	56
	Concluding remarks: Harmful information is not background noise	60
	Asks, aims and recommendations	61
	Endnotes	63

Chapter 2	Harmful information and the erosion of trust in humanitarian response: The role of truth, trust and technology	69
	Introduction: The shifting ground of trust	71
	2.1 Incentivized to hostility	71
	2.2 Trust: What is being lost?	72
	2.3 Trust in institutions	75
	2.4 Integrity, perception and the fragile foundation of trust	80
	2.5 Decline of expertise and its impact on trust	83
	2.6 Community engagement: A bridge to trust	85
	2.7 Facts and feelings: A perception challenge	89
	2.8 Responding to harmful information: Building trust in crises	91
	2.9 Transparency and identifiability: Foundations for trust	92
	Concluding remarks: Trust, truth and preparedness	94
	Asks, aims and recommendations	96
	Endnotes	98

Chapter 3	Global and local: Dynamics of harmful information in a connected world	103
	Introduction: Harmful narratives that thrive	105
3.1	Local and global interplay	105
3.2	When words harm: The International Red Cross and Red Crescent Movement's call to action	110
3.3	The evolving nexus between cyberattacks and harmful information	111
3.4	Addressing harmful information, strengthening early warning and response	115
3.5	Risk communication and community engagement	123
	Concluding remarks: Navigating humanitarian action in a post-trust information era	127
	Asks, aims and recommendations	129
	Endnotes	131
Chapter 4	From context to consequence: Humanitarian sector voices on the impact of harmful information	137
	Introduction: Harmful information and operational realities	139
4.1	Narratives of harm	142
4.2	Harms and impacts of harmful information on humanitarian action	143
4.3	<i>Do no harm</i> in a harmful information age	150
4.4	Vulnerabilities and amplifiers	152
4.5	From fragmentation to focus: Avoiding overwhelm and building humanitarian capacity	153
4.6	When viruses go viral: Building resilience against harmful information in a pandemic	157
	Concluding remarks: Navigating a hostile information landscape	158
	Asks, aims and recommendations	160
	Endnotes	162
Chapter 5	Navigating regulation, rights and societal resilience	167
	Introduction: Information landscape and humanitarian contexts	169
5.1	Defining harmful information: A strategic and contextual challenge	170
5.2	The risks of information control in emergencies	171
5.3	Digital ceasefires and harmful information	173
5.4	Sovereignty in cyberspace	175
5.5	Media as a pillar of societal resilience	175
5.6	A threat to humanitarian action and to humanity itself	186
5.7	Red Cross and Red Crescent Appeal to States	189
5.8	UN action on AI and information integrity	194
5.9	Content moderation and the power of platforms	194
5.10	From self-regulation to state oversight: The evolving governance of online content	196
5.11	Framing a response: Supply and demand solutions to disinformation	197
5.12	Civic trust and societal resilience	200
	Concluding remarks: A collective responsibility for preserving principled humanitarian action	202
	Asks, aims and recommendations	203
	Endnotes	205

Chapter 6	Rooted in resilience: Community-first approach to harmful information	211
	Introduction: The importance of community for resilience	213
	6.1 Determining community	213
	6.2 Community engagement and accountability	214
	6.3 What communities say: Primary research insights	225
	6.4 Community-led solutions	228
	6.5 Unpacking risk: A gender and diversity lens	232
	6.6 Building sustainable information resilience	234
	6.7 The community within: How volunteers build trust and humanitarian reach	238
	6.8 Effective communication in the context of harmful information	242
	6.9 Critical reflections: Challenges in practice	247
	Concluding remarks: From communication to collaboration	248
	Asks, aims and recommendations	250
	Endnotes	252
Chapter 7	Upholding humanitarian principles in the age of echo chambers	257
	Introduction: Outpaced – humanitarian action in the era of instant narratives	259
	7.1 Reaffirming and applying humanity in an age of distrust	261
	7.2 Dehumanization in the digital age	262
	7.3 Promoting tolerance in diverse and divided societies	264
	7.4 Neutrality at a cost: The price of not taking sides	267
	7.5 Patriotism and humanitarian principles	273
	7.6 Independence as a condition for trust	274
	7.7 Impartiality as a compass	275
	7.8 Organizational integrity: Internal alignment with principles	276
	7.9 A framework for applying humanitarian principles in the information age	282
	7.10 Influencers: Connectors and dividers	290
	7.11 Reaffirming humanitarian principles in the age of digital tools	292
	Concluding remarks: Principled action requires more than declarations	293
	Asks, aims and recommendations	295
	Endnotes	297
Chapter 8	Truth, trust and resilience in the humanitarian sector	301
	Introduction: The high stakes of the information crisis	303
	8.1 Why connection matters in humanitarian response	307
	8.2 What's ahead? Evolutions and known unknowns	309
	8.3 Defining humanitarian resilience in the information age	313
	8.4 Trust brokers – the human bridge to credibility	330
	8.5 Humanitarian principles as a compass	332
	8.6 Recommendations for resilience	339
	8.7 Prerequisites and cross-cutting enablers	345
	Conclusion: Together, we can uphold and reclaim space for humanity	348
	Endnotes	350

Annex I	353
Annex II	359
Concluding remarks: Lessons and looking ahead	369
Asks, aims and recommendations	371
Endnotes	372





Acknowledgements

The International Federation of Red Cross and Red Crescent Societies (IFRC) gratefully acknowledges the support provided by the following organizations for the production and publication of the IFRC *World Disasters Report 2026*:



 Cruz Roja Española



 Global Disaster Preparedness Center

 Australian Red Cross

 ÖSTERREICHISCHES ROTES KREUZ

 Canadian Red Cross

 Röda Korset

 Deutsches Rotes Kreuz

 日本赤十字社
Japanese Red Cross Society

 Finnish Red Cross

 Røde Kors

 Croce-Rossa Svizzera
Schweizerisches Rotes Kreuz
Croce Rossa Svizzera



Australian Government
Department of Foreign Affairs and Trade

Translation costs for the Russian and Chinese editions were generously provided by the Academy of the Russian Red Cross and the International Academy of Red Cross in China.

Lead Editor and Author: **Charlotte Lindsey Curtet**

Designer: **Yann le Floc'h**

Project Manager: **Heather Marie Leson**

Copyeditor: **Jen Claydon**

The *World Disasters Report 2026* features nearly 100 written contributions from more than 60 practitioners and researchers across the humanitarian sector, as well as from governments, academia, civil society and beyond. These contributions appear throughout the report in Contributor Insight boxes, acknowledging the individual authors and their respective organizations.

The **IFRC Solferino Academy** led a two-month rapid community intelligence study with 40 volunteers and staff from 10 National Societies. Acting as community researchers, they interviewed 132 volunteers and community members affected by humanitarian crises. Findings were synthesized by volunteer academics from the University of Michigan, Northumbria University and Open Lab (Newcastle University), ensuring that local nuance was preserved. Selected quotes appear throughout the report.



Foreword

By Jagan Chapagain

Secretary General, International
Federation of Red Cross and
Red Crescent Societies



In every crisis I have witnessed – and in every response by the International Red Cross and Red Crescent network to a disaster, public health emergency, mass population movement or the humanitarian consequences of armed conflict – information is as essential as food, water and shelter. It guides people to safety, connects them to loved ones and gives them the knowledge to safeguard themselves and their communities.

But information can also cause harm. When false, misleading or deliberately manipulated, it can deepen fear, fuel discrimination, obstruct humanitarian access and cost lives. We have seen this too often: during disease outbreaks, when rumours outpace health advice; after disasters, when mistrust hampers aid delivery; and in armed conflicts, when inflammatory narratives escalate violence.

Harmful information is not new, but today it moves with unprecedented speed and reach. Digital platforms open vital channels for community voices, yet also provide fertile ground for lies. In this environment, trust is fragile.

In humanitarian contexts, harmful information flourishes where trust is weak. Trust strongly influences whether people believe harmful information, share it and spread it – shaping its impact and reach. Harmful information thrives on fear and uncertainty, weakening the relationships that make humanitarian access and action possible. Without trust, people are less likely to prepare, seek help or follow life-saving guidance; with it, communities act together, absorb shocks and recover more effectively. Maintaining trust is not optional – it is a humanitarian necessity.

The *World Disasters Report 2026* calls on governments, humanitarian actors, media, technology companies and communities, to recognise that the trustworthiness of information is a matter of life and death. Just as we plan for logistics, shelter and health care in emergencies, we must also plan for the information environment. This requires investing in community engagement, prioritizing listening over speaking, building resilience against harmful narratives and consistently upholding humanity, neutrality, impartiality and independence in every interaction and message.

In 2024, 32 Red Cross and Red Crescent volunteers and staff were killed while on duty; in 2025, a further 27 lives were lost. This is a painful reminder of the risks borne by those who put humanity first, often in the most dangerous circumstances. Their deaths underscore the urgent need to protect humanitarian staff and volunteers, and to safeguard the space they need to save lives. Anything that deepens mistrust, fuels hostility or turns the emblem of protection into a target must be confronted with urgency. Honouring their memory compels us to meet this crisis with urgency and resolve.

Hope must also be part of our response. Harmful information thrives on fear and confusion, but hope – rooted in trust, solidarity and human dignity – can be just as contagious. Balancing threats with hope and fostering narratives of agency and possibility offers a vital counterforce. Around the world, communities respond to crises not only with resilience but also with creativity and compassion, often leading the way in finding solutions. By amplifying authentic stories and voices, we do more than counter falsehoods: we inspire action, strengthen trust and remind people that even in the darkest moments, there are paths forward.

The IFRC and its member Red Cross and Red Crescent Societies have always worked at the heart of communities, earning trust through presence, transparency and respect. In today's information crisis, this role matters more than ever. Our commitment is clear and unwavering: act with humanity and based on evidence, uphold trust and put people

in need first – so that in moments of fear and uncertainty, the voices that carry furthest are those that heal, aid and protect. Each of us has a role to play. By staying informed, questioning, verifying and amplifying only what is trustworthy and constructive, we strengthen the information environment and ensure that hope speaks louder than harm.





Acronyms

AI	Artificial intelligence
CAR	Central African Republic
CBS	Community-based surveillance
CDAC	Communicating with Disaster Affected Communities
CEA	Community engagement and accountability
CREC	Centre for Research on the Epidemiology of Disasters
CSO	Civil society organization
DRC	Democratic Republic of the Congo
DREF	Disaster Response Emergency Fund
DRM	Disaster risk management
ICRC	International Committee of the Red Cross
ICT	Information and communication technology
IDMC	Internal Displacement Monitoring Centre
IFRC	International Federation of Red Cross and Red Crescent Societies
ITU	International Telecommunication Union
MDH	Misinformation, disinformation and hate speech
MHPSS	Mental health and psychosocial support
NGO	Non-governmental organization
OCHA	Office for the Coordination of Humanitarian Affairs (UN)
OECD	Organisation for Economic Co-operation and Development
Q&A	Questions and answers
RCCE	Risk Communication and Community Engagement
SDG	Sustainable Development Goal
UNDP	UN Development Programme
UNHCR	UN High Commissioner for Refugees
WHO	World Health Organization



Introduction

**At crisis point:
Countering harmful
information,
defending humanity**





Introduction



**At crisis point:
Countering harmful
information,
defending humanity**

The *World Disasters Report* is the flagship publication of the International Federation of Red Cross and Red Crescent Societies (IFRC), designed to drive policy change, shape thinking and strengthen practice across the humanitarian sector. This 2026 edition focuses on **harmful information** in humanitarian contexts.

The IFRC defines harmful information as “information that has the potential to cause, contribute to, or result in harm to an individual or entity”. The term focuses attention on the harm itself, rather than on classifying the type of information being spread, which is often difficult to discern and constantly evolving. Harmful information includes misinformation, disinformation, malinformation, hate speech and other damaging narratives (see [Annex I: Glossary, on page 353](#)).

Harmful information is a pressing and pervasive challenge across all sectors of society, and a critical issue for organizations operating in humanitarian crises. Its impact on people in need, communities, responders, institutions and public trust in humanitarian action is profound.

What makes information harmful? Is its impact escalating, and why? Why does it matter so critically in humanitarian contexts? What can be done about it? These are the central questions the *2026 World Disasters Report* addresses, offering key insights, practical guidance and recommendations to help the sector navigate this challenge with greater clarity, resilience and accountability.

Twenty years ago, the *2005 World Disasters Report* introduced a new paradigm: **information as a life-saving resource**. Influenced by the devastating loss of life in the 2004 Indian Ocean tsunami, it exposed deep inequities in access to information and called for more equitable, meaningful exchange with communities. While it made minor reference to the issues of rumours and myths affecting humanitarian response, its central message marked a profound shift – **information was no longer only a support to aid, but aid itself**. Timely, accurate information was reframed as essential to survival, safety and dignity: a critical resource and a moral imperative, especially for communities with the fewest resources, for whom it may be the only form of disaster preparedness available.¹

In 2005, information was recognized as a form of disaster response. The concern then was one of omission – what people did not know – rather than the deliberate spread of damaging falsehoods. Twenty years on, that has changed. In 2026, this report frames harmful information as a de facto *crisis*: one that undermines access of populations to humanitarian aid, erodes trust, misleads communities, deepens vulnerabilities and destabilizes social cohesion. It is a crisis because it is chronic and evolving – escalating, mutating and persisting across time and contexts. Harmful information does not simply ‘strike’ once; it spreads, multiplies and compounds harm in many forms. It directly undermines the humanitarian sector’s ability to fulfil its core role: to alleviate suffering and protect life and human dignity, especially for people in the most vulnerable situations, while promoting respect for international laws and upholding humanitarian principles and standards. Because harmful information can obstruct access, distort needs, fuel mistrust and even incite violence, it is no longer a peripheral concern. Responding to it is not optional – it is integral to protecting people and upholding humanitarian purpose.

Managing harmful information is no longer just a communication challenge; it is an operational and ethical imperative that demands a whole-of-society response, encompassing public policy, institutional policies, preparedness, education and social awareness. Meeting this challenge requires a shift in how the humanitarian sector prepares for, mitigates and actively counters harmful information. The sector is neither

fully equipped to respond nor solely responsible, making collective response essential. Some 60 organizations provided written contributions to this report, underscoring how seriously its impacts are now regarded across the sector and beyond. Across consultations, all recognized the growing danger of harmful information and many voiced hopes that this report can help build understanding and drive action.

Historically, humanitarian standards and policies have often emerged in response to operational failures, access constraints or new technologies. Today, the challenge is no longer about the availability or accessibility of information, but about its reliability. This calls for clear sector-wide standards and guidance on how humanitarian actors should prepare for and respond to harmful information that are integrated into preparedness, risk assessment, protection and accountability frameworks and always adapted to local realities.

A focus on technology and the future of humanitarian intervention was the theme of the *2013 World Disasters Report*. It observed: “Humanitarians always lament the lack of information when they face hard decisions yet they take pride in acting during the ‘fog of war’. The advent of digital information and communication systems may reduce that fog but make the decisions even more consequential.”² What once inspired optimism – that technology could generate new information, provide early warning and place affected communities at the centre of humanitarian action as engaged participants – has increasingly been tempered by recognition of its limitations and risks. The sheer volume of witnesses, bystanders and malicious actors using these same tools to deliberately undermine humanitarian response and erode trust has, in many cases, not lifted the ‘fog of war’ but created a smokescreen or even a deeper darkness – one that shifts power, deepens vulnerabilities and undermines resilience.

Information as a basic need must disrupt this dynamic. What was acknowledged two decades ago must now be reaffirmed and actively put into practice.

Today, as the humanitarian sector confronts a new wave of technological change – artificial intelligence (AI) – the stakes are rising once again. AI is accelerating the production and spread of information at unprecedented speed and scale, lowering the barriers to entry for malicious actors to manipulate content and influence opinion. Governance frameworks are struggling to keep pace.

Understanding how both information and technology are evolving is now essential to shaping effective responses to harmful information in humanitarian contexts. What has changed in the two decades since information was first recognized as a life-saving resource? What can we now say about the impact of harmful information on humanitarian response – and, more importantly, on the people the response is meant to serve?

The scale and speed of harmful information today far exceed anything previously experienced. Navigating this flood of misleading, instrumentalized and targeted content has become critical to maintaining trust, protecting people’s safety and ensuring humanitarian response reaches people in need.

The report draws on lived experience and offers practical recommendations for preparedness, response and addressing harmful information in humanitarian contexts. Covering a wide range of crises – health emergencies, disasters, migration, climate shocks, armed conflict – it identifies key risks and provides guidance for strengthening resilience, fostering engagement and building trust. It includes community intelligence research contributed by 40 volunteers from 10 National Red Cross and Red Crescent

Societies, supported by the IFRC and academic partners. The report provides asks, aims and recommendations grounded in actual crisis response; these are practical, shaped by lessons learned and illustrated with real-world responses. It presents an analysis and typology of harm (see [Chapter 1, section 1.7, on page 46](#)), illustrating how harmful information impacts humanitarian action – and indicates the type of harm associated with each ‘Contributor Insight’ box. This underscores the need for a stronger evidence base to guide strategies for measuring and addressing the harms and impacts of harmful information to inform responses and policy.

In this landscape of harmful information, the report places **trust, proximity, community and resilience** at the centre of its analysis.

From the IFRC’s perspective, self-protection is a critical component of resilience building, regardless of the vulnerabilities individuals face. Even in the most complex environments, human capacities are the starting point for addressing risks.

The IFRC defines resilience as: “the ability of individuals, communities, organizations or countries exposed to disasters, crises and underlying vulnerabilities to anticipate, prepare for, reduce the impact of, cope with and recover from the effects of shocks and stresses without compromising their long-term prospects.”³

Academic critiques caution that framing resilience as a responsibility solely at the levels of individuals or communities risks obscuring systemic failures and reinforcing inequalities. This is particularly relevant in humanitarian contexts and within the information ecosystem: resilience cannot rest only on the shoulders of the people already most affected. It must be supported by broader systems that reduce vulnerability, enable equitable access to information and uphold rights. This requires investments in public trust, inclusive governance, meaningful accountability mechanisms and fair access to digital infrastructure and digital literacy. Without these foundations, resilience efforts risk becoming superficial – placing undue burden on the people and communities already most at risk.

The report is composed of eight chapters:

- the information ecosystem, the erosion of trust and global-local narratives ([Chapter 1, on page 29](#), [Chapter 2, on page 69](#) and [Chapter 3, on page 103](#))
- the consequences of harmful information on humanitarian responses ([Chapter 4, on page 137](#))
- regulation and rights ([Chapter 5, on page 167](#))
- community-first approaches ([Chapter 6, on page 211](#))
- challenges related to the Fundamental Principles of the International Red Cross and Red Crescent Movement ([Chapter 7, on page 257](#))
- a forward-looking reflection on resilience in the humanitarian sector ([Chapter 1, section 1.7, on page 46](#)).

Each chapter is written to stand alone while contributing to a coherent whole, recognizing that most readers will engage with selected sections rather than read the report in its entirety.

[Annex I on page 353](#) provides a glossary and [Annex II on page 359](#) focuses on the role of data in humanitarian decision-making.

It should go without saying – but now, more than ever, it must be said unequivocally: those working in humanitarian action understand the stakes of harmful information – the ability to reach people in need, to be trusted to act impartially and to save lives. Mistrust is deepening. Needs are outpacing resources. Access is under threat. Harmful information is not just compounding these challenges – it is actively undermining humanitarian action, distorting public understanding and putting lives at risk. And yet, the humanitarian commitment endures. **Humanity must be the disruptor** – a counterforce to apathy, manipulation and inaction. Law applies. People are in need. The question is no longer whether the sector wants to respond, but whether it is adequately supported, resourced and protected to do so.

In a crowded and contested information space, principled humanitarian action cannot be sidelined. It requires more than statements of intent. It demands investment, collaboration and clear standards that span from policy to practice to the very front lines of humanitarian response. The ability to act with integrity – and to be trusted to do so – depends on it.

“ **Rumours stop with the wise. With skills and abilities, we can address the issue. While other teams were retreating, we could continue rescuing people because we knew how to handle it.**”

Community member, China

This insight underscores the essence of this report: a call to confront the threat of harmful information directly and to reaffirm the role of humanitarian actors in assisting and protecting people, supporting their agency and restoring trust even in the most fractured environments.

Our shared humanity must guide the way forward – to meet today’s challenges with resilience, integrity and purpose.

Endnotes

- 1 IFRC's *World Disasters Report 2005: Focus on information in disasters* emphasized the divide between those who have information and those who do not, noting that too little information was being shared with the very people who aid organizations aim to support. It examined both the quality of the information – right and wrong – and the need for more effective communication.
- 2 IFRC. *World Disasters Report 2013: Focus on technology and the future of humanitarian action*. (2013) p.189. www.ifrc.org/sites/default/files/WDR-2013.pdf
- 3 IFRC. *IFRC Framework for Community Resilience* (2014), p.6. www.ifrc.org/sites/default/files/IFRC-Framework-for-Community-Resilience-EN-LR.pdf





SYCHO UNNY

ADIDAS
THE STRIPES BRAND

CIVIL

Chapter 1

Crisis, chaos and confusion: Understanding harmful information in humanitarian contexts





Chapter 1

Crisis, chaos and confusion: Understanding harmful information in humanitarian contexts



Contents

	Introduction	31
1.1	Defining harmful information	34
1.2	The evolving information ecosystem	34
1.3	Who are the threat actors?	37
1.4	Information integrity in crisis situations	38
1.5	From broadcast to two-way engagement	41
1.6	Who is most vulnerable to harmful information – and why?	44
1.7	What is the impact of harmful information?	46
1.8	Artificial intelligence and harmful information	50
1.9	Narratives shape perception, trust and action	53
1.10	Controlling accessibility and content moderation	55
1.11	Analysis and monitoring of harmful information	56
	Concluding remarks: Harmful information is not background noise	60
	Endnotes	63

Introduction

For the humanitarian sector, the 2004 Indian Ocean tsunami marked the first major disaster to receive widespread digital coverage, while the wars in Afghanistan and Iraq (in the early 2000s) became the first armed conflicts dissected in real time by thousands of online commentators. This signalled the start of a decentralized digital information era, shaped less by traditional media and more by fast, participatory online spaces. Early digital content often failed to reflect local realities, particularly in contexts where local languages were absent from platforms. Yet the emerging blogger community played an active role in reporting, verifying content and calling out manipulated imagery,¹ fabricated reporting² or exaggerated harm. While imperfect,³ these efforts exposed new possibilities for scrutiny and public accountability.

These online blogger communities emerged as a powerful force within the information ecosystem. Enabled by technological innovation and driven by social shifts, individuals were able to share unfiltered perspectives, often in response to growing dissatisfaction with mainstream media. Blogs offered immediacy and more personal, alternative narratives, fostering informal networks of discussion ('communities') and shaping early forms of participatory media. Ultimately, they laid the groundwork for today's decentralized and user-driven communication models.

For humanitarian operational and communication professionals, this period felt fast-paced and unpredictable, heralding changes that were once unimaginable. Humanitarian crises – especially wars – have always been accompanied by the use of information to deceive, manipulate, discredit, disrupt or advance political or ideological agendas.⁴ What has changed is the nature, speed, scale and accessibility of harmful information. Today, a wide range of actors – professional and amateur, anonymous and overt – create and spread harmful information across digital and offline channels. The internet, mobile phones, social media platforms and messaging apps serve as powerful amplifiers. In many spaces, fact, truth and accuracy have been displaced by opinion, emotion and perception – often driven by financial, political or ideological motivations. Polarization and distrust are often not just by-products but deliberate objectives. This dynamic extends well beyond warfare: harmful information now shapes how communities understand, prepare for and respond to disasters, health emergencies, migration and other humanitarian challenges.

Instead of facilitating access to trustworthy, life-saving information, the information space increasingly feeds on people's fear and uncertainty. Crisis information is routinely manipulated, politicized or distorted – including minimizing a disaster's cause or severity. For example, in 2024, false information spread online claimed that Hurricane Helene in the US was geo-engineered.⁵ Such theories circulated widely on social media, illustrating a growing trend of disaster misinformation and disinformation, and mistrust in official information.

If the Indian Ocean tsunami was the first disaster to be digitally witnessed at scale, COVID-19 was the first to be digitally *lived*. Every aspect of the crisis – health guidance, rumours, fear and solidarity – spread online as quickly as the virus itself. COVID-19 – an unparalleled health emergency in living memory – combined a pandemic and **infodemic**,⁶ affecting people worldwide. While the health crisis has largely been contained, the infodemic has left lasting consequences: deepened distrust in authorities, increased vaccine hesitancy and refusal, and the rise of an industry of influencers and conspiracy theorists spreading narratives about the dangers of everything from medical treatments and health care professionals to cell phone towers.⁷



Infodemic

An **infodemic** is too much information, including false or misleading information, in digital and physical environments during a disease outbreak.

The impact of such information is immediate and often harmful. Correcting it is extraordinarily difficult. Exploiting social divisions, political sensitivities and deep-rooted prejudices, harmful information increasingly drives people to turn against authorities, reject aid and dehumanize or mobilize against those perceived as 'other'.



What happens if misinformation lands on me now, somehow it breeds fear on me, and now I find it difficult to identify which one makes sense to me, which one I should follow. Because I get all of them from the media still. I have gotten this one which tells me to do this. I've also gotten this one which is telling me if you do this, you're going to suffer from this, you're going to get this. So I remain there. So it becomes really very, very difficult and confusing."

Community member, Uganda

During the 2024 Valencia flood response in Spain, misinformation claimed that aid was being diverted to migrants rather than reaching disaster-affected communities, fuelling xenophobic narratives. Similarly, in Southport (UK) in 2024, following the killing and injuring of several children at a dance workshop, false rumours and inflammatory content spread that suggested the attacker was a Muslim asylum seeker. This narrative inflamed anti-immigrant and Islamophobic sentiment, triggering violent protests and riots outside a mosque and sparking wider unrest across multiple cities.

Contributor Insight 1.1



Harmful Information and the DANA floods in Valencia (Spain) 2024 (part 1 of 2)

In late October 2024, Valencia was struck by one of Spain's deadliest floods in recent history, caused by a DANA (*Depresión Aislada en Niveles Altos*, or 'cold drop'). Intense rain fell over several hours, overwhelming infrastructure, leaving 236 people dead and devastating communities.

Specific incidents of harmful information that spread on social media included negative comments, insults and threats directed at our workers and volunteers on the streets and, to a lesser extent, acts of vandalism against our offices and vehicles (such as graffiti and flat tyres).

Misinformation about the Spanish Red Cross has circulated for years, but during the DANA it mainly took the form of hoaxes and rumours claiming that the National Society was:

- not present on the ground helping people in Valencia
- only in Valencia taking staged photos and videos, rather than providing real assistance – “your clothes are clean, your work tools are new and recently purchased” (see Figure 1.1)
- helping migrants instead of Spanish citizens affected by the DANA
- misusing funds – lying about their destination and amounts allocated or refusing in-kind donations to keep the money.

Fig 1.1

Posts on X, November and December 2024

A smaller number of recurring posts attacked the Spanish Red Cross' international work (e.g., for Ukraine or Gaza) or resurfaced older posts related to migrants in Spain or blood donations.

In this case, harmful information did not directly hinder the humanitarian response; it was not an obstacle to providing assistance. However, it undoubtedly had an impact on the people involved in the response by:

- creating an extraordinary workload dedicated to denying, explaining or minimizing the impact of harmful information
- causing emotional distress and even doubts about the organization, both within society and among Spanish Red Cross members
- undermining public trust, with some donors questioning the organization.

During the DANA, the groups most vulnerable to misinformation were our staff and volunteers (both on the ground and off), our partners and allies, and, of course, our beneficiaries. Migrants were also targeted in the harmful information – these posts inevitably fuelled xenophobia – but in this case, they were used more as a tool to target the Spanish Red Cross. These groups were emotionally affected to varying degrees and some also experienced a certain loss of trust in the organization.

María Trénor Alvargonzález
 Director of Communication
 Spanish Red Cross

Kenan Terzic
 Social Media Manager
 Spanish Red Cross



These impacts of harmful information are often felt most acutely at the community level, though in some contexts they are being actively fuelled by external actors. This unfolds against a backdrop of declining trust in traditional institutions and growing perceptions of non-governmental organizations (NGOs) as partisan or politically aligned. This erosion of trust is further exacerbated by what has been termed the “death of expertise”,⁸ as journalists, academics and public institutions – once seen as trusted sources of information – are no longer widely trusted in many countries. Meanwhile, AI is being used to produce content that imitates authoritative, expert voices but may be inaccurate, misleading or stripped of its context.

1.1 Defining harmful information

Today’s information ecosystem is highly complex, encompassing many forms of harmful content, including misinformation, disinformation, malinformation, hate speech, propaganda, foreign information manipulation and interference. These forms of harmful information often overlap and reinforce one another, with intent the key factor that distinguishes them.

This report uses the term ‘harmful information’ to focus on its impact and the responses – rather than rigid classifications, which are often politically charged and context dependent. While there is no universally agreed definition, for the purposes of this report, harmful information refers to **information that has the potential to cause, contribute to or result in harm to an individual or entity.**

Legitimate criticism, even when uncomfortable or challenging, is not considered harmful. However, organizations must still acknowledge and address it, as unaddressed criticism can be exploited by others.

Understanding harm is essential not only to assess its impact on individuals, communities, organizations and societies, but also to design appropriate, proportionate and principled strategies for response.

1.2 The evolving information ecosystem

Harmful information is often framed as an online phenomenon, but humanitarian crises have long been shaped by rumours, myths and propaganda spread through offline means – word of mouth, pamphlets, radio and television broadcasts, community meetings and other official or semi-official channels. These offline dynamics remain deeply influential, especially in contexts with limited digital access or low media literacy. Importantly, harmful information often moves fluidly between online and offline spaces, amplifying its reach and creating real-world consequences for individuals and communities.

Contributor Insight 1.2



South Sudan – impact of harmful information on humanitarian response

Incident: In Juba and surrounding areas, rumours claimed that international NGOs were distributing poisoned food as part of a political agenda.

Impact: Aid recipients hesitated to collect aid supplies, and local staff faced verbal threats in at least two documented instances.

Response: The Sentinel team used WikiRumours to quickly verify and publicly refute the false claims, working with local radio stations and community chiefs to reassure residents and restore trust in humanitarian food distribution activities.



Incident: Community members reported hearing that armed youth were advancing on villages in Lainya County, causing families to flee overnight.

Impact: Although the rumour was later proven false, it triggered panic-driven displacement and led to the temporary suspension of local humanitarian activities.

Response: Community ambassadors verified the facts on the ground and corrective broadcasts were issued via radio and SMS the next morning. The rapid clarification helped resume humanitarian activities and deter further displacement or conflict.

Anahi Ayala Iacucci, Nabeel Chudasama,
Nabeela Jivraj, Zainah Alsamman
Grand Challenges Canada

Christopher Tuckwood
The Sentinel Project

The digital information environment – spanning both digital and cyber domains – has become increasingly complex, contested and politicized. States approach this environment from distinct angles. Some focus on **infrastructure** – the **cables**: the physical architecture that enables data flows, such as data centres, cloud servers and undersea cables. Others concentrate on **content**: the narratives, data and discourse circulating within and across platforms. These two lenses – cables and content – are not mutually exclusive, but they shape how influence and control are asserted over the information space, with implications for sovereignty, surveillance and resilience.

Control over content can involve shaping, restricting or distorting information flows for political, ideological or security purposes. This may be achieved through disinformation campaigns, censorship, platform regulation, information manipulation or influence operations. As a result, the boundaries between cybersecurity, information influence operations and geopolitical competition are increasingly blurred. Strategic decisions regarding cables and content have direct consequences for access to information,

freedom of expression and communities' ability to engage safely and meaningfully in the digital space.

At the same time, threat actors' tactics, techniques and procedures are no longer confined to technical systems. They increasingly leverage harmful information motivated by financial, political or ideological agendas, making information manipulation as consequential as cyberattacks on physical systems. Psychological factors are at the core of this threat: harmful information exploits fear, emotion, identity and grievance, deepening social division and making it far harder to counter with facts alone.

Contributor Insight 1.3

Information first responders: Building narrative resilience into humanitarian response

In an age when crises are amplified by misleading information and manipulation, the humanitarian sector must treat the information environment as operational terrain. False narratives no longer circulate on the margins: they now shape perceptions, drive behaviours and determine the legitimacy of aid itself. From manipulated evacuation route guidance to targeted information attacks on humanitarian actors, harmful content is no longer a side effect of crises – it is part of the crisis.

During the early weeks of the COVID-19 pandemic, false cures, conspiracy theories and xenophobic narratives competed strongly with verified public health guidance, often outpacing and overshadowing it in the battle for attention. In more recent crises, from Sudan to Ukraine, misinformation has obstructed safe passage of civilians, fuelled local suspicion toward aid workers and led to violent reprisals. These examples underscore that the costs of harmful information are not abstract – they are immediate and sometimes life-threatening. Waiting to react means being late to the race.

We are experiencing a crisis in our information ecosystem where distorted narratives erode trust, deepen polarization and fracture the relationship between responders and communities. Edelman's Trust Barometer 2025 found that nearly 70% of global respondents worried about being purposefully misled by authoritative sources such as government or media. An alarming 25% of respondents said they were willing to spread disinformation to achieve change, a number that rises to 34% among young adults.

Meanwhile, recent data from the Reuters Institute shows 58% of global respondents are concerned about how they can distinguish truth from falsehood online. In many regions, online influencers and national politicians are now seen as the leading sources of inaccurate information – surpassing traditional media, foreign powers or rogue actors.

For the humanitarian sector, the message is clear: the fight for trust cannot begin after a crisis erupts. Just as supply chains are pre-positioned, communicators, digital monitoring protocols and trusted local messengers must be identified and deployed in advance,

ready to scale up in the event of a crisis, but more importantly an ever-present player on the pitch.

Such 'information first responders' are essential to compete in contested digital spaces and inoculate audiences against rapidly spreading falsehoods. This requires mobilizing authentic individuals who are relatable voices within key communities, building multilingual content pipelines and training field teams to identify and counter harmful narratives. It also requires deploying social media monitoring, real-time alert systems and rapid myth-debunking toolkits. Although AI can cause enormous harm by supercharging the generation and dissemination of fake content, it also presents opportunities to provide the superpower to create accurate and engaging content.

Building resilience to manipulation in a polluted information ecosystem means more than getting the facts right. It means designing communication systems that are agile, community led and built for trust at their core. If humanitarian aid is to remain a neutral and stabilizing force, trust must be treated as a mission-critical capability – one that needs resourcing and defending.

Oliver Hayes OBE

Head of Counter-Disinformation

EMEA

Dave Fleet

Head of Global Digital Crisis and Counter-Disinformation

Edelman

Other factors shaping the information ecosystem are discussed in [Chapter 2, on page 69](#) and [Chapter 3, on page 103](#).

1.3 Who are the threat actors?

A diverse range of threat actors create and amplify harmful information from lone individuals to paid contractors, coordinated troll networks to outsourced call centres, propagandists to inauthentic accounts, national to transnational entities. Each requires a differentiated response. Whether motivated by profit, ideology, coercion or ego, these actors operate within a vast and expanding influence economy. Some act defensively, others offensively, amplifying noise or falsehoods with strategic intent.

The introduction of AI has further lowered the barrier to entry, enabling more sophisticated, scalable manipulation and deepening the asymmetry between those spreading harmful information and those trying to contain it. What most clearly differentiates these actors is their intent – whether to deceive, disrupt, distract, divide or dominate. That intent not only shapes the forms their actions take but also determines the countermeasures required. This is not a level playing field: those spreading harmful content are faster, louder and have fewer constraints than those working to uphold truth and trust.

Fig 1.2 Threat actors involved in harmful information



1.4 Information integrity in crisis situations

The UN has identified strengthening information integrity as “one of the most urgent challenges of our time” – fundamental to human rights, peace and sustainable development. Information integrity refers to the accuracy, consistency and reliability of information. In response to growing threats from misinformation, disinformation, hate speech and the misuse of digital technologies, including AI, the UN Global Principles for Information Integrity⁹ were launched.

These principles are:

- Societal Trust and Resilience
- Healthy Incentives
- Public Empowerment
- Independent, Free and Pluralistic Media
- Transparency and Research.

Preceding these principles, a 2023 UN Policy Brief addressed information integrity on digital platforms, calling for follow-up steps including consultations and the possible development of a code of conduct to help “guide Member States, the digital platforms and other stakeholders in their efforts to make the digital space more inclusive and safe for all, while vigorously defending the right to freedom of opinion and expression, and the right to access information”.¹⁰ The consultations ultimately led to the development and launch of the principles. In September 2025, the UN released the first in a new Issue Brief series titled *From Principles to Practice: Strengthening Information Integrity*.¹¹

Contributor Insight 1.4

Building on UN norms for information integrity

Before designing a plan to address harmful information, we need to understand the global frameworks that guide our actions. Any strategy must align with three key UN pillars:

- 1 **UN Global Principles for Information Integrity** – Recommend action around five principles: public trust, healthy incentives, public empowerment, a free media space, and transparency and research. These principles provide a blueprint for multi-stakeholder collaboration.
- 2 **UNESCO Guidelines for the Governance of Digital Platforms** – Set out five duties, from conducting impact assessments to ensuring meaningful redress.
- 3 **Global Digital Compact** – Adopted in September 2024, commits governments to an open, safe, rights-based digital future, with specific measures to protect information integrity.

These frameworks are not constraints – they are foundations for innovative, multi-faceted action.

Guiding principles and the need for a multi-stakeholder approach

- The UN Global Principles call on all actors – governments, tech companies, advertisers, media and more – to strengthen information integrity.
- They urge the UN to scale up action, which we are doing through system-wide coordination, knowledge-sharing and a working group on information integrity.
- No single actor can succeed alone; coalitions are essential to building global resilience.

Four practical action tracks based on lessons learned

- 1 Multi-stakeholder coalitions and action** – Establish a standing forum of trusted actors, from governments to youth groups, to share information, flag risks and coordinate rapid responses – especially during elections or public health crises.
- 2 Capacity building for resilience** – Equip journalists, educators, fact-checkers and community leaders with the skills, resources and funding to detect and counter harmful narratives, with a focus on under-resourced areas.
- 3 Research and risk assessment** – Create a regional research hub to monitor narratives, publish open data briefs, and provide cross-border early warning.
- 4 Prevention, mitigation and response protocols** – Develop clear, human-rights-based procedures for governments, media, platforms and civil society to act quickly and transparently when harmful content emerges.

Hiroyuki Saito

Director

UN Information Centre, Dakar

Contributor Insight 1.5



Data quality as a lifeline in disaster management

Misinformation, even when not created and spread deliberately, can still negatively affect the effectiveness and efficiency of humanitarian services and disaster management. During the 2003 Bam earthquake in Iran, which claimed thousands of lives and affected over 250,000 people, a major emergency response was launched at both national and international levels. The operation received huge national and international attention and support, and was largely successful, but it also faced challenges including in information management.

Inaccurate and exaggerated reports on the scale and value of international humanitarian financial donations and aid caused three main challenges. First, inaccurate news and

information disrupted planning, implementation and reporting. Second, exaggerated claims about the amount of international aid created doubts among potential donors as to whether further aid and contributions were needed, thereby undermining global resource mobilization and fundraising efforts. Third, inflated figures on the amount and value of international aid heightened expectations among affected people and beneficiaries, creating management challenges. These experiences highlight the imperative that all stakeholders in disaster response and humanitarian services comply with minimum quality standards for data and information – namely **accuracy, timeliness, relevance and reliability** – before sharing and spreading them.

Beyond accuracy and quality, information must also be **comprehensive and integrated** for effective disaster management and risk reduction. Incomplete data and information can impede the achievement of the intended outcomes. Traditionally, the disaster management community has focused on natural **hazards** as the main driver of disasters. Later, it became clear that natural hazards alone should not be blamed for the occurrence of disasters; rather, human and societal vulnerabilities including **physical, social, economic and environmental** vulnerabilities, along with exposure of people, systems and assets to hazards, are the main drivers. These vulnerabilities often stem from human behaviour and unsound policies and practices. Therefore, it is imperative to integrate data on physical, socio-economic and environmental vulnerabilities with information on hazards in order to build a fuller understanding of **disaster risk**. This approach aligns with the first priority of the Sendai Framework for Disaster Risk Reduction (2015–2030): Understanding disaster risk.

Mostafa Mohaghegh

Senior Coordinator of the Asian and Pacific Centre for the Development of Disaster Information Management (APDIM)

UN Economic and Social Commission for Asia and the Pacific (UNESCAP)

1.5 From broadcast to two-way engagement

The promise of new tools for community engagement was meant to mark a broader shift toward more inclusive communication. Digital platforms enabled a move from one-way, broadcast messaging – press releases, websites, official statements – towards two-way engagement. But in practice, this much-valued shift has become increasingly complex. Anyone can now respond to, challenge, reinterpret or amplify a message instantly and publicly.

While two-way engagement invites dialogue, it also opens the door to criticism, harmful content and narrative disruption, including from global audiences and malicious threat actors far removed from the realities on the ground. Social media rarely fosters meaningful exchange, especially when responses spiral far beyond the original message. For humanitarian organizations, committed to neutrality and impartiality, an unpredictable, emotionally charged environment is difficult to monitor and manage, and failure to navigate it can have direct operational consequences.

As of June 2025, platforms such as Google (26 years old), Facebook (21), YouTube (20), Twitter/X (19), VK (18), WhatsApp (16), Instagram (14), Signal (12), Telegram (11) and TikTok (9 in China, 8 globally) reach billions of people worldwide. These platforms have transformed how people access and share information, connect and engage. But they are also powerful vectors for harmful content. Their algorithms prioritize engagement – often the most shocking or polarizing material – because it generates more clicks, attention and ultimately profit. The result is not a glitch but a feature: systems that reinforce bias, create echo chambers, distort perception and deepen division.

The deliberate spread of harmful information erodes trust, casting doubt on humanitarian intentions, principled action and legitimacy. In some cases, it has led directly to threats and violence: refugees and migrants being menaced, humanitarian staff being attacked and volunteers facing hostility.



During the Ebola outbreak in Guinea in July 2014, the President of the Red Cross Society of Guinea warned that violence against volunteers was hindering access to affected communities and disrupting safe burials. These attacks stemmed from misconceptions and fear surrounding Ebola, despite efforts to counter rumours through radio, television and community engagement. By working with religious leaders, families and local stakeholders, the National Society developed respectful burial practices adapted to cultural norms – allowing people to honour the dead.¹²

Social media users increasingly expect rapid, authentic responses. This creates pressure on organizations to engage in real time, often in tension with centralized approval processes, coordination needs with teams on the ground and the imperative for accuracy. For example, during safeguarding or integrity incidents, the demand for swift and transparent communication can clash with the requirement to verify facts, uphold duty of care and ensure due process. Delays or overly cautious messaging may then be perceived as prioritizing organizational reputation or funding relationships.

In an environment where harmful information moves faster than facts and institutional credibility is under constant scrutiny, even well-intentioned engagement can appear reactive, defensive, disingenuous or lacking empathy. Building genuine two-way engagement requires more than responsiveness – it demands intention, transparency and sustained presence. Yet, when malicious actors exploit an organization's communications, such as using comment sections to disrupt, derail messages or discredit an organization, two-way engagement may not shift the narrative. Many organizations avoid responding for fear of escalation, but silence can leave harmful narratives unchallenged and weaken credibility. Comment sections, once viewed as spaces for engagement, have become increasingly challenging to moderate. The sheer volume of harmful content they attract, combined with limited capacity to manage it, has led some news and media organizations¹³ to disable them to reduce reputational risks and curb the spread of harmful information.

A good example of responding comes from Haiti during the 2010 earthquake. At that time, the IFRC established a Beneficiary Communication Programme that created ten access points for individuals and communities to share feedback on the humanitarian aid they were receiving and express their needs. These access points included live radio programmes, text messaging, information centres, direct engagement with communities

and call centres. This initiative enabled the IFRC to continuously adapt its programming and make more informed decisions based on direct community input.

Knowing when – and how – to communicate is increasingly difficult. Delays or silence often create a vacuum that others, whether ill-intentioned or misinformed, will quickly fill.

Contributor Insight 1.6



A communications failure and moment of reckoning

The civil unrest of July and August 2024 in Bangladesh will be remembered not only for its political consequences – the fall of a government and the arrival of an interim one – but also for the way it tested institutions. Among them was the Bangladesh Red Crescent Society, an organization with a long-standing legacy of service, suddenly caught in the crossfire of public distrust and internal challenges.

Although the National Society responded during the unrest – quietly, courageously and across multiple districts – the dominant narrative in the media and among segments of the public was one of absence, silence, even complicity. Accusations of inaction spread quickly and damaging speculation about political alignment overshadowed the life-saving work taking place on the ground. As the situation stabilized and the fog of unrest lifted, it became clear that this wasn't just a communications failure: it was a moment of reckoning. While the Bangladesh Red Crescent Society volunteers were providing first aid to protesters, law enforcement personnel and bystanders alike, and while ambulances transported hundreds of injured people to safety, the broader public narrative told a different story. Many simply didn't know what the National Society was doing. And in the absence of clear, timely and confident communication, that mistrust deepened.

In reality, the Bangladesh Red Crescent Society had mounted a significant humanitarian response. Twenty-one teams were deployed across nine branches, hundreds of people received first aid, and food assistance reached more than 2,000 families. But as the National Society quickly recognized “we hadn't told this story in time, or in the right way. And in a crisis where perception shapes reality, that silence came at a cost”.

Given the volatile and politically charged context, the Bangladesh Red Crescent Society leadership had made a strategic decision to maintain a low public profile – intentional discretion – to minimize the risk of politicization, safeguard volunteers and avoid being drawn into partisan narratives. While this approach was understandable from a risk management perspective, this silence and lack of public visibility created an information vacuum – one that was quickly filled by speculation, criticism and political narratives beyond the organization's control. The consequences went beyond reputational damage. In some districts, volunteers were harassed, partnerships with local NGOs were strained and questions from donors and Movement partners raised concerns about the National Society's ability to operate in politically sensitive contexts while maintaining the trust of all stakeholders. Recovering from this reputational risk requires more than image

management. It has called for a long-term effort to reinforce institutional independence, diversify governance and rebuild public trust.

Alberto Bocanegra

Head of Delegation for Bangladesh

IFRC

1.6 Who is most vulnerable to harmful information – and why?

Despite the unprecedented volume of information available today, many people still live in information vacuums where vital, life-saving information is inaccessible or unreliable. Harmful information can affect individuals, communities, institutions and entire societies, but certain groups are disproportionately at risk due to structural, contextual or situational vulnerabilities. As political and security analysts Singer and Brooking observe: “Like any viral infection, information offensives work by targeting the most vulnerable members of a population – in this case, the least informed.”¹⁴

In humanitarian contexts, those in the most vulnerable situations may include:

- **Crisis-affected populations:** People living through armed conflict, disaster, displacement, migration or health emergencies often face disrupted communication channels, heightened anxiety and scarce or manipulated information.
- **Groups that are marginalized or socially excluded:** Communities facing discrimination based on race, ethnicity, gender, religion, disability or legal status are frequently targeted by harmful narratives and often lack access to reliable sources of information.
- **Youth and older populations:** Young people may be more exposed to harmful information due to time spent online, peer influence or online trends. Older populations may struggle with digital literacy and navigating new information environments.
- **Humanitarian staff, volunteers and organizations:** These are increasingly targeted by narratives questioning their neutrality, independence, impartiality or integrity, undermining trust, access and security.
- **Journalists, human rights defenders and civil society leaders:** People in these roles are often deliberately discredited, silenced or endangered in polarized environments.
- **People with limited access to information or media literacy:** Without trusted information channels or critical thinking skills, these individuals are more likely to believe and spread false content.



I don't think it's necessarily low-income people or the elderly who are more susceptible. Rather, it's those who are more easily influenced by their own emotions and by information – this has nothing to do with age. You'll find that some people are especially opinionated and unwilling to change their views. In some cases, certain 'leaders' see their own judgement as superior to others and insist they're right. Their will, or their subjective descriptions, can drive the spread of information – and they believe it themselves. For example, I might be a 20-year-old rescuer, but if the person is 50 years old, even if I tell them what's correct and what's not, they may question me and say, 'I've eaten more salt than you've eaten rice; I know better'. So it's not age that determines susceptibility to misinformation – it's subjective bias."

Community member, China

Identifying and understanding these vulnerabilities, along with the specific risks of harmful information in different contexts, is now essential to mitigating their impact as part of humanitarian response. In a survey conducted by IFRC with 400 staff members, **50% reported observing misinformation, disinformation or hate speech affecting humanitarian work**, with the most frequently cited issues relating to vaccinations, health and migration.

Contributor Insight 1.7

UNHCR staff survey

Threats to information integrity on digital platforms, such as misinformation, disinformation and hate speech (MDH), are worsening and leading to real-world harms, especially in the most vulnerable humanitarian contexts of armed conflicts, disasters and other emergencies. These harms can be physical, such as violence and killings, economic, social and psychological, eroding trust and safety capacities. Humanitarian actors also face significant risks: staff safety and security, operational neutrality, public support and fundraising capacity can all be threatened. Ultimately, information inequality threatens the ability of refugees to access life-saving protection information on digital platforms.



To better understand the impact of MDH, UN High Commissioner for Refugees (UNHCR) conducted two global surveys in late 2023 – one with forcibly displaced and stateless communities, and another with UNHCR staff. Among forcibly displaced and stateless respondents, 85% reported seeing hate speech or misinformation targeting their communities and 72% said they had personally been targeted. Most UNHCR staff reported witnessing MDH affecting the organization's mandate: 37% saw such content monthly, 60% had seen hate speech with direct impact, 20% had been personally targeted, and one-third said colleagues had been affected. Nearly all (92%) expressed concern about MDH's impact on the people UNHCR protects. On a more positive note, one in three

staff noted that refugee-led or community-based groups are actively working to counter MDH – collaboration with these actors has proven vital in effective responses.

Gisella Lomax

Senior Advisor, Information Integrity

UNHCR

1.7 What is the impact of harmful information?

Harmful information threatens the lives, safety and dignity of people in humanitarian crises. It can escalate violence during armed conflict and disasters, distort realities on the ground, mislead people about the availability of aid and influence life-saving decisions – such as whether to stay or flee, or accept or reject medical care.



So misinformation is very dangerous – very, very dangerous, not only in the bigger spectrum. It can also be in our families. It can break families. It can confuse and hijack people’s future. It can wreak havoc on people. It can bring about violence. It can cause a lot. It’s a terror of its own, in my opinion.”

Community member, Uganda

UNHCR has documented a wide range of offline harms linked to harmful information in humanitarian contexts, including xenophobia, racism, persecution, violence, killings, forced displacement, trafficking and exploitation, barriers to accessing rights and services, damaged reputation, erosion of trust and legitimacy, diminished ability to protect and support refugees, threat to the physical security of humanitarian workers and decreased donor support.¹⁵

The World Health Organization (WHO) has highlighted the psychological and social harm caused by harmful information, both within communities affected by armed conflict and among humanitarian responders. As noted in a WHO review, “Based on the available evidence, people are feeling mental, social, political and/or economic distress due to misleading and false health-related content on social media during pandemics, health emergencies and humanitarian crises.”¹⁶ This occurs when harmful information discourages people from seeking or accessing humanitarian services and undermines organizations’ ability to deliver and implement effective interventions.

The International Committee of the Red Cross (ICRC) has noted that sharing certain forms of harmful information may violate international law. International humanitarian law imposes “important limits on publishing or sharing certain forms of [harmful information]”,¹⁷ such as that which encourages violations of international humanitarian law, threatening violence which spreads terror among civilians, unduly interfering with humanitarian or medical work, and publishing images of prisoners of war.









Under international human rights law and other rules of international law, advocating hatred that constitutes incitement to hostility, discrimination or violence and inciting genocide, may amount to violations.¹⁸ The UN Special Rapporteur for Freedom of Expression has cautioned that state-sponsored disinformation “has a potent impact on human rights, the rule of law, democratic processes, national sovereignty and geopolitical stability because of the resources and reach of states and because of their ability to simultaneously suppress independent and critical voices in the country so that there can be no challenge to the official narratives.”¹⁹

The IFRC is concerned that some donors have given undue weight to rumours instead of engaging directly and formally with the affected National Red Cross and Red Crescent Society. This has led to opinions being formed based on disinformation rather than verified facts. While these rumours may have served the interests of certain individuals or institutions, they have negatively impacted communities and hindered the ability of National Societies to expand and deliver essential humanitarian services.

Laws, policies and plans underpin all aspects of disaster risk management (DRM), protecting and preparing communities all around the world. Robust legal and policy frameworks are therefore a crucial piece of the puzzle for promoting information integrity and addressing the challenges posed by harmful information in DRM. To respond effectively, humanitarian actors must understand how harmful information disrupts response efforts. A typology of harm is important in building an evidence base that supports efforts to identify, measure and mitigate these effects. Each of the following types of harm can significantly undermine humanitarian action and all must be better understood, monitored and addressed.



Table 1.1 **Typology of harm** ↔

Type of harm	Examples of harm
 Physical	Physical injury, loss of life, or incitement to violence, panic or unsafe behaviours, medical treatment avoidance. Physical attacks on humanitarian vehicles, facilities or offices.
 Psychological	Emotional or mental trauma, fear, anxiety, disorientation, discrimination, shame, distrust, manipulation, bullying, stalking, harassment.
 Social	Disruption to social cohesion and trust within communities, erosion of relationships, stigma, social fragmentation, divisions, polarization. Forced or induced displacement of populations leading to family separation, loss of community bonds and weakening of collective support systems.
 Societal	Shrinking space for humanitarian action, undermining trust in institutions, reputational damage, operational disruption, impaired service delivery, silencing of sectors of society, erosion of rights, restricted access to information, exclusion, limits to freedom of expression. Use of harmful information to justify legal persecution, criminalization of speech, or abuse of judicial systems.
 Informational	Distortion, suppression or manipulation of information; loss of access to accurate, timely or trustworthy information; saturation with falsehoods (information overload); erosion of the shared understanding needed for decision-making.
 Deprivational/ financial/ economic	Livelihood disruption, economic loss, loss of access to essential resources or services, financial losses, theft, looting, fraud, scams, diversion of resources, inability to procure necessities, restrictions on funding, extortion, loss of reputation, loss of donor support.
 Digital/ technological	Attacks on digital identity, doxing, algorithmic amplification of harmful content, deepfakes, bot-driven abuse, platform manipulation (see Annex I: Glossary, on page 353).
 Longitudinal/ intergenerational (cross-cutting dimension)	Lasting effects on children exposed to harmful narratives, breakdown of intergenerational trust, perpetuation of stereotypes or trauma, loss of hope.

Contributor Insight 1.8



Online racism and mental health after disasters

In the aftermath of disasters, communities seek solace and solidarity. Yet, for many racialized groups, tragedy is compounded by online racism. Following the June 2025 Air India crash, which claimed at least 290 lives,²⁰ Indian communities around the world were subjected to a torrent of racist abuse online – mocked with slurs, stereotypes and dehumanizing jokes.²¹ Similarly, after the 2023 earthquakes in Turkey (which killed more than 50,000 people including 7,200 Syrian asylum seekers) surviving Syrians, already marginalized, faced scapegoating and xenophobic attacks both online and offline. Accusations circulated that they were looting or ‘taking resources’ meant for local citizens.²²

This form of racism is not only offensive – it is psychologically harmful. Studies consistently link racial discrimination to heightened rates of anxiety, depression, post-traumatic stress disorder and suicidal ideation.²³ When racism follows disaster, it disrupts collective mourning, isolates survivors and undermines long-term recovery.²⁴ For diaspora and asylum-seeking communities, it reinforces a sense of unbelonging in the countries where they have sought refuge or migrated.²⁵

These patterns reveal systemic failures: social media platforms that allow hate to flourish, governments that underfund anti-racism work, and humanitarian systems that often overlook the intersection of race, mental health and vulnerability. Addressing the mental health consequences of racism following disasters requires a coordinated and proactive approach. This is needed because “human well-being occurs in the context of, and is dependent on, an individual’s personal and a people’s collective ability to meet challenges and adversities in their environment through the intelligent, skilful, and ethical use of strengths and resources available to them. Person and community coexist – dynamically, interdependently, and integrally”.²⁶

Social media platforms must be held accountable for the spread of hate speech. This includes implementing advanced detection systems – potentially AI-powered – backed by transparent reporting mechanisms and clear timelines for content removal. At the same time, the mental health needs of the human moderators tasked with viewing and removing racist and traumatic content must also be proactively addressed.²⁷ All post-disaster response programmes, including mental health and psychosocial support, should integrate culturally competent, anti-racist care and support. Practitioners should be trained to recognize and respond to the specific psychological toll of racism, particularly in the wake of a crisis. Governments should lead public education campaigns rooted in evidence and community engagement, to counter post-disaster scapegoating, xenophobia and misinformation. These should be embedded in schools, workplaces and public institutions to have lasting impact. Ultimately, this is about upholding human dignity and humanity.

Yasin Duman

Research Specialist

Red Cross Red Crescent Movement Mental Health and Psychosocial Support (MHPSS) Hub

Shona Whitton

Technical Advisor

Red Cross Red Crescent Movement Mental Health and Psychosocial Support (MHPSS) Hub

1.8 Artificial intelligence and harmful information

People's opinions are increasingly shaped by systems they neither fully understand nor can meaningfully challenge. Many assume they are seeing the same content as others on their feeds, but this is rarely the case. AI-driven recommendation algorithms curate what content surfaces, often reinforcing emotional reactions and echo chambers. This dynamic subtly but powerfully erodes individual agency in how information is encountered, interpreted and acted on.

The harmful information landscape has rapidly evolved with the proliferation of AI tools. The publication of the first *International Scientific Report on the Safety of Advanced AI* (commonly referred to as the 'Bengio Report') in January 2025 underscored that, if discerning truth from falsehood was already difficult, **AI has fundamentally shifted the balance of power in favour of those who control the production and dissemination of information.** The report notes that AI reshapes how information is produced, who is empowered or disempowered by it and how it is manipulated. AI can generate persuasive, human-like content, rapidly and at scale – despite lacking deep conceptual understanding. As such content is often indistinguishable from human-produced material, people tend to overestimate their ability to detect it, increasing their vulnerability to manipulation. Emotionally charged or personalized content, especially when combined with social media data, can strongly shape perceptions and beliefs.²⁸

The broader impact of AI-driven disinformation is still debated. Some studies suggest its spread and effects may be limited; others warn it may be more concentrated, harder to trace and prone to unintended consequences. What is certain is that the information space is becoming more complex and contested. Although no scientific consensus exists on the full societal impact of false information, its viral potential is well documented. Efforts such as adding watermarks or filtering content show some promise in detecting and mitigating AI-generated manipulation but remain limited. Moreover, interventions must balance curbing manipulation with protecting free expression. Reports point to an increase in the prevalence of AI-generated deepfake content (see box) but overall, scientific evidence remains limited. Anecdotal reports of harm caused by AI-generated fake content is growing, but without better evidence the full scale and impact remains difficult to assess.²⁹



AI-generated fake content

Audio, text, or visual content, produced by generative AI, that depicts people or events in a way that differs from reality in a malicious or deceptive way, e.g., showing people doing things they did not do, saying things they did not say, changing the location of real events, or depicting events that did not happen.³⁰



Deepfake

Synthetic media – most often video, audio or images – created using artificial intelligence techniques, particularly deep learning, to realistically manipulate or generate content that portrays events or people in ways that did not actually occur.

A subset of AI-generated fake content. Examples include video face swaps, fake voice recordings.

As AI-generated content becomes increasingly indistinguishable from human-created material, detection remains a persistent challenge. Media authentication techniques such as digital watermarks offer some protection but are easily bypassed or removed, especially in high-risk or adversarial environments.³¹

Contributor Insight 1.9

Detection of synthetic content in critical contexts

Generative AI presents a new opportunity to create and spread harmful information. From AI fabricated 'leaked conversations' of election candidates to AI-manipulated videos of key military figures communicating false information, such content can be used to mislead and manipulate. Deceptive synthetic content disrupts the information ecosystem in various ways. It can be used to create highly realistic depictions of events that have not happened. It can also be exploited to discredit genuine evidence and authentic videos by enabling political figures to claim real footage is fake and has been AI-generated. It can distort the public's perception of ongoing humanitarian crises by fabricating and disseminating AI-generated videos portraying real events.

The threats posed by generative AI can be countered through technical solutions, such as AI detection tools that identify unnatural patterns or inconsistencies in content to determine the presence of potential AI manipulation. One example is the Deepfakes Rapid Response Force, an initiative of the NGO WITNESS, which supports detection efforts to mitigate information crises in real time. The initiative connects frontline information actors with leading AI detection experts to assist them in analysing cases of suspected AI-generated content.

The work of the Deepfakes Rapid Response Force provides a sociotechnical perspective to make AI detection more effective in real-life scenarios, highlighting the need for tools that can process low-quality, highly-compressed content, function across diverse languages and cultural contexts, and deliver analysis in a clear and actionable manner. To ensure that AI detection technology can effectively support humanitarian actors in crisis situations,

the detection solutions must be technically advanced and also tailored to the needs, constraints and experiences of users operating in complex, high-stakes contexts.

Zuzanna Wojciak

Program Associate, Technology Threats and Opportunities

WITNESS

There is no consensus on whether more realistic fake content leads to more effective manipulation, or whether the main barrier is distribution. Some experts argue that the greater challenge lies in spreading fake content at scale – not creating it.³² Research also suggests that ‘cheapfakes’ (simple, less sophisticated manipulations of audiovisual content) may be just as harmful as sophisticated deepfakes, reinforcing the idea that distribution and reach may matter more than quality.³³ While social media platforms deploy moderation, labelling and source credibility checks to limit the spread of manipulated content, these measures raise concerns about free expression. At the same time, research shows that algorithms often “prioritise engagement and virality over accuracy or authenticity of content”, potentially aiding the rapid spread of AI-generated content to manipulate public opinion.³⁴

The UN report *Governing AI for Humanity* emphasizes that it is “more useful to look at risks from the perspective of vulnerable communities and the commons”.³⁵ It draws on the AI Risk Global Pulse Check, a survey of 348 AI experts that captures emerging AI-related trends and risks. The report stresses, however, that risk management must go beyond simply listing or prioritizing risks. It advocates for framing risks through the lens of vulnerability – shifting the focus from **what** the risk is (e.g., ‘risk to safety’) to **who** is at risk, **where** these risks occur and **who** should be accountable. This approach highlights the vulnerability of individuals, political systems, society, the economy and the environment. In the context of safety, the report underscores the importance of ensuring reliability and interpretability of AI systems and of assessing and mitigating risks to individual and collective rights, national and international security, and public safety across diverse contexts.³⁶

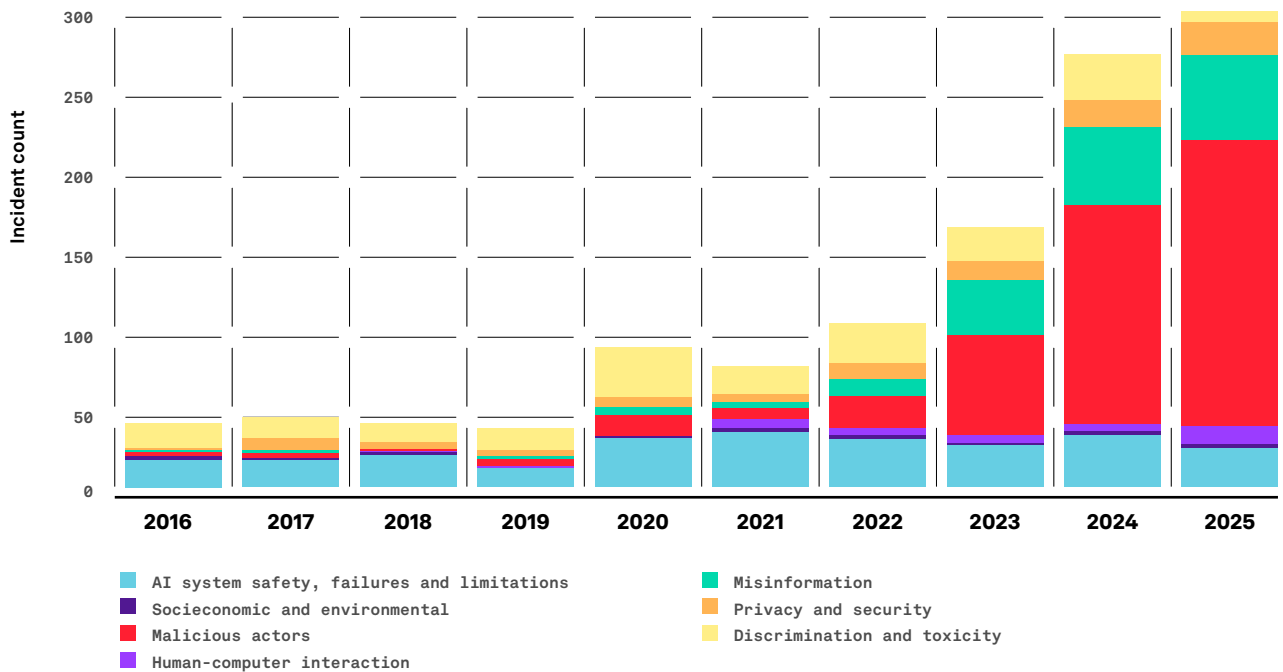
Access to AI tools also remains uneven, reinforcing existing digital divides. People without reliable connectivity, digital literacy or language inclusion continue to be excluded or misinformed.

What is clear is the need for systematic monitoring.

Fig 1.3

AI Incident Tracker

This tracker illustrates how the severity of AI incidents has changed over time.



Sources: MIT AI Risk Initiative — AI Incident Tracker & Scalable AI Incident Classification dashboards (based on the AI Incident Database)

See [Chapter 5, on page 167](#) for reflections on regulation, rights and policy.

1.9 Narratives shape perception, trust and action

Narratives are the frameworks through which people make sense of their experiences and the world around them. They connect the personal to the political, the local to the global. Strong narratives simplify complexity, offering coherence – often at the expense of nuance. Narratives matter. In harmful information campaigns, they often carry historical or cultural weight, with many actors playing a long game.

An event holds power only if people **believe** it happened. In today's information environment, perception can outweigh truth: fabricated incidents can generate real consequences, while verified ones may be ignored. Outcomes are increasingly shaped less by *facts* and more by psychological, political and algorithmic influence. Trust is easily eroded by distortion and distraction. Political discourse often borrows tactics from the use of information in warfare, while armed conflict is shaped by battles over online narratives.³⁷ The risk is that we all become participants when we click, like or share – feeding a wider contest over attention, belief and bias.

In armed conflicts, civilians are often inadvertently drawn into the information dynamics, exposing them to serious risks to their safety – for example, by posting geolocated photos or videos of shelling on social media, relaying updates about territorial control or documenting attacks to seek help or accountability.

These overlapping, often invisible, struggles shape how we perceive the world and act within it. In this ‘network of networks’, there is no neutral ground³⁸ – the objective is influence, control and power. As Singer and Brooking note, influence is achieved not only through force but through shaping how people interpret the world, triggering emotional reactions that drive action and cultivating a sense of shared identity and belonging. Those who succeed are the ones who construct and spread compelling narratives – consistently, repeatedly and globally.³⁹

Effective narratives in the digital landscape rely on simplicity, allowing messages to be absorbed quickly. This is why images resonate so powerfully. Resonance and novelty drive engagement. Yet simplicity is not something the humanitarian sector excels at; some argue the sector has even complicated public understanding of the humanitarian endeavour.⁴⁰

Geopolitical tensions are intensifying polarization within societies and across borders, fuelling division, mistrust and competing narratives that complicate humanitarian action and undermine social cohesion. Returning to the earlier discussion of cables and content, control over both is increasingly treated as a strategic priority, with direct consequences for humanitarian response.

Contributor Insight 1.10



The impact of harmful information on civilians: Insights from two contexts

Across the world, harmful information deepens mistrust in institutions, polarizes communities and undermines people’s ability to make informed decisions. In 2025, the CDAC Network conducted research in two very different crisis settings – Sudan and Lebanon – to better understand how harmful information distorts civilian decision-making, erodes trust and disrupts daily life. Despite the contexts differing significantly, the patterns are consistent, offering insights into the broader dynamics of harmful information in humanitarian crises.

In both countries, harmful information shaped critical, often irreversible, decisions. In Lebanon, refugee families delayed or rushed returns to Syria based on social media rumours of border closures or new UN compensation packages, none of which could be verified. In Sudan, manipulated narratives about territorial control misled civilians into returning to contested areas or postponing evacuations, placing them in direct danger.

Harmful information also fractures communities. In Lebanon, economic uncertainty and political disillusionment created fertile ground for scapegoating. Refugees were portrayed as favoured, host communities as neglected, while misinformation circulating in

WhatsApp groups inflamed resentment. In Sudan, NGO workers, mutual aid groups and ethnic minorities were branded as politically aligned, undermining trust and triggering violence against members of civil society.

The impact on livelihoods was equally stark. In Sudan, rumours targeting displaced groups led to exclusion from markets, housing and informal jobs. In some areas, entire labour sectors – such as butchery or tea vending – were destabilized as fear and suspicion drove down participation and income. These dynamics exacerbated existing inequality and fuelled intercommunal tensions.

Finally, the research revealed how harmful information undermines both institutional and informal trust. In both countries, teachers, activists and aid workers who once acted as information bridges now hesitate to speak out, fearing backlash. Across both contexts, the findings are clear: harmful information does not just confuse, it divides, isolates and destabilizes, with profound consequences for crisis-affected populations.

Ila Schoop Rutten

Information Integrity Lead

Communicating with Disaster Affected Communities (CDAC) Network

1.10 Controlling accessibility and content moderation

Content moderation, which combines human oversight with automated tools such as fact-checking and content removal, is conducted according to platforms' own terms of service. In this role, platforms act as gatekeepers, determining what is permissible and facing criticism both for failing to act decisively against harmful content and for over-censoring. Recently, major platforms have scaled back their moderation efforts, driven by economic pressures, reputational concerns and political or ideological influences. Increasingly, they rely on user reporting or content de-ranking as a way to comply with legal requirements.

Although the internet was founded on openness and trust, it remains vulnerable to shutdowns, blackouts and targeted restrictions.

Contributor Insight 1.11



Locally led action in the struggle for humanitarian connectivity

Since 2021, millions of people across Myanmar have been disconnected from the internet, especially in conflict-affected regions. Access Now and the #KeepItOn coalition – uniting

more than 334 national, international, regional and local organizations – recorded 85 shutdowns in Myanmar in 2024 alone.⁴¹

This is hardly an isolated case. In 2024, the #KeepItOn coalition documented 296 shutdowns in 54 countries, an all-time-high and a 35% increase in the number of countries where shutdowns occurred compared with the previous peak in 2022.

The 2024 data confirms that conflict was the leading trigger for internet shutdowns for a second consecutive year, with an expanding arsenal of tools deployed to cut connectivity – jamming devices, cutting cables, destroying infrastructure and sabotaging internet service providers. Humanitarian crises are increasingly becoming complex emergencies with some countries affected by a layering of disaster, climate change and conflict.

Myanmar is a striking example, hit by Cyclone Mocha in 2022 and a devastating earthquake in 2025, all while continuing to suffer from internal violence. When Cyclone Mocha hit western Myanmar, the lack of internet connectivity exacerbated the storm's impact. People were unable to access vital information on evacuation efforts, warnings and even relief efforts because the internet was disconnected. Three years later, the earthquake in the Sagaing region claimed over 3,000 lives. The region had already been subjected to a prolonged connectivity blackout since 2021. The shutdown, which impacted over one-third of the country, once again complicated rescue and relief operations.

The essential role of connectivity for the civilian population during disasters underscores the importance of resilient and grassroots-led information management. Shutdowns, especially during complex crises, hampering timely, safe and effective humanitarian response may ultimately be responsible for preventable deaths. This environment can also hinder telecom operators from deploying enough crisis response infrastructure, as necessary authorizations are delayed or blocked due to political or tactical considerations.

While humanitarian reporting by international actors focused on the visible outcomes of disasters and conflict, local advocacy groups consistently raised alarms about connectivity disruptions, their consequences and the harms generated.

Wai Phyoo

Asia Pacific Policy Analyst

[Access Now](#)

Faiz Naeem

Asia Pacific Program Associate

[Access Now](#)

Giulio Coppi

Senior Humanitarian Officer

[Access Now](#)

1.11 Analysis and monitoring of harmful information

Analysing harmful information and its risks is essential to understanding the tactics, techniques and procedures used by those who create and disseminate it. Such analysis must consider contextual, historical, political, social, cultural and economic factors. Several analytical frameworks exist, including the 5 Ws⁴² (commonly used by journalists), ABCDE,⁴³ DISARM⁴⁴ and MITRE ATT&CK,⁴⁵ each requiring different levels of training and analytical capacity. This is explored more fully in [Chapter 3, on page 103](#).

Effective analysis also relies on monitoring both the operational and information environments to detect harmful content, while integrating community-based approaches to better understand its dynamics and impact. Many humanitarian organizations are using a range of tools – commercially available or internally developed – that combine automated and manual methods to scrape public data. This is too often with a narrow focus on reputation, rather than assessing how harmful information affects their security, operations and the broader humanitarian context.

The humanitarian sector urgently needs more systematic monitoring and social listening – including sentiment analysis – to detect, understand and frame responses to harmful information. In this regard, monitoring remains fragmented, underfunded and often ad hoc. By incorporating community-based approaches and sentiment analysis into social listening, organizations can detect and track not only what is being said but also how people feel – fearful, angry, mistrustful or supportive – providing deeper insights into shifting perceptions and risks.

Without structured, real-time detection and analysis of both online and offline narratives, humanitarian actors risk being reactive rather than proactive, allowing harmful information to spread unchecked and erode trust before it can be countered. Systematic monitoring, social listening and sentiment analysis, grounded in ethical safeguards and community participation, would allow organizations to detect emerging narratives early, anticipate risks, adapt communications and programmes in real time, and strengthen resilience at both community and institutional levels.

However, there is general acknowledgement that optimal monitoring tools have yet to be deployed, not because of a lack of tools but primarily due to resource constraints. Concerns also persist around the legal and ethical implications of data scraping, which can be perceived as intrusive or akin to surveillance.

Contributor Insight 1.12

Detection and monitoring of harmful information in humanitarian and crisis contexts

In humanitarian and crisis-affected contexts, harmful or misleading information about humanitarian efforts can jeopardize the safety of humanitarian personnel, undermine operational security and render aid delivery ineffective. At LinkAlong we have worked extensively on this challenge, developing an AI-based open-source monitoring platform tailored to detect and track such threats.

Due to recent advances in AI-based natural language processing, there is now a significant opportunity to systematically monitor large volumes of open-source content in an automated way. Leveraging large language models, it is possible to detect, classify, aggregate and characterize harmful information with high precision and in near real time. A typical workflow involves three steps:

- **Data collection** from diverse sources, including open social media, public forums and relevant news outlets.
- **Deep text analysis** using AI to identify narratives, classify types of harmful content and assess potential risks.
- **Dissemination of results** through timely, structured reporting to stakeholders, enabling rapid and informed responses.

However, several challenges hinder widespread adoption. First, large language models still struggle with low-resource languages and dialects common in crisis regions. Second, access to relevant social media data is increasingly restricted, while at the same time harmful narratives often migrate to encrypted or closed platforms such as Telegram or WhatsApp. Third, regional differences in social media usage require customized solutions, which can limit scalability. Fourth, most organizations lack the budget for sustained monitoring, and donors often undervalue this activity despite its strategic importance.

Efforts are also fragmented: many organizations develop isolated, homegrown tools that are costly to maintain and miss opportunities for information sharing, even though information threats and narratives often overlap.

The urgent need is for greater **cooperation between organizations** to pool resources, share information threat intelligence and jointly leverage sustainable solutions. The technical foundation already exists; the remaining challenge is building the business case and convincing relevant stakeholders that collaborative monitoring is an operational necessity, not a luxury.

Karl Aberer

Professor, Swiss Federal Institute of Technology in Lausanne

LinkAlong⁴⁶

Contributor Insight 1.13

Building resilience: Catalogue of fact checks, false claim markers and provenance support

Among monitoring tools developed to provide structured, machine-readable data on false and misleading claims circulating online, NewsGuard's False Claim Fingerprints database tracks provably false narratives across 102 countries in local languages. It covers topics such as humanitarian crises and wars, health emergencies, earthquakes, climate-related disasters and migration. It also monitors information operations by malign actors worldwide. The database serves as an early warning system and fact-based 'guardrails' for high-risk false claims, providing a 'Fingerprint' for each (detailing the evidence for why the claim was determined false) with human-verified URLs,⁴⁷ keywords, search terms, social posts, multimedia assets and language excerpts.

NGOs and other institutions can use the database to detect false claims targeting them – or the vulnerable communities they serve – before those claims enter the mainstream. NewsGuard detects and flags high-risk claims before any public reporting or fact-checking in two-thirds of cases, and is the first to debunk false claims by authoritarian governments 83% of the time. Public relations teams can integrate NewsGuard’s journalistic assessments into their communication responses, using accurate, transparent, evidence-backed insights to defend their reputation when under attack. As more internet users turn to generative AI chatbots for information and news, such a database can be used to verify whether false claims have contaminated AI-generated responses, enabling organizations to respond quickly and strategically.

Chine Labbe

Senior Vice President, Partnerships and
Managing Editor, Europe and Canada

NewsGuard Technologies⁴⁸

Contributor Insight 1.14

Securing communication in humanitarian contexts

In an era where harmful information can spread rapidly and undermine trust in humanitarian responses, ensuring that communication channels are secure, reliable and trusted is paramount. Twilio is a customer engagement platform, partnering with governments and service organizations worldwide to deliver critical information to those in need. When technology platforms directly affect people’s lives, regulatory discussions on safety and security become essential.

Companies can provide valuable insights into how communications platforms operate and the potential effects of regulation on both innovation and user safety. Twilio’s approach is to work proactively with industry stakeholders to safeguard users while preserving open, reliable communications – especially where safety and well-being are at stake. Twilio has robust policies and enforcement mechanisms to prevent misuse of its platform, including violations of its Acceptable Use Policy. Automated systems monitor for security incidents and misuse of services.

Technology and digitalization are having a meaningful impact on humanitarian crisis response. A [European Union report](#) found that digitalization can streamline processes for refugees and migrants, particularly supporting their integration into host communities.

Twilio has a [multi-year partnership with IFRC](#), which provides cash and voucher assistance to people affected by disaster or conflict. The process, enabling the IFRC to confirm the identity of displaced people at scale, was enhanced by the [integration of a secure digital](#)

registration system through SMS and WhatsApp. This collaboration shows how cross-sector partnerships to bring aid can be underpinned by secure communication platforms.

Chiara Kunnie

New Business Manager, Social Impact Sales

Twilio

Alexandros Koronakis

Senior Director, Head of Government,
Policy & Regulatory Affairs, EMEA

Twilio

Concluding remarks: Harmful information is not background noise

Harmful information is not background noise: it actively shapes how people understand crises, who they trust and whether they can access humanitarian assistance and protection. It influences safety and security both directly and indirectly. The struggle over harmful information is as much about cables as it is about content – the infrastructures and narratives that shape access, trust and power.

As the information ecosystem becomes increasingly complex, so too must the capacity to read it, respond to it and protect affected populations, individuals and organizations from its harms. Navigating this ecosystem is now a core part of what it means to act in humanitarian crises. It must inform how responses are designed and implemented, while also driving advocacy for broader systemic change.

The responsibility cannot rest solely with humanitarian organizations. Addressing harmful information requires enhancing resilience, building trust and deepening community engagement. It also demands coordinated, multi-stakeholder action – by governments, technology companies, media, communities and civil society organizations.

Confronting harmful information is not optional. It requires nothing less than systemic change. It is essential to protecting lives, upholding humanitarian principles and ensuring that humanity itself remains the strongest counterforce to manipulation, mistrust and division.

The task now is to ensure that the necessary safeguards, engagement, proximity, trust and resilience are not left to chance or profit, but are deliberately harnessed to protect people and enable principled humanitarian action.

Asks, aims and recommendations

Asks

Confront harmful information as a systemic humanitarian crisis that undermines safety, dignity and access. Integrate universally recognized rights, preparedness, accountability and transparency into crisis response, and act with the same urgency as for other humanitarian threats. Strengthen collaboration for sector-wide monitoring through rumour tracking, social listening and sentiment analysis.

Aims

Protect people: Prioritize reliable and accurate information in crisis preparedness, response and recovery to safeguard safety, dignity and access.

Detect early and adapt: Monitor narratives, perceptions and sentiment to enable coordinated, transparent and timely responses.

Safeguard humanitarian action: Shield staff, volunteers, affected populations and operations from harmful information, cyber threats and disruptions to critical infrastructure, while preserving trust and legitimacy.

Recommendations

States and policy-makers

- Integrate harmful information management into crisis preparedness and response frameworks.
- Invest in early-warning and verification systems to deliver timely, reliable, life-saving information.
- Uphold cyber norms and protect humanitarian organizations and critical infrastructure from malicious ICT use.
- Establish clear legal and policy frameworks that support and protect humanitarian action from interference and harmful information campaigns.

Technology platforms

- Prioritize rapid moderation and fact-checking in humanitarian crises – before, during and after crises – with effective escalation channels for humanitarian organizations. Ensure tools function in low-bandwidth, multilingual and resource-constrained contexts.
- Integrate purposeful friction into digital platforms – such as accuracy prompts, warning labels,

‘read-before-you-repost’ nudges, and slower sharing pathways for unverified content – particularly in the context of humanitarian crises, to limit the rapid spread of harmful information and encourage more deliberate, informed user engagement. Adapt algorithms to reduce amplification of harmful narratives targeting humanitarian organizations, principled humanitarian action and affected populations.

- Report transparently on moderation actions, algorithmic adjustments and impacts in humanitarian contexts.
- Co-design crisis-response protocols with humanitarian actors to ensure interventions are timely, context specific and aligned with internationally recognized rights and standards.

Humanitarian actors

- Treat harmful information management as an operational risk – not just a communications challenge – integrated into humanitarian diplomacy and protection dialogue with states, regional organizations and multilateral forums; and into programmes, risk frameworks and preparedness planning.
- Train staff and volunteers in rumour verification, digital literacy and safe information programmes; share insights across trusted organizations.
- Strengthen monitoring by combining AI-enabled tools with

human expertise for real-time detection, mapping of harmful information ecosystems and early warning; foster collaboration across the humanitarian sector to pool resources and capacity.

- Document and analyse harmful information incidents systematically to build an evidence base for policy, adaptation and advocacy.
- Engage communities transparently, co-creating messages with affected communities and reinforcing neutrality, impartiality and independence.

Communities and local leaders

- Develop and sustain local rumour tracking and verification systems.
- Act as trusted intermediaries to strengthen solidarity, resilience and public confidence in humanitarian response.
- Provide feedback loops to humanitarian actors and authorities on trust gaps and unmet concerns.

- Ensure inclusivity by making sure youth, minorities and people with disabilities are represented in community information systems.
- Foster dialogue, peer-to-peer engagement and community-led initiatives to counter harmful narratives and prevent polarization or stigma.

Endnotes

- 1 For example, in 2006, Reuters severed ties with a freelance photographer after discovering digitally manipulated images, most notably smoke plumes, added to photographs related to the Israel–Lebanon conflict: 'Reuters drops photographer over doctored images,' *The Guardian*, 7 August 2006. www.theguardian.com/media/2006/aug/07/reuters.pressandpublishing
- 2 During the Arab Spring, the blog *A Gay Girl in Damascus*, purportedly written by a Syrian-American woman, was revealed to be a fictitious persona created by a male university student in Scotland. See Mackey, R. 'Gay Girl in Damascus' Blog a Hoax, American Soys,' *The New York Times*, 13 June 2011, www.nytimes.com/2011/06/13/world/middleeast/13blogger.html
- 3 In July 2006, during the Israel–Lebanon conflict, two Lebanese Red Cross ambulances were struck by munitions, injuring crew members. Blogger communities claimed the damage was staged due to the appearance of the vehicles. Investigations confirmed the attack was real. See UN Human Rights Council, Report of the Commission of Inquiry on Lebanon pursuant to Human Rights Council resolution S-2/1, A/HRC/3/2, 23 November 2006, para.173, https://digitallibrary.un.org/record/587605/files/A_HRC_3_2-EN
- 4 See, for example, Rid, T. *Active Measures: The Secret History of Disinformation and Political Warfare*. (2020)
- 5 See, for example, West, DM. *Fighting hurricane misinformation requires aggressive pushback*. Brookings. (2024) www.brookings.edu/articles/fighting-hurricane-misinformation-requires-aggressive-pushback
- 6 See World Health Organization. *Infodemic*. www.who.int/health-topics/infodemic
- 7 False claims that 5G mobile phone towers were linked to the spread of COVID-19 led to arson attacks on telecommunications infrastructure in the UK and threats against telecom workers. Scientific authorities consistently debunked any connection between 5G technology and the virus. See Hern, A. 'Arsonists Attack Phone Mast Serving NHS Nightingale Hospital.' *The Guardian*, 14 April 2020. www.theguardian.com/technology/2020/apr/14/arsonists-attack-phone-mast-serving-nhs-nightingale-hospital
- 8 Nichols, T. *The Death of Expertise: The Campaign Against Established Knowledge and Why It Matters*. (2024)
- 9 The 'UN Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action' were officially launched on 24 June 2024, during a press conference led by UN Secretary-General António Guterres. www.un.org/en/information-integrity
- 10 See UN. *Information Integrity on Digital Platforms*. (Our Common Agenda, Policy Brief 8). Executive Office of the Secretary-General. (2023) p.2. https://digitallibrary.un.org/record/4012857/files/^EOSG_2023_8^--EOSG_2023_8-EN.pdf
- 11 UN Department of Global Communications. Issue Brief 1 – *From Principles to Practice: Strengthening Information Integrity*. (2025) www.un.org/sites/un2.un.org/files/un-information-integrity-issue-brief-01.pdf
- 12 IFRC, ICRC and Red Cross Society of Guinea, *Red Cross Red Crescent denounces continued violence against volunteers working to stop spread of Ebola*. Joint statement. 11 February 2015. www.icrc.org/en/document/red-cross-red-crescent-denounces-continued-violence-against-volunteers-working-stop-spread
- 13 See, for example, www.mdpi.com/2673-5172/2/4/34; www.niemanlab.org/2015/09/what-happened-after-7-news-sites-got-rid-of-reader-comments/; www.axios.com/2023/07/06/online-comments-medianews-group-newspapers-end
- 14 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) p.270
- 15 Ayala Iacucci, A. *Misinformation, Disinformation, and Hate Speech in Humanitarian Contexts*. Grand Challenges Canada, Creating Hope in Conflict initiative. (2024) p.38 https://humanitariangrandchallenge.org/wp-content/uploads/2024/11/MDH-Scoping-Study_Full-report.pdf This scoping study provides a comprehensive overview of operational challenges, strategic approaches and innovative opportunities for countering misinformation, disinformation and hate speech in conflict-affected humanitarian settings.
- 16 Borges do Nascimento, J., Pizarro, AB., Almeida, JM. et al. Infodemics and health misinformation: a systematic review of reviews. *Bull World Health Organ*. 2022;100(9), 544–561. doi:10.2471/BLT.21.287654, p.557
- 17 ICRC, *Addressing Harmful Information in Conflict Settings: A Response Framework for Humanitarian Organizations*. (2025) p.8 www.icrc.org/en/publication/addressing-harmful-information-conflict-settings-response-framework-humanitarian
- 18 Ibid.
- 19 UN General Assembly, Disinformation and freedom of opinion and expression during armed conflicts. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. A/77/288, August 2022, para. 59, <https://docs.un.org/en/a/77/288>
- 20 Radford, A., Tanno, S. and Kemp, O. 'At least 290 killed in Air India crash.' CNN. 12 June 2025. <https://edition.cnn.com/world/live-news/ahmedabad-india-plane-crash-06-12-25>
- 21 Silva, A. 'Social media flooded with racist comments after Air India crash, but Indians are speaking out.' *ABC News*. 17 June 2025. www.abc.net.au/news/2025-06-17/air-india-plane-crash-indian-racism-social-media-comments/105420614
- 22 Mellersh, N. Anti-Syrian sentiment grows in Turkey following earthquake. *Infomigrants*. 16 February 2023. www.infomigrants.net/en/post/46835/antisyrain-sentiment-grows-in-turkey-following-earthquake; Ozduzen, O., Korkut, U. and Ozduzen, C. 'Refugees are not welcome': Digital racism, online place-making and the evolving categorization of Syrians in Turkey. *New Media & Society* 2020:23(11), 3349–3369. <https://doi.org/10.1177/1461444820956341>
- 23 Carter, RT. Racism and Psychological and Emotional Injury: Recognizing and Assessing Race-Based Traumatic Stress. *The Counseling Psychologist* 2007:35(1), 13–105. <https://doi.org/10.1177/0011000006292033>; Hankerson, SH., Moise, N. and Wilson, DR. The Intergenerational Impact of Structural Racism and Cumulative Trauma on Depression. *The American Journal of Psychiatry* 2022:179(6), 434–440. <https://doi.org/10.1176/appi.ajp.21101000>; Polanco-Roman, L., DeLapp, RC. and Dackis, MN. Racial/ethnic discrimination and suicide-related risk in a treatment-seeking group of ethnically minoritized adolescents. *Clinical Child Psychology and Psychiatry* 2023:28(4), 1305–1320. <https://doi.org/10.1177/13591045221132682>
- 24 Rodriguez-Díaz, CE. and Lewellen-Williams, C. Race and Racism as Structural Determinants for Emergency and Recovery Response in the Aftermath of Hurricanes Irma and Maria in Puerto Rico. *Health Equity* 2020:4(1), 232–238. <https://doi.org/10.1089/heaq.2019.0103>
- 25 Gatwiri, K. and Anderson, L. Boundaries of Belonging: Theorizing Black African Migrant Experiences in Australia. *International Journal of Environmental Research and Public Health* 2020:18(1), 38. <https://doi.org/10.3390/ijerph18010038>
- 26 Quiñones-Rosado, R. Liberation Psychology and Racism. In Comas-Díaz, L. and Torres Rivera, E. (Eds.), *Liberation Psychology: Theory, Method, Practice, and Social Justice*. American Psychological Association. (2020) pp.53–68 <https://doi.org/10.1037/0000198-004>
- 27 Perrigo, B. Exclusive: New Global Safety Standards Aim to Protect AI's Most Traumatized Workers. *Time*. 19 June 2025. <https://time.com/7295662/ai-workers-safety-rules>

- 28 Bengio, Y, Mindermann, S., Privera, D. et al. *International Scientific Report on the Safety of Advanced AI*. International Scientific Panel on the Safety of Advanced AI. (2025) <https://internationalaisafetyreport.org>
- 29 Ibid
- 30 Ibid, p.62
- 31 Ibid, p.65
- 32 Ibid, p.69
- 33 Ibid
- 34 Ibid, p.69
- 35 UN. *Governing AI for Humanity: Final Report*. (2024) p.75 www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf. The UN High-Level Advisory Body on Artificial Intelligence (HLAB-AI), established by the UN Secretary-General and composed of 32 global experts, aims to align AI development with human rights and the Sustainable Development Goals.
- 36 Ibid, pp.28, 30, 57
- 37 See, for example, Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) p.262
- 38 Ibid
- 39 Ibid, p.160
- 40 See, for example, Slim, H. *Solferino 21: Warfare, Civilians and Humanitarians in the Twenty-First Century*. (2022)
- 41 Access Now. 'Lives on hold: internet shutdowns in 2024'. 24 February 2025. www.accessnow.org/internet-shutdowns-2024; DW. 'Myanmar: Why is junta shutting down internet?' 27 February 2025. www.dw.com/en/myanmar-why-is-junta-shutting-down-internet/a-71770580; Myanmar Internet Project, Access Now, et al. 'Myanmar must lift internet restrictions following devastating earthquake.' Joint statement. 31 March 2025. www.accessnow.org/press-release/call-for-lifting-of-internet-restrictions-myanmar.
- 42 The '5 W's' (Who, What, When, Where, Why) are a standard journalistic framework for ensuring completeness and clarity in reporting; sometimes 'How' is added as a sixth element.
- 43 A: Actor, B: Behaviour, C: Content, D: Degree, E: Effect. Pamment, J. *The EU's Role in Fighting Disinformation: Crafting A Disinformation Framework*. (2020)
- 44 DISARM (*Disinformation Analysis and Risk Management*) is an open-source framework developed by the DISARM Foundation drawing on cybersecurity methodologies to help organizations identify, analyse and coordinate responses to disinformation campaigns through a shared taxonomy of tactics, techniques and procedures. www.disarm.foundation
- 45 MITRE ATT&CK (Adversarial Tactics, Techniques and Common Knowledge) is an open knowledge base developed by the MITRE Foundation that documents real-world adversary tactics and techniques used in cyberattacks. <https://attack.mitre.org>
- 46 LinkAlong Sàrl, a Swiss-based company founded at École Polytechnique Fédérale de Lausanne – one of Europe's leading science and engineering institutions – combines advanced AI with human expertise to detect weak signals, anticipate risks and accelerate insight-driven decisions.
- 47 URL stands for uniform resource locator: the web address typed into a browser to reach a specific page or resource on the internet.
- 48 NewsGuard is a US company employing an international team of journalists to produce information reliability data for consumers, brands and institutions. It maintains a continuously updated, machine-readable database of provably false claims circulating online.



1030
1030

ETHIOPIAN RED CROSS SOCIETY



Chapter 2

Harmful information and the erosion of trust in humanitarian response: The role of truth, trust and technology





Chapter 2

Harmful information and the erosion of trust in humanitarian response: The role of truth, trust and technology



Contents

	Introduction: The shifting ground of trust	71
2.1	Incentivized to hostility	71
2.2	Trust: What is being lost?	72
2.3	Trust in institutions	75
2.4	Integrity, perception and the fragile foundation of trust	80
2.5	Decline of expertise and its impact on trust	83
2.6	Community engagement: A bridge to trust	85
2.7	Facts and feelings: A perception challenge	89
2.8	Responding to harmful information: Building trust in crises	91
2.9	Transparency and identifiability: Foundations for trust	92
	Concluding remarks: Trust, truth and preparedness	94
	Endnotes	98

Introduction: The shifting ground of trust

“To be persuasive we must be believable;
to be believable we must be credible;
to be credible we must be truthful.”

So said Edward R Murrow, the American broadcaster and correspondent during the Second World War. The principle still holds: truth (accuracy and honesty) and credibility (competence, consistency, reliability) remain essential to building institutional trust. Yet in today’s information landscape, applying this principle has become far more complex and contested in an age shaped by harmful information.

In times of crisis or uncertainty – and these are profoundly uncertain times – people increasingly turn to information sources they perceive as relevant and aligned with their personal and lived experience, rather than those grounded solely in factual accuracy. Truth alone no longer always persuades. Emotion, identity and repetition can entrench misbeliefs in powerful, sometimes harmful ways. In such an environment, even reaching agreement on what constitutes a fact is difficult. For humanitarian organizations – whose access, acceptance and ability to operate depend on trust – navigating this fragmented, emotionally charged information space has become not only an operational challenge, but also a security risk.

2.1 Incentivized to hostility

This online dynamic has been described as the ‘filter bubble’,¹ a term used to explain how search engines and social media platforms serve content that algorithms *think* we want to see, based on our searches, likes and clicks. Eli Pariser warned that being confined to such echo chambers reduces exposure to diverse perspectives and increases the risk of isolation in our views or rejection of opposing viewpoints and sources of information due to confirmation bias.²



Confirmation bias

Confirmation bias refers to the tendency to seek out, favour and recall information that supports our existing beliefs, while ignoring or dismissing contradictory evidence.

Although some analyses show that users do encounter opposing views online, these interactions often trigger annoyance or hostility rather than reflection or debate.³ Algorithms reinforce this by amplifying divisive content, since hostility drives

engagement and engagement drives profits for platform companies. As Pariser notes, platforms are effectively “incentivizing us to fight with each other when we are online”. This dynamic is compounded by the **online disinhibition effect**, whereby people tend to express opinions more freely online due to factors like anonymity (hidden identities), invisibility (not being seen by others we communicate with) and asynchronicity (not engaging in real time).⁴

These forces contribute to an erosion of shared reality. Increasingly, there are claims that the impact of disasters are inflated, that scientifically validated medical treatments are unsafe or that documented atrocities are fabricated or exaggerated. This reflects what many describe as a ‘post-trust’ world – one in which people are especially vulnerable to harmful information and deepening polarization over what is considered true or false. Such polarization increases the risk of social unrest, violence and even armed conflict. There is also a secondary level of manipulation in today’s post-trust world, where institutions themselves come under scrutiny and their legitimacy is increasingly contested. In some contexts, polarization between opposing political, social or geopolitical dynamics has created conditions in which public authorities may challenge or overturn established processes or outcomes, despite the absence of clear evidence of interference or wrongdoing. The erosion of institutional credibility poses profound questions for public trust. In turn, as humanitarian actors mandated by public authorities, National Red Cross and Red Crescent Societies face heightened reputational risks if confidence in governance deteriorates.

The internet makes it easier than ever to find information that confirms existing beliefs – a phenomenon sometimes referred to as the ‘Google delusion’. Digital spaces amplify confirmation bias by helping users to find like-minded communities and cite information that reinforces their views while disregarding contradictory information⁵ or expert advice. While the internet offers access to vast knowledge, it also accelerates the spread of harmful information, encourages argument over dialogue and conspiracy over evidence.⁶

Researchers Kavanagh and Rich (2018) have described this phenomenon as *truth decay*, defined by a set of four related trends: (1) increasing disagreement about facts and analytical interpretations of facts and data; (2) a blurring of the line between opinion and fact; (3) the increasing relative volume, and resulting influence, of opinion and personal experience over fact; and (4) declining trust in formerly respected sources of factual information.⁷ Instead of using facts to inform our beliefs, information – regardless of its truthfulness – is increasingly used to justify the beliefs already held by individuals and the groups they affiliate with. As they conclude: “We’re no longer willing to agree on something as seemingly fundamental as what counts as evidence, facts, or truth anymore”.⁸

2.2 Trust: What is being lost?

Trust is grounded in expectations, involves vulnerability and builds gradually, yet once broken may be lost completely. Rousseau et al. (1998) define trust as “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviour of another”.⁹ This highlights a key dimension: trust requires accepting vulnerability to those in whom it is placed.

People often trust what aligns with their worldview or community, rather than what is factually accurate. Harmful information campaigns exploit this tendency, distorting truth and diminishing the visibility and influence of credible sources. In contexts where independent media and public trust in institutions remain strong, Edward R Murrow's notion of truth-based credibility still underpins effective communication. But where information is instrumentalized, persuasiveness often relies less on truth and more on emotional resonance, social identity or perceived authority.



I think the major thing that people lack is they should take time to verify information from the correct sources because in most cases what people see or just hear is what they end up going to redistribute to other people.”

Community member, Zambia

Public trust has also been eroded by the widespread perception that governments largely failed to manage crises such as the COVID-19 pandemic and climate change.¹⁰ This is especially true for communities that are marginalized and at risk, who often view authorities less as providers of needed services, and more as sources of demands. Ironically, these are the very people most at risk during crises, yet their lived experiences of exclusion often make them more likely to distrust those offering help.¹¹

As psychologist Joe Pierre observes, vaccine hesitancy driven by misinformation “may very well end up being more lethal than any other misbelief in our lifetimes”.¹² Although social media platforms took steps to revise their algorithms to deprioritize some medical misinformation, the damage was already significant.¹³

The IFRC's 2022 *World Disasters Report: Trust, Equity and Local Action – Lessons from the COVID-19 pandemic to avert the next global crisis* highlights that multiple studies suggest that “trust was one of the best predictors of a country's success or failure in handling the COVID-19 pandemic”.¹⁴ Many communities had extremely limited access to health education and information.¹⁵ Information from social media platforms or state news outlets was often distrusted, and many people did not understand the rationale behind public health measures. In several countries, governments issued contradictory advice and oversimplified or false beliefs took hold – for example, claims that COVID-19 was only dangerous to older and clinically vulnerable people¹⁶ or that “COVID-19 does not kill black people”.¹⁷ These narratives, compounded by vaccine-related disinformation (such as sterility or microchip claims) and entrenched anti-vaccine beliefs, drove harmful behaviours including vaccine refusal and avoidance of life-saving health measures.

Pierre defines beliefs as “cognitive representations of past, present, and future reality, encompassing our inner experiences, the world around us, and the world beyond.”¹⁸ He models beliefs as probability judgements, emphasizing that many of our beliefs are held with excessive conviction, often at the expense of acknowledging appropriate levels of uncertainty. People often tend to adopt an absolute ‘belief that’ stance, even on matters that would more appropriately warrant a probabilistic ‘belief in’ that acknowledges uncertainty and complexity. Deciding who to trust or mistrust often comes down to assessments of source credibility – based on perceived trustworthiness and expertise. When these perceptions are distorted, the very foundation of humanitarian and public health response is at risk.

Contributor Insight 2.1



Trust, misinformation and the power of local connection in crisis response

In today's increasingly connected world, misinformation, disinformation and harmful speech pose serious threats to humanitarian access, public health and social cohesion. Understanding how these dynamics take root and evolve is essential for designing effective, community-led programmes and responses.

A powerful example emerged during the COVID-19 pandemic, involving the work of several National Red Cross and Red Crescent Societies with both Indigenous and last mile communities across Latin America. In many remote areas, communities resisted public health interventions due to deep mistrust of authorities and cultural understandings of disease that diverge from medical models. In Peru, for instance, health workers were labelled as 'white devils' and accused of bringing the virus into communities. These beliefs were not simply a result of misinformation – they reflected traditional conceptions of illness and were exacerbated by limited access to accurate, culturally relevant information. Often, public health messaging was shared in a single language and failed to reflect the communities' own perspectives on health and healing. In this vacuum, misinformation spread rapidly by word of mouth and was reinforced by local leaders.

To address these challenges, health authorities turned to the National Society – fostering dialogue to uncover and bridge differences in perceptions of illness and treatment. By engaging trusted religious and community leaders, they were able to counter harmful rumours and promote the value of vaccination in a way that honoured local beliefs and customs.

This experience illustrates that building trust requires more than simply transmitting information. It requires a sincere commitment to listening, understanding cultural perspectives, identifying shared values and communicating transparently throughout the response. Relying on generic messaging or assumptions risks alienating the very communities we aim to support and ignoring the complex realities they face.

One of the persistent challenges for National Societies lies in how they are perceived. Their auxiliary role is often poorly communicated or understood, leading to confusion and mistrust. In politically polarized contexts, a National Society may be viewed either as an extension of the government or as an opposing force – and sometimes both simultaneously. These contradictory perceptions can place volunteers at risk and hinder humanitarian access.

To effectively respond to these challenges, National Societies need targeted training and practical tools to monitor and respond to harmful information. Clear guidelines, shared terminology and stronger coordination within the Movement and with external partners are essential. Documented evidence of the impacts of harmful information on access and safety can also help guide effective mitigation strategies.

Ultimately, the most sustainable solutions begin at the community level. Empowering people through digital and information literacy, combined with culturally sensitive engagement, builds the foundation for trust, critical thinking and informed decision-making.

Monitoring community trust and reinforcing transparency can help strengthen the relationship between National Societies and the communities they serve.

Diana Medina

Asia Pacific Regional Coordinator, Community Engagement and Accountability (CEA)

IFRC

2.3 Trust in institutions

Trust is critical to the legitimacy, effectiveness and acceptance of humanitarian action. This was strongly emphasized at the 2019 International Conference of the Red Cross and Red Crescent Movement which recognized that trust in principled humanitarian action is indispensable to serving at-risk people and encouraged all members of the conference to act to preserve and develop this trust.¹⁹

The decision to place trust on the agenda reflected broader global concerns: the erosion of trust in institutions and governments, growing public scrutiny, and rising demands for accountability. Where there is limited understanding of the Movement's rules and regulations, reputational risks grow and trust declines. As the Commission report²⁰ stressed: "Trust is the most critical currency for the future of humanitarian action, and one which stems from humility and being truthful and transparent." It identified three key priorities:

- 1 Community engagement and accountability
- 2 A conducive environment for principled humanitarian action
- 3 Integrity and risk sharing.

While harmful information was not yet identified as a central concern, the focus on trust, truth and transparency reflected growing recognition that both the Movement and states need to do more to earn trust. It also underscored a broader challenge: preserving the space for principled humanitarian action – something the Movement cannot achieve alone.²¹ This was reaffirmed in the 2024 Council of Delegates resolution on "respect and support for principled humanitarian action"²² which recalls collective Movement commitments to strengthen integrity, accountability and trust.

Crucially, trust in humanitarian action does not rest solely on humanitarian actors. It also depends on a conducive environment – one in which principled, effective and accountable action is actively supported. Such an environment is shaped by the legal, political and operational frameworks established by states and other actors. This includes respect for humanitarian principles, flexible and needs-based funding, protection of humanitarian access and clear delineation of roles and responsibilities of various actors. As highlighted in the Commission discussions in 2019, states play a decisive role: their policies and practices can either foster trust and enable principled humanitarian action or hinder it. States also have a responsibility to support and facilitate the work of the National Society in their country, in their auxiliary role to public authorities. Creating such an environment requires sustained dialogue, shared responsibility and

a commitment to removing obstacles that undermine trust. Without these conditions, even well-intentioned humanitarian efforts risk being delayed, politicized or perceived as partial – ultimately eroding the very trust they aim to build. By the time of the 34th International Conference four years later, harmful information was recognized as part of this deeper crisis of trust.

Erosion of trust in non-governmental organizations (NGOs) has been documented in the annual Edelman Trust Barometer, which surveys trust in institutions across 28 countries. An analysis of 25 years of survey data provides valuable insights for humanitarian actors navigating this fragile landscape of shifting perceptions of trust in NGOs.

Contributor Insight 2.2

Earning trust in an age of accountability

For over a decade, NGOs have occupied a unique position in the global trust landscape.

According to Edelman Trust Barometer data, NGOs were the most trusted institution across 21 markets from 2012 to 2018, only to be matched by business in 2019. In 2021, however, NGOs were overtaken by business which remains the most trusted institution in 2025. Nonetheless, trust in NGOs remains strong, consistently ranking higher than both government and the media.

The public's resilient trust in NGOs may stem from perceptions of moral integrity. Since Edelman began measuring ethics and competence scores in 2020, NGOs have consistently been viewed as the most ethical institution. At the same time, the public is less confident when it comes to an NGO's ability to deliver results. As of 2025, NGOs rank 15 points below business on competence.

This duality – high ethical standing but perceived operational softness – presents a central challenge: while people believe NGOs are doing the right things, the public doesn't show the same confidence when it comes to their efficacy.

A similar tension appears in views on treating everyone equally and fairly. Since 2020, a global majority of those surveyed agree that NGOs are effective agents of positive change, yet only a minority believe NGOs serve the interests of everyone equally and fairly. While the public recognizes that NGOs are doing good work, this finding demonstrates concern that the work may not be distributed evenly across society.

Such perceived limitations may help explain why trust has shifted toward business, which is seen as more capable of delivering results. Business has also improved its ethical standing, positioned as the institution better equipped to drive impact at scale.

For NGOs, maintaining strong public trust for over a decade is a testament to the mission-driven ethos, but in an era of rising stakeholder expectations this must be reinforced through measurable, equitable impact.

To strengthen trust, NGOs can:

- Show results clearly. Share not just what the organization is doing, but what's working and what impact it's having.
- Partner with others. Work closely with businesses, governments, and local groups to combine strengths and reach more people.
- Build trust locally. Support local leaders and listen to feedback from communities to make sure efforts align with real needs.
- Tell better stories. Focus on sharing stories of tangible success and impact, not just values.
- Be flexible and try new things. Keep improving by learning, adapting and testing new ways to solve problems.

Edelman Trust Institute

Organisation for Economic Co-operation and Development (OECD) research highlights that trust in public institutions rests on two core pillars: **competence** – the ability to meet expectations through operational efficiency, capacity and sound judgements – and **intentions** – acting in good faith, guided by values and principles. Trust is dynamic: individuals form and adjust their trust based on personal experiences and the information available to them. By contrast, distrust often arises not from actual performance but from perceptions of bias, cynicism, disengagement or expectations of betrayal.²³

The effectiveness and even the feasibility of humanitarian action depends on trust across a broad range of stakeholders, including affected populations, governments, non-state actors, donors and technology companies. Without trust, humanitarian action risks losing the access, support and legitimacy it critically needs.

Contributor Insight 2.3



Addressing harmful information: Lebanese Red Cross experience in humanitarian crises

During Lebanon's recent period of overlapping crises, the Lebanese Red Cross, with over 12,000 volunteers nationwide, faced significant challenges to provide life-saving services due to the spread of harmful information. At the height of the COVID-19 pandemic, false claims circulated that Lebanese Red Cross volunteers were spreading the virus. Following the 2020 Beirut Port explosion, disinformation suggested that aid distribution was biased. Similarly, rumours falsely alleged that the National Society was selling blood units. During the Cholera outbreak, misleading narratives claimed that vaccines and oral rehydration

salts were dangerous. These narratives both endangered the volunteers and staff and weakened public trust.

Harmful information disrupted humanitarian operations and disproportionately affected vulnerable groups including older people in remote areas, migrants and displaced communities, as well as Lebanese Red Cross volunteers. Many had limited access to verified digital information or were disconnected from official communication channels.

In response, the National Society drew on its long-standing reputation and deployed targeted communication strategies. During the COVID-19 pandemic response, it launched a non-emergency hotline to address public questions and during the Cholera outbreak used in-person awareness sessions to counter rumours. Tools such as community rumour tracking, hotline feedback, timely press releases and real-time social media presence proved essential. However, gaps in digital literacy – particularly in remote areas – hindered response effectiveness.

The Lebanese Red Cross' impartiality was also questioned in some regions, particularly in areas hosting refugees. When rumours emerged that it favoured refugees in aid distributions, the organization responded swiftly with factual clarifications to reaffirm its impartiality and neutrality.

Ultimately, building resilience to harmful information requires more than reactive communication. The Lebanese Red Cross' recognized status – from the highest authorities to local actors – as a neutral and trusted humanitarian partner was key to maintaining public confidence and operational access. The core humanitarian principles of impartiality, neutrality and independence are increasingly challenged by the spread of false narratives. While communication guidelines are in place, the National Society continues to strengthen and refine specific safeguards to counter harmful content. Above all, its volunteers – trusted and embedded in local communities – remain the organization's most effective asset in upholding community trust and ensuring the continuity of principled, life-saving humanitarian response.

Georges Kettaneh

Secretary General

Lebanese Red Cross

Contributor Insight 2.4

Canadian Red Cross: Moving beyond brand; measuring trust in the face of harmful information

The Canadian Red Cross conducts annual survey-based research to better understand what drives trust among Canadians and how that trust connects to the organization's humanitarian mission and work. This research was designed to move beyond traditional

brand strength measurement tools such as familiarity, attitude and remarkability, to gain deeper insights into the factors that sustain trust.

In the face of rising levels of harmful information, maintaining high levels of trust is crucial for the Canadian Red Cross to serve as a credible source of information at both the community and partnership levels. In recent years, trust levels have declined across all sectors in Canada, consistent with global trends linked to the rise of harmful information online and the weakening of traditional broadcast and print media.

2025 marked the fourth consecutive year of the research, which also includes pulse surveys following significant issues and emergency responses.

Results have consistently shown that among peer organizations and across sectors, the Canadian Red Cross leads in trust. This strong reputation positions the organization as a trusted voice in times of crisis at the community level and as a partner of choice within civil society and with the government.

Trust research measures three core areas:

- 1 Ability (Competence) – “I can rely on the Canadian Red Cross to operate competently and effectively.”

The collective knowledge, skills and competencies that enable the Canadian Red Cross to deliver services and meet its goals and responsibilities.

- 2 Benevolence (Goodwill) – “I believe the Canadian Red Cross cares about its stakeholders.”

The degree to which the Canadian Red Cross demonstrates a duty of care and humanity towards all stakeholders.

- 3 Integrity (Character) – “I can rely on the Canadian Red Cross to do the right thing.”

The extent to which the Canadian Red Cross consistently adheres to widely accepted ethical principles and moral values.

Understanding what drives trust and how to strengthen and maintain it is now integrated into the Canadian Red Cross’s strategic approach, from the governance level (Board of Directors) through to executive leadership teams in Operations, Stakeholder Relations, Philanthropy and Communications.

Nathan Huculak

Chief Communications Officer

Canadian Red Cross

2.4 Integrity, perception and the fragile foundation of trust

Cases of misconduct such as abuse, exploitation, fraud or mismanagement have severely eroded public trust, especially when humanitarian organizations respond without transparency or empathy. Scandals involving sexual abuse in organizations have demonstrated how quickly confidence can collapse.²⁴ The politicization of aid – when governments, armed actors or other parties manipulate humanitarian action for political purposes – further undermines perceptions of neutrality and leaves communities sceptical of humanitarian motives.

In today's landscape of digital transparency and constant scrutiny, humanitarian actors are no longer the sole or even primary narrators of their work. They now compete for legitimacy not only with governments and non-state actors but also with communities themselves, who increasingly speak, organize and question responses in real time.

Addressing integrity issues is therefore essential. Allegations of partiality, corruption or mismanagement can deflect attention away from life-saving work and damage trust both internally within humanitarian organizations and externally with the public. The overall decline of trust in institutions is reflected in heightened scrutiny of the integrity of National Societies and other Movement components. This demands both proactive promotion of mandates, principles and activities, and robust, strategic approaches to issues management, reputational challenges and related risks.

At the same time, humanitarian organizations face unprecedented demands for reporting, compliance and proof of impact. Failures must be acknowledged and addressed openly and transparently. In parallel, there must also be concerted efforts across sectors to counter the intentional spread of harmful information, which is both unethical and dangerous.

Finally, a lack of, or slow progress in, localizing humanitarian aid has further fuelled mistrust. Many communities perceive humanitarian action as top-down and dominated by international actors with limited local representation or contextual understanding. In some contexts, the sector is seen as Western-dominated, a perception reinforced by standards, codes of conduct and coordination mechanisms developed by organizations from the Global North.

Contributor Insight 2.5



Costa Rican Red Cross: Strengthening emergency operations through verified information flows

Costa Rica is a small country with an extraordinary geographic diversity. Within a territory of just over 51,000 square kilometres, it brings together mountain ranges, active volcanoes, fertile valleys, tropical forests, extensive coastlines on both the Pacific Ocean and the Caribbean Sea, and a vast maritime territory that exceeds its land area. This natural richness shapes the livelihoods, culture and identity of its people, and positions Costa Rica as a globally recognized leader in environmental conservation and sustainable development.

At the same time, this diverse geography and exposure to dynamic climatic patterns place the country in a region where natural hazards are a recurring reality. Atmospheric and geographic factors converge to generate extreme weather events such as floods, storms and droughts, which recurrently trigger emergencies and affect communities across the territory.

According to the National Institute of Statistics and Censuses, 85.4% of households in Costa Rica have internet access. Widespread access to digital networks means that Costa Ricans benefit from improved education opportunities and enhanced disaster risk management tools, such as early warning systems. However, it has also led to the rapid diffusion of false information online, including during emergencies, where malicious actors exploit public fear and uncertainty.

The most frequent challenges occur when inaccurate information is circulated through messaging platforms. These messages are often designed to go viral, generate alarm and, in turn, sow doubt in the public about the capacity of the National Risk Management System to address threats. They may also attempt to confuse first responders and disrupt the allocation of resources during humanitarian response efforts.

The Costa Rican Red Cross recognizes that strengthening disaster management capacities is essential to address challenges such as false information in communities, which has a direct influence on operational decision-making and resources for humanitarian response.

Since 2023, the National Society has implemented a continuous improvement process within its Emergency Operations Centre, focusing on four key areas: process optimization; updated Disaster Response Plan; capacity building (with some 50 people trained, including in information management); and strengthened national coordination.

The centre now emphasizes information and situation analysis, processing data from its reception through verification, filtering and interpretation to the generation of actionable outputs. Key products include:

- 1 Operational decision-making: Filtered, verified and analysed information enables timely, accurate decisions for the allocation of critical resources, minimizing the risk of using these on situations that do not need a humanitarian response.

- 2 Information dissemination strategies and products:
 - a Increased visibility of the Costa Rican Red Cross across television, radio and social media, using trained spokespersons to provide reliable information, counter harmful content and promote accurate guidance. This includes reinforcing the importance of seeking information from official national and community channels aligned with the National Risk Management System.
 - b Statistical reporting: Data-driven infographics shared through institutional channels keep the public and institutions informed about humanitarian responses.
 - c Public consultation channels: The National Society contact lines allow people to access accurate information and provide feedback.
- 3 Reports and briefings: Situation reports, data platform updates and information on financing, such as DREF, are built on reliable, timely data, reducing errors.

Combating misinformation as anticipatory action

The Costa Rican Red Cross implements early action plans (EAPs) under DREF funding, with a strong focus on managing accurate information, for example, related to volcanic ash dispersion. Costa Rica has over 400 volcanic structures, three of which – Rincón de la Vieja, Poás and Turrialba – are classified by the Volcanological and Seismological Observatory of Costa Rica at activity level 2 ('warning'), placing nearby communities and their livelihoods at risk from potential eruptions and associated impacts.

Increased volcanic activity can affect communities' health and livelihoods, and myths or false information can amplify these impacts. To address this, early action involves providing timely, accurate information to exposed communities to prevent health risks, livelihood losses and guide risk mitigation measures. This includes:

- Digital communication shared by the Costa Rican Red Cross on its social media accounts, developed in collaboration with technical-scientific institutions.
- Media engagement to explain EAP measures and current volcanic activity.
- Community outreach through in-person engagement and accountability approaches – such as focus groups, talks and Q&A sessions – in coordination with local authorities.

Success stories

- 1 **Flood emergencies in November 2024.** During the national emergency caused by severe rains, floods and landslides, the National Society implemented a communication strategy centred on transparency, prevention and timely access to verified information. In total, 92 official updates were issued through different channels, and daily media engagement helped to inform communities about the humanitarian response and to

disseminate key prevention and safety messages, helping to counter rumours and misinformation.

- 2 Volcanic activity, March 2025.** When Poás volcano activity significantly increased and the alert level was raised to level 3 ('precaution'), the EAP for volcanic ash dispersion was activated as planned. This early communication provided communities with actionable guidance to protect their health and livelihoods, successfully replacing misinformation with scientific, timely information.

Costa Rican Red Cross

The current humanitarian 'reset' or 'renewal' – driven by the unprecedented reduction in humanitarian funding in 2025 and beyond – and resulting in widespread restructuring or programme cuts – is amplifying harmful information. Sudden programme closures, staff layoffs and uncertainty – often communicated inadequately to affected communities – creates a void that harmful information, distrust and political narratives quickly fill. When communities are not meaningfully engaged or when changes are poorly explained, speculation arises that aid is being politicized, diverted or withdrawn entirely. Critics and bad-faith actors exploit this uncertainty to undermine trust in humanitarian action. Without clear, transparent communication and visible inclusion of local voices, the reset or renewal risks becoming fertile ground for suspicion, polarization and harmful narratives.

2.5 Decline of expertise and its impact on trust

The erosion of trust in expert knowledge, described as the "death of expertise" (Nichols), has deepened public scepticism toward authoritative information. The internet's promise of open access and free expression has blurred the line between reliable information and harmful information, enabling both to coexist: one guiding us and the other misleading us.²⁵ **Emotion often shapes belief before evidence is even considered.** While questioning claims and evaluating sources is essential, it increasingly occurs outside traditional frameworks of expertise.

Humanitarian organizations are generally perceived as expert actors, valued for their technical skills, operational experience and legal and policy knowledge. Yet this legitimacy is not fixed: it depends on context, conduct and how well organizations meet growing expectations around localization, accountability and effective response.

Meanwhile, the information landscape has shifted dramatically. Social media has empowered communities to challenge traditional gatekeepers of knowledge and narrative, elevating peer voices over institutional or credentialed experts. As outlined in *The Death of Expertise*: "Facts and reason are under siege on multiple fronts".²⁶ Notably, Edelman's 2006 Trust Barometer marked the first time that 'a person like me' emerged as one of the most credible sources of information.

In this environment, expertise alone is no longer enough. Trust must be continuously earned through transparency, humility and responsiveness. This shift raises critical questions explored in the next section: Where is trust in humanitarian organizations genuinely misplaced – requiring constructive criticism, evaluation and open dialogue? Where is distrust deliberately manufactured to erode trust and legitimacy? And how can we tell the difference?

Contributor Insight 2.6

Designing trust

Does trust matter when you are hungry? When you are focused on being efficient? When your information is backed by science?

Yes, it matters, because aid delivered with respect for people's full capacity builds something that can last beyond funding cycles. It becomes embedded in communities, generates local solutions and avoids antagonism that could increase risks, not reduce them.

Too often, conversations about trust in the humanitarian sector focus on external threats – the circumstances outside the control of humanitarian organizations: malicious actors, polarizing algorithms, shrinking aid budgets or shifting societal norms. While it is good to understand these changes and challenges, much of what builds – or erodes trust – lies within our control. We can do a lot ourselves in the sector to warrant trust.

Trust is relational, conditional and multidimensional. This is why trust can sometimes feel too abstract or too intangible to be actionable. The Trust Framework tries to overcome this. It was developed under the [Rooted In Trust](#) project through consultations in more than ten different crisis contexts. It breaks trust down into four core elements – accuracy, proximity, intention and control – each with three sub-elements that shape how people perceive the organizations around them. This framework can reveal disconnects we might otherwise overlook, for instance, that while our information is factually correct, it may come across as too polished and out of touch with the current conversations. It might reveal that people perceive local staff and local partners as 'them' and not 'us', as they are not seen as independent enough from the international aid system they are associated with.

Designing programming around trust is not about ignoring risks. It is more like a parent watching their child cycle alone to school for the first time: fully aware of the risks but choosing to let them go on their own, so the child can grow in independence. Trust requires a level of independence, space not control, so people can contribute meaningfully, signal gaps, adapt and change course when needed. Because with agency comes responsibility. With trust, far more becomes possible.

Stijn Aelbers

Humanitarian Consultant

2.6 Community engagement: A bridge to trust

“So then, the advantage we had ... is that when we were finally able to go in to carry out an assessment and have contact with the community leaders, we would compare the information coming out in the press or the information sometimes given to us by the municipalities with what the people in the community themselves told us. And without a doubt, we went with what they said, because there were huge differences in terms of losses, impact, destruction, displaced families, etc., compared to what was being reported. ... that’s why we always made a point of seeking information from the people who were suffering at that moment, the ones who were actually facing the emergency.”

Community member, Bolivia

Community engagement and accountability (CEA) is a vital bridge to building and sustaining trust. Trust grows through proximity, inclusive participation, timely and transparent communication and shared decision-making with people and communities. CEA also ensures that communities have access to accurate, relevant and potentially life-saving information – making it essential not only for effective humanitarian response but also for the safety and security of staff and operations.²⁷



Movement-wide commitments for community engagement and accountability²⁸

Community engagement refers to ways of working collaboratively with people and communities to ensure that Red Cross and Red Crescent actions are effective, inclusive, sustainable and accountable, and that they contribute to supporting and enabling people and communities to lead and shape positive, sustainable changes in their own lives and on their own terms. This includes processes to systematically listen to, engage and communicate with people and communities to better understand their diverse needs, vulnerabilities and capacities; to gather, respond to and act on feedback and input about their priorities and preferences; and to provide safe and equitable access and opportunities to actively participate in decisions that affect them.

This also includes the responsibility to communicate transparently and in an appropriate, accessible manner about Movement principles and values, the aims and objectives of Movement action and advocacy, what people and communities can expect from us, and how they can participate in and provide their inputs about issues and decisions that affect them.



Community engagement played a pivotal role in rebuilding trust during the 2018 Ebola outbreak in the Democratic Republic of the Congo (DRC). Fear, rumours and misinformation fuelled mistrust, leading at times to violence against Red Cross teams, particularly those conducting safe and dignified burials. This resistance blocked access to life-saving care and may have contributed to further transmission of the virus in vulnerable communities.

To respond, the DRC Red Cross placed community engagement at the centre of its strategy. Volunteers maintained daily dialogue with communities to address concerns and rebuild trust. Using a first-of-its-kind feedback system, they collected over 394,000 pieces of community feedback, enabling humanitarian actors to adapt approaches to local realities. Volunteers went door to door sharing information on how to prevent, detect and respond to Ebola, hosted radio talk shows, organized mobile cinemas and conducted outreach to vulnerable groups, including people with disabilities, children and women's associations. This approach had a measurable impact: the success rate of safe and dignified burials remained consistently high at 80%. Community resistance to safe and dignified burials dropped drastically from 79% in the first two months of the operation to just 8% in September 2019. Trust, dialogue and locally driven action proved essential to stopping the spread of Ebola.²⁹



Yemen Red Crescent Society volunteers – trusted by the communities they serve

Through consistent and transparent external communication, the Yemen Red Crescent Society has established itself as a trusted organization within Yemeni communities. With a strong volunteer base across its 22 branches, it relies on individuals who are directly connected to the communities they serve. Typically, one male and one female volunteer are responsible for every 20 households, fostering gender-based engagement and local familiarity.

These volunteers play a key role in selecting new staff and volunteers, helping ensure that individuals' political, tribal or religious backgrounds do not interfere with their ability to work effectively with community leaders and members. This careful and inclusive selection process helps maintain harmonious working relationships and reinforces the community's trust in the Yemen Red Crescent Society. By adopting community-based programme approaches, the National Society ensures its interventions are culturally relevant and locally accepted – critical to effective disaster risk reduction and other humanitarian initiatives.³⁰

Trust is difficult to establish and even harder to rebuild once lost. The IFRC identified trust as one of the strongest predictors of a successful emergency response, making preparedness fundamentally dependent on cultivating trust within communities and societies. To support this, the IFRC developed the **Community Trust Index** – an evidence-based tool to measure and strengthen trust between National Red Cross and Red Crescent Societies and the communities they serve. The index tracks changes in trust over time and identifies the factors that influence it across different contexts. By uncovering barriers to trust, it guides tailored activities that promote trust-based behaviours, improve community acceptance and enhance programme effectiveness. It also links trust to compliance, governance and inclusive, community-informed decision-making, reinforcing accountability and impact.

Contributor Insight 2.7

Trust in local organizations endures amid a shifting information landscape

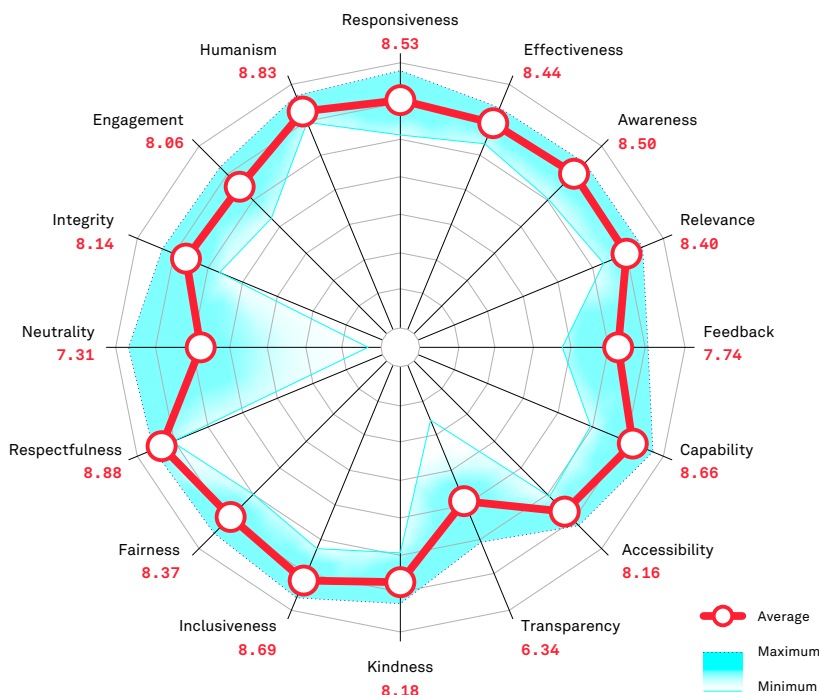
An analysis of **Community Trust Index** data from seven countries, collected between 2022 and 2024, provides important insights into how trust in National Red Cross and Red Crescent Societies is evolving in a world increasingly shaped by misinformation and disinformation. The Index measures two key dimensions of trust: **competence** – reflecting perceptions of an organization’s effectiveness, skills and capacity; and **values** – encompassing ethics, integrity and alignment with community interests.

Overall, communities report high levels of trust in their National Societies, with an average score of 8.15 out of 10. However, a closer look reveals emerging ethical concerns, particularly around transparency, neutrality and openness to community feedback.

Transparency is emerging as a significant challenge. In nearly all contexts studied (Figure 2.1), it ranked as the weakest driver of trust, with an average score of just 6.34 out of 10. Many communities view humanitarian actors as reluctant to openly share mistakes or acknowledge shortcomings – an impression amplified by the digital era’s demand for real-time accountability. This trend is particularly pronounced in Mongolia, Ecuador and Zimbabwe, where gaps in transparent communication may create opportunities for misinformation to spread.

Fig 2.1

Cross-country comparison of the Community Trust Index across all measured sub-dimensions in seven countries



At the same time, perceptions of neutrality are increasingly fragile in today's polarized and politically charged environments. Neutrality ranked as the second weakest driver of trust, with an average score of 7.31 out of 10. In contexts where humanitarian organizations were once broadly regarded as impartial, these assumptions are now being challenged. For example, in Zimbabwe, neutrality scores dropped sharply to 2.2 out of 10 among the general population and just 1.03 among aid recipients. In such settings, the principle of neutrality risks being undermined by contested narratives and politicized messaging, threatening the very foundation of trust.

Compounding these concerns is a limited openness to community feedback. While essential for correcting misinformation and fostering dialogue, feedback ranked as the third weakest driver of trust, with an average score of 7.74 out of 10 – indicating that many communities feel discomfort or hesitation in voicing concerns or complaints. This lack of effective feedback mechanisms hampers the ability of humanitarian actors to respond and adapt to local needs. Without clear, accessible and inclusive channels for engagement, trust becomes more vulnerable to erosion, rumours can take hold and humanitarian organizations risk being perceived as disconnected or unaccountable.

Together, these findings underscore that in today's complex information landscape, delivering aid alone is not enough. To maintain and strengthen trust, humanitarian organizations must commit to transparent, neutral and responsive engagement that meets community expectations and builds resilience against misinformation.

Youth and the trust gap: a growing disconnection

Younger respondents (ages 18–30) consistently expressed lower trust in National Societies than older generations. While this may reflect broader generational shifts in how institutional legitimacy is perceived, it also signals a crucial vulnerability: if humanitarian actors are not perceived as trustworthy by youth – who are often more active in digital spaces – misinformation is more likely to fill the vacuum. Younger audiences may be particularly sensitive to perceived hypocrisy, opaque communication or misaligned values and less forgiving of mistakes that go unacknowledged.

The role of engagement in countering distrust

One of the strongest protective factors against erosion of trust is direct engagement. Communities with established relationships with National Societies reported trust levels somewhat higher (8.5 versus 7.8) than those without such connections. This reinforces the view that localized, participatory engagement serves as a buffer against the destabilizing effects of harmful information. By contrast, the absence of dialogue or proximity can foster suspicion – leaving humanitarian access, neutrality and security more exposed.

Gefra Fulane

Research Coordinator, CEA

IFRC, Geneva

2.7 Facts and feelings: A perception challenge

In an era of increasing volumes of harmful information and rising public distrust, humanitarian organizations face a growing challenge: principled, fact-based communication often struggles to compete with emotionally charged narratives, polarizing content and opinion-driven discourse. Even when – perhaps especially when – grounded in evidence and neutrality, humanitarian messaging is often drowned out by louder, more emotionally resonant voices, particularly online. This creates a significant perception challenge. Trust and integrity may be questioned and principles seen as detached from realities or lacking empathy.

The result is a trust and perception problem: neutrality, facts and rationality are increasingly misunderstood or mistrusted. When emotion and polarization drive engagement, even well-intentioned communications risk being reframed as political or partial. This distorts how communities perceive humanitarian work and undermines the credibility of those committed to principled humanitarian action. Navigating this landscape requires sustained community engagement, especially offline, where dialogue can foster proximity and trust. Online, it may also involve recognizing the limits of engagement and, in some cases, closing comment sections or refraining from responding when dialogue cannot be conducted in good faith.

Among the seven Fundamental Principles of the International Red Cross and Red Crescent Movement,³¹ neutrality and impartiality are often the least understood and provoke the most controversy. This is somewhat ironic, given that neutrality's purpose is precisely to avoid entanglement in controversy. At its core, neutrality is not an end in itself, but a method to preserve trust and access in polarized or politicized environments.

Reconciling how to carry out principled humanitarian action lies at the heart of what humanitarianism is – and is not. Critiques of neutrality often argue that “in situations of injustice, neutrality only helps the oppressor, never the victim.”³² Such critiques serve as warnings against moral or political indifference and silence or passiveness in the face of injustice, violence, oppression or discrimination. At times, accusations of silence or complicity are directed at components of the Movement, especially during armed conflicts. Failing to speak can leave an information vacuum that others may fill, eroding trust. Yet speaking out can be equally fraught, as statements may be perceived as ‘taking sides’ – especially in limited, emotionally charged online comment sections. In this way, humanitarian principles are pulled into political narratives that can distort their purpose. See [Chapter 7, on page 257](#) for more on the fundamental principles, including dilemmas such as equivalence.

Neutrality is often defined by what it is not. Overlooked is that neutrality is not neutral toward suffering. It does not mean moral indifference or denial of injustice. Rather, it means refraining from taking sides in hostilities or engaging in political, racial, religious or ideological controversies. This discipline can enable humanitarian actors to gain the confidence of all parties and access to people in need.

In today's information age, where warring parties actively use online platforms to frame polarizing narratives and civilians engage in digital discourse, neutrality also requires refraining from being drawn into a war of words. It must be upheld both online and offline.

Ultimately, neutrality is about building trust as a means to an end: enabling the delivery of impartial humanitarian action. To safeguard it, principled humanitarian actors must consistently advocate for the protection of a neutral and impartial humanitarian space, free from political influence. States also have a key role in preserving this humanitarian space. The ability to protect and assist victims of armed conflicts, in accordance with the 1949 Geneva Conventions and their Additional Protocols, often depends on strict adherence to the principles of humanity, impartiality, neutrality and independence. These principles remain the essential tools to access the most at-risk people and to gain their trust.

That does not mean that the Movement components always ‘get it right’ in what they say – or choose not to say. Timing, consistency and clarity remain essential and difficult. These dilemmas are explored further in [Chapter 7, on page 257](#) on the fundamental principles.

Contributor Insight 2.8

Principled humanitarian action under threat

The Fundamental Principles of the International Red Cross and Red Crescent Movement – humanity, impartiality, neutrality, independence, voluntary service, unity and universality – form the ethical backbone of principled humanitarian action. Among these, **neutrality** and **impartiality** are particularly vital to ensuring trust, access and effectiveness in volatile environments.

Neutrality requires that the Movement refrains from taking sides in conflicts or engaging in political, racial, religious or ideological controversies. This principle is essential for maintaining credibility and securing access to all parties in a crisis. **Impartiality**, meanwhile, demands that aid be provided solely based on need, without discrimination of any kind. Together, these principles safeguard the Movement’s ability to serve vulnerable populations equitably and without bias.

However, upholding these principles is increasingly difficult. The politicization of aid, the spread of misinformation and rising social polarization have created environments where neutrality and impartiality are often misunderstood or manipulated. Humanitarian actors face difficult dilemmas such as whether to delay aid due to safety concerns, how to respond to conditional funding that excludes host populations or how to maintain neutrality while engaging with political stakeholders. For example, accepting funding that excludes certain groups can compromise impartiality and give the appearance of political bias. Similarly, a National Society publicly endorsing a political party undermines neutrality, eroding public trust and donor confidence. Even volunteers expressing political views while in uniform – or even privately on social media – may jeopardize the Movement’s perceived neutrality.

Navigating these dilemmas requires transparency, sound judgement and unwavering commitment to the seven fundamental principles. The International Red Cross and Red Crescent Movement must continuously educate stakeholders and the public about the

meaning and value of principled humanitarian action, especially in today's highly complex and polarized world.

Lessons learned:

- Across the world, there is a growing disregard for the core principles that guide humanitarian action: humanity, neutrality, impartiality and independence.
- Increasing polarization, the politicization of aid and dehumanizing rhetoric are fuelling distrust, jeopardizing access to communities and placing the most vulnerable people at greater risk.
- Saving lives is not a political act: it is a humanitarian imperative and one that National Red Cross and Red Crescent Societies are mandated by their states to uphold.
- Neutrality and impartiality are not abstract ideals – they are essential to operating in complex, polarized environments. When these principles are ignored or weakened, the consequences extend beyond endangering aid workers – they endanger the people who rely on protection and assistance.

Anna Bowen

Regional Humanitarian Diplomacy Coordinator

IFRC, Regional Office for Europe

2.8 Responding to harmful information: Building trust in crises

Responding to harmful information requires more than increasing the volume of communication or correcting falsehoods. These approaches alone do not build trust and can even backfire, particularly during armed conflicts, political unrest or situations marked by discrimination. Effective responses go beyond countering narratives; they focus on disrupting harmful information and strengthening resilience at both organizational and societal levels. Trust is built through proximity, understanding, access to services and a sense of shared ownership.³³ Crucially, it generally cannot be created in the middle of a crisis; it must be cultivated over time.



Trust is the foundation of humanitarian action. Without it, our services cannot reach those who need them most. It enables cooperation, open communication and gives communities the confidence to call on us in times of crisis, knowing we will preserve life and support recovery. In my time as a volunteer and Red Cross Action Team member, I have seen how trust allows us to reach the most vulnerable people without resistance, strengthen resilience and encourage participation in rebuilding and preparedness.

Saving lives is not only about speed or skill, it is about the depth of trust we have built, because in emergencies, trust is truly transformative.”

Hodan Ismail-Shukry, Volunteer, **Kenya Red Cross Society**

The COVID-19 pandemic, the most extensive disaster in living memory, exposed deep fractures in trust within and between countries. Harmful information became a major barrier, undermining public health efforts and fuelling social division. Where trust was strong, compliance with public health measures such as social distancing and vaccination was higher. Where trust was weak, these same measures became politicized, contested and fragmented. The experience underscores a critical lesson: trust is foundational to crisis preparedness and response and it emerged as one of the strongest predictors of successful response. Preparedness is inherently local and built through proximity and sustained engagement, and reinforced through transparency long before crisis strikes.

Humanitarian scholar Hugo Slim observes that trust in humanitarian action is under pressure and organizations must confront difficult questions: Can parties to conflict trust humanitarians to be neutral and impartial? Can vulnerable people trust humanitarian organizations to deliver aid fairly and treat them with dignity? Can donor governments and generous publics be sure their support will be well spent? Can humanitarian organizations trust each other to work in common cause?³⁴ Slim argues that humanitarian organizations themselves have contributed to the trust deficit. Research by Grand Challenges Canada also makes this link, noting that a lack of transparent communication, conflicting mandates and failure to protect communities when needed can exacerbate distrust. These shortcomings not only erode public confidence but also create fertile ground for harmful information to thrive – further deepening scepticism and reluctance to trust humanitarian organizations.³⁵

2.9 Transparency and identifiability: Foundations for trust

Digital virality is rarely organic; it is often deliberately engineered by a small number of influential accounts that dominate the attention economy, shaping beliefs and behaviours across platforms. As Singer and Brooking note: “On social media, everyone may be entitled to their own facts, but rarely do they form their own opinions. There’s someone else manufacturing the beliefs that go viral online.”³⁶ In this environment, identifiability becomes a cornerstone of trust. Trust depends not just on **what** is said, but on **who** is saying it – whether an individual, an institution, a troll or an AI system. In particular, “It helps answer some of the questions that trust inspires us to ask: Is there a recognisable and persistent identity to the institutions and individuals behind the myriad websites one might visit?”³⁷ Identifiability supports reputation, accountability and helps distinguish credible sources from malicious actors.³⁸

The Bengio Report on AI safety (see [Chapter 1, on page 29](#)) warns that the rapid spread of AI-generated content could further erode trust in the information environment.

General-purpose AI can now produce (and spread) both accurate and false content – including synthetic media – at an unprecedented scale. This raises the risk that people may begin to distrust information altogether. This dynamic could undermine public debate (and in democracies, democratic processes). Malicious actors can exploit this ambiguity and distrust through the so-called ‘liar’s dividend’ – dismissing real evidence as fake. However, societies may adapt over time, developing new norms and tools to evaluate credibility, much as they did with past technologies like image, video and audio editing.³⁹ But until such norms take hold, uncertainty and distrust are likely to deepen.

For humanitarian actors, real-time monitoring of both the information and operational environment (online and offline) is essential to understand the type, narratives, impact and potential harm of harmful information. Yet the accessibility of such tools and systems available for verifying information has not kept pace with the speed and scale of harmful information – particularly given the financial constraints faced by humanitarian actors.

Contributor Insight 2.9

Verifiable provenance and the challenge of trust in the digital age

In humanitarian crises and beyond, images and videos are essential tools for response coordination, communicating with affected communities, advocacy and accountability. But as synthetic media and manipulated content become more accessible and harder to detect, questions around what and who to trust have never been more urgent.

‘Verifiable provenance’ can be a key part of a broader toolkit for building digital trust and resilience. This refers to the ability to cryptographically verify the origin, history and integrity of digital content. Initiatives such as the [Coalition for Content Provenance and Authenticity \(C2PA\)](#) are developing cross-sector open technical standards to attach secure metadata to digital files. These standards can help determine whether an image was altered, who published it and what tools were used – offering markers for authenticity and digital trust. However, provenance is not a silver bullet. It comes with its own risks, such as exposing sensitive information, enabling surveillance, increasing reliance on major platforms and proprietary tools, and reinforcing a trust gap between content that is verifiable and content that is not. This is why it is necessary to [shape the provenance ecosystem](#) – including standards, legislation and implementation – with a focus on privacy, equitable access, power dynamics and the protection of human rights.

In contexts where generative AI can hinder humanitarian relief or where digital evidence is critical to justice, verifiable provenance tools can help safeguard the truth. But these tools must be deployed in ways that uphold human rights, reflect local realities and actively support humanitarian response efforts.

Jacobo Castellanos

Coordinator, Technology, Threats and Opportunities

WITNESS

Transparency is a prerequisite for trust and a core element of the Red Cross and Red Crescent Code of Conduct. It underpins accountability by ensuring clear and honest information flows between humanitarian organizations and affected people. Transparency is embedded in key sectoral standards, including the Sphere Common Standards, which emphasize feedback mechanisms and participation, the Good Humanitarian Donorship Principles and the IFRC's Guidelines for Emergency Assessment.



Movement-wide commitments for community engagement and accountability⁴⁰

Accountability refers to the mutual responsibility of all components of the Movement to use their power and resources ethically and responsibly to put the interests of the people and communities they aim to serve at the centre of decision-making, thereby ensuring that humanitarian actions lead to the best possible outcomes and results for them, while protecting and preserving their rights and dignity and increasing their resilience to face situations of vulnerability and crisis. This includes people's rights to equitable access to assistance in proportion to their needs, priorities and preferences, the right to information, and the right to provide feedback and participate in decisions that affect them.

Accountability also includes the responsibility to ensure Movement staff and volunteers comply with all legal and ethical obligations to protect and safeguard the integrity and dignity of the people they seek to serve, prevent sexual exploitation and abuse as well as corruption and misuse of resources, and to take timely and appropriate corrective action to deal with situations that compromise principled humanitarian action. It is also essential to ensure that corrective actions are applied consistently, without double standards – either in the measures themselves or in the expectations placed on National Society conduct. In practice, this can be challenging and is not always straightforward to apply.

Concluding remarks: Trust, truth and preparedness

We are already living through an information crisis. Future emergency preparedness must include equitable access to reliable information, supported by early warning systems that serve all communities, especially the most at risk. Harmful information is not only eroding trust in humanitarian action, it is fuelling societal division and undermining cooperation at a time when unity is urgently needed to face global challenges.

As the World Economic Forum warned, misinformation and disinformation are no longer just communications problems – they are systemic threats that aggravate nearly every

other global risk. The 2025 *Global Risks Report* underscores that: “Misinformation and disinformation and societal polarization remain key current risks” and explains how the accelerating spread of false or misleading information amplifies other major risks – from state-based armed conflict to extreme weather events.⁴¹

In such a fragmented and contested information space, the voices of affected populations risk being drowned out, distorted or co-opted. Trust cannot be demanded; it must be built through repeated action, transparency, accountability and integrity. Resilience against harmful information requires more than just correcting falsehoods, it requires sustained engagement, openness and meaningful participation. Humanitarian organizations must also actively communicate who they are, what they do, their impact and why they act – making a clear, compelling case for principled humanitarian action and for preserving a neutral and impartial humanitarian space, free from political influence. Despite the barriers posed by polarized perceptions, principled humanitarian organizations can maintain – and even rebuild – trust and acceptance in the face of harmful information. Looking ahead, a critical question is whether trust and proximity at the community level can serve as a firewall against the spread or impact of harmful information. [Chapter 3, on page 103](#) and [Chapter 6, on page 211](#) explore this question.

Asks, aims and recommendations

Asks

Place trust at the centre of humanitarian action by ensuring that the principles of rights, dignity, inclusivity, accountability, transparency and meaningful feedback guide how organizations communicate and engage with communities.

Aims

Reduce uncertainty through timely, transparent and consistent communication and engagement before, during and after crises.

Strengthen legitimacy and accountability by aligning with community priorities and tracking trust.

Protect staff and volunteers with safeguards, skills and inclusive engagement to operate safely in contested information spaces.

Empower communities: Build two-way feedback and participation systems that allow people to voice concerns, influence decisions and counter harmful narratives.

Recommendations

States and policy-makers

- Establish rights-based policy frameworks that safeguard access to reliable information in humanitarian crises.
- Embed trust-building into crisis preparedness and response plans.
- Invest in early-warning and monitoring systems to detect and counter harmful narratives.
- Support National Societies in their auxiliary role, ensuring independence, impartiality, neutrality and integrity are respected.
- Ensure rapid rebuttals of falsehoods that threaten humanitarian access, action and security.

Technology platforms

- Detect and mitigate harmful content that undermines humanitarian action and trust.
- Share relevant data and insights safely with humanitarian actors to support real-time response.
- Ensure moderation, fact-checking and AI tools work in low-bandwidth, multilingual environments.
- Collaborate with humanitarian actors on verification and labelling mechanisms to amplify credible content.

Humanitarian actors

- Develop policies, guidelines, standards and metrics (e.g., Community Trust Index and community engagement and accountability (CEA)) to measure and track trust and harmful information, and guide responses.
- Embed trust as an operational asset across preparedness, response and recovery.
- Communicate transparently and inclusively, correcting falsehoods rapidly and co-creating messages with communities.
- Train staff and volunteers in rumour management, digital safety (including protection from online harassment) and culturally sensitive engagement.
- Strengthen real-time analysis of and response to community feedback, adapting visibly to concerns, demonstrating impact and sharing lessons learned to build collective resilience.

Communities and local leaders

- Act as trusted intermediaries by amplifying verified information and countering rumours.
- Partner with humanitarian actors to co-create rumour-tracking and verification systems.
- Sustain trust through dialogue in schools, faith institutions and community centres.
- Engage in participatory feedback and research to ensure that responses reflect local priorities.

Endnotes

- 1 See Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You*. (2011)
- 2 Ibid
- 3 Pierre, J. *False: How Mistrust, Disinformation, and Motivated Reasoning Make Us Believe Things That Aren't True*. (2025) p.39
- 4 See John Suler, Rider University psychologist (2004) referenced in Pierre, J. *False: How Mistrust, Disinformation, and Motivated Reasoning Make Us Believe Things That Aren't True*. (2025) p.40
- 5 Ibid, pp.50, 59
- 6 Nichols, T. *The Death of Expertise: The Campaign against Established Knowledge and Why it Matters*. (2024) pp. 50–58, 134–139
- 7 Kavanagh, J. and Rich, M.D. *Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life*.
- 8 Cited in Pierre, J. *False: How Mistrust, Disinformation, and Motivated Reasoning Make Us Believe Things That Aren't True*, (2025), p.59
- 9 Rousseau, DM., Sitkin, SB., Burt, RS. et al. Not So Different After All: A Cross-Discipline View of Trust. *Academy of Management Review* 1998:23(3), p.395
- 10 Edelman. Edelman Trust Barometer (2022)
- 11 Bavel, JJV., Baicker, K., Boggio, PS. et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat Hum Behav* 2020:4, 460–471. <https://doi.org/10.1038/s41562-020-0884-z>
- 12 Pierre, J. *False: How Mistrust, Disinformation, and Motivated Reasoning Make Us Believe Things That Aren't True*, (2025), p.45
- 13 Ibid, p.52
- 14 IFRC. *World Disasters Report 2022: Trust, Equity and Local Action – Lessons from the COVID-19 pandemic to avert the next global crisis*. (2022) p.68, www.ifrc.org/document/world-disasters-report-2022
- 15 Royston, G., Pakenham-Walsh, N. and Zielinski, C. Universal access to essential health information: accelerating progress towards universal health coverage and other SDG health targets. *BMJ Global Health*. 2020;5:e002475. <https://doi.org/10.1136/bmjgh-2020-002475>
- 16 King's College London. COVID-19 Monitor: Public attitudes and behaviours. (2020)
- 17 Erlach, E., Nichol B, Reader S, et al. Using Community Feedback to Guide the COVID-19 Response in Sub-Saharan Africa: Red Cross and Red Crescent Approach and Lessons Learned from Ebola. *Health Secur*. 2021;19(1), 13–20. doi: 10.1089/hs.2020.0195
- 18 Pierre, J. *False: How Mistrust, Disinformation, and Motivated Reasoning Make Us Believe Things That Aren't True*, (2025), p.7
- 19 33rd International Conference of the Red Cross and Red Crescent. Resolution 6: Act Today, Shape Tomorrow (Point 1). (2019) (33IC/19/R6) https://rcrcconference.org/app/uploads/2019/12/33IC-R6-Act-today-shape-tomorrow_CLEAN_ADOPTED_en.pdf
- 20 A dedicated commission was convened to explore how the Movement and states can work together to protect and strengthen trust in principled humanitarian action. 33rd International Conference of the Red Cross and Red Crescent. (2019) *Summary Report from Commission III: Trust in Humanitarian Action*. https://rcrcconference.org/app/uploads/2020/05/33IC-Commission-III-Trust-in-humanitarian-action-report_FINAL-EN.pdf
- 21 33rd International Conference of the Red Cross and Red Crescent. (2019) *Summary Report from Commission III: Trust in Humanitarian Action*. https://rcrcconference.org/app/uploads/2020/05/33IC-Commission-III-Trust-in-humanitarian-action-report_FINAL-EN.pdf
- 22 Council of Delegates of the International Red Cross and Red Crescent Movement. Resolution 5: Call for Respect and Support for Principled Humanitarian Action (2024) (CD/24/R5)
- 23 OECD. An updated OECD framework on drivers of trust in public institutions to meet current and future challenges. (2021) p.41 <https://doi.org/10.1787/b6c5478c-en>
- 24 See, for example, Slawson, N. 'Oxfam government funding cut off after Haiti scandal.' *The Guardian*. 16 February 2018. www.theguardian.com/world/2018/feb/16/oxfam-government-funding-cut-off-after-haiti-scandal
- 25 Levitin, DJ. *A Field Guide To Lies: Critical Thinking in the Information Age*. (2016) pp.x-xi, 123–8, 129–130, 152–4
- 26 Nichols, T. *The Death of Expertise: The Campaign against Established Knowledge and Why it Matters* (2024) p.xiii
- 27 As outlined in Council of Delegates. Resolution 1: Movement-wide Commitments for Community Engagement and Accountability. (2019) (CD/19/R1, Annex)
- 28 Ibid
- 29 IFRC. 'Placing communities at the centre of the Ebola response.' Article. 13 September 2019. www.ifrc.org/article/placing-communities-centre-ebola-response
- 30 IFRC, ICRC, German Red Cross et al. *Navigating Fragility, Conflict and Violence to Strengthen Community Resilience: A Handbook for Disaster Risk Reduction Practitioners*. (2024) p.38 www.climatecentre.org/wp-content/uploads/Full-Handbook_RCC-Navigating-fragility-conflict-and-violence-to-strengthen-community-resilience.pdf
- 31 The seven fundamental principles are humanity, impartiality, neutrality, independence, voluntary service, unity and universality. These are referenced in the Preamble to the Statutes of the International Red Cross and Red Crescent Movement adopted by the 25th International Conference of the Red Cross in October 1986.
- 32 This quote is commonly attributed to Desmond Tutu, though the exact origin is debated. Tutu, D. *No Future Without Forgiveness*. (Doubleday, 1999)
- 33 IFRC. *World Disasters Report 2022: Trust, Equity and Local Action*. Executive summary. (2022) p.9 www.ifrc.org/document/world-disasters-report-2022
- 34 Slim, H. 'Trust Me – I'm a Humanitarian.' ICRC Humanitarian Law and Policy Blog. 24 October 2019. <https://blogs.icrc.org/law-and-policy/2019/10/24/trust-humanitarian>
- 35 Ayala Iacucci, A. *Misinformation, Disinformation, and Hate Speech in Humanitarian Contexts*. Grand Challenges Canada, Creating Hope in Conflict initiative. (2024) p.31 https://humanitariangrandchallenge.org/wp-content/uploads/2024/11/MDH-Scoping-Study_Full-report.pdf
- 36 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) p.127
- 37 Nissenbaum, H. *Will Security Enhance Trust Online, or Supplant It?* In Kramer, RM. and Cook, KS. *Trust and Distrust in Organizations: Dilemmas and Approaches*. (2004) p.166
- 38 Ibid
- 39 Bengio, Y., Mindermann, S., Privera, D. et al. *International Scientific Report on the Safety of Advanced AI*. International Scientific Panel on the Safety of Advanced AI. (2025) <https://internationalaisafetyreport.org>

40 Council of Delegates. Resolution 1: Movement-wide Commitments for Community Engagement and Accountability. (2019) (CD/19/R1, Annex)

41 World Economic Forum. *Global Risks Report 2025*. (2025) p.13 https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf



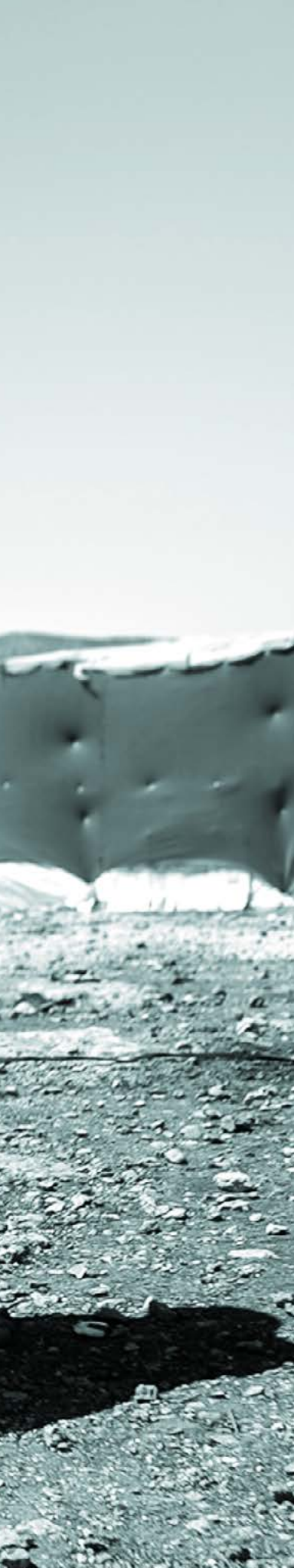
Chapter 3



**Global and local:
Dynamics of harmful
information in a
connected world**



Chapter 3



Global and local: Dynamics of harmful information in a connected world

Contents

	Introduction: Harmful narratives that thrive	105
3.1	Local and global interplay	105
3.2	When words harm: The International Red Cross and Red Crescent Movement's call to action	110
3.3	The evolving nexus between cyberattacks and harmful information	111
3.4	Addressing harmful information, strengthening early warning and response	115
3.5	Risk communication and community engagement	123
	Concluding remarks: Navigating humanitarian action in a post-trust information era	127
	Endnotes	131

Introduction: Harmful narratives that thrive

Climate change, vaccine hesitancy and migration are among the issues on which harmful information thrives, crossing borders and continually being reshaped by local contexts, narratives and political agendas. These same issues also contribute to humanitarian crises. Humanitarian organizations therefore face a delicate balancing act: they must meet urgent needs while navigating what can and cannot be said, avoiding controversies that could inflame tensions or compromise perceptions of neutrality.

At the heart of this challenge lies the power of narrative. Narratives shape how people interpret information and decide whom to trust and whether to engage with humanitarian actors. They are reinforced by lived experience, public discourse, media coverage and digital engagement. When there is a gap between dominant narratives and people's daily realities, trust erodes and alternative or harmful narratives often emerge. These narratives frequently define identity by portraying certain individuals or groups as outsiders, threats or illegitimate – as the 'other'. Such framing generates fear, fuels exclusion, dehumanizes and normalizes hostility. For example, displaced populations may be cast not as victims of a crisis but as threats to national stability or competition for scarce resources. In this way, narrative shapes perception and the notion of the 'other' shapes belonging – who is trusted or untrusted, legitimate or illegitimate.

Narratives also simplify complexity, evoke emotion and may weaken social cohesion. In today's information ecosystem, humanitarian organizations must contend not only with physical threats but also with digital and cyber ones. They are increasingly framed as ineffective responders, political tools or even foreign agents. This makes it harder to ensure that accurate, trusted information reaches the people who need it most and can be distinguished from harmful content. In contested or sensitive environments, people may instead turn to digital communities for answers – even when these communities amplify harmful messaging driven by actors lacking expertise or clear political or economic agendas.

Malicious actors exploit people's limited capacity to process information by amplifying feelings of fear, grievance and crisis. This does not just create short-term confusion – it entrenches long-term mistrust. The consequences are real. Harmful narratives influence how communities perceive risk, how they respond during emergencies and how they prepare for future shocks, including disasters. In moments of crisis, harmful narratives can fuel denial, panic or resistance – ultimately undermining life-saving interventions.

3.1 Local and global interplay

The relationship between local realities and global narratives is a continuous feedback loop. What resonates globally may not resonate locally, while local experiences – though vital – too rarely shape global communication approaches. Global messaging risks missing the mark when it fails to reflect lived realities, priorities or the linguistic, cultural and political contexts of communities. Even when factually accurate, such messaging may

be perceived as tone-deaf, abstract or slow to adapt. Yet local dynamics have already influenced – and must continue to influence – global humanitarian positioning and diplomacy. These moments must be captured and applied.

Community insights and lived experiences are critical to strategies for countering or responding to harmful information, including by challenging flawed assumptions, surfacing blind spots and revealing opportunities for trust-building that global actors might otherwise overlook. Importantly, communities are not passive. Local voices are actively pushing back, using digital platforms to counter harmful information, reclaim narratives and assert agency. These efforts should be recognized, amplified and better connected to global strategies.

The information landscape is complex: harmful narratives resonate differently depending on political, historical, social, cultural and economic contexts. Understanding this nuance is essential for shaping responses that are relevant, respectful and effective. Narratives that target humanitarian organizations are often deliberately designed to erode trust, challenge neutrality and disrupt access to affected populations and other stakeholders. They frequently exploit allegations of misconduct, misuse of aid or funds, operational inefficiencies and failures to uphold humanitarian principles – particularly neutrality and impartiality.

In some settings, humanitarian actors are framed as being too closely aligned with governments, undermining perceptions of independence. In others, they are portrayed as detached from communities, reinforcing views of elitism, foreignness or lack of local legitimacy. Such portrayals often feed into broader narratives of neocolonialism or external interference, further complicating acceptance and access. These risks demand both strategic foresight and renewed commitment to inclusion, transparency and sustained community engagement – ensuring that humanitarian action is both principled and trusted at every level.

Contributor Insight 3.1



Crossing borders, losing truth: The humanitarian cost of misinformation in migration

Trust is essential to humanitarian action. Without the trust of affected populations, public authorities and donors, humanitarian organizations cannot reach or respond effectively to the needs of the most vulnerable people, including many migrants. Yet in an era of misinformation, that trust is increasingly at risk, especially for migrants who experience harm, discrimination and exploitation along their journeys. When misinformation erodes trust and fuels fear, migrants may avoid seeking critical assistance and support, with life-threatening consequences.

Evidence from research led by the Red Cross Red Crescent Global Migration Lab (the Lab),¹ involving over 20,000 migrants across 34 countries, highlights how misinformation

impacts migrants' willingness and ability to access humanitarian assistance and protection. Studies related to the COVID-19 pandemic, trust in humanitarian action, and missing migrant women and children consistently show that inaccurate or inaccessible information undermines safety and dignity throughout migration journeys.²

During the COVID-19 pandemic, a study by the Lab involving more than 3,200 migrants across eight countries confirmed that a major barrier to accessing health services – including COVID-19 treatment and vaccines – was the lack of accessible information, particularly in migrants' languages and via trusted channels.³ As one migrant in the UK explained, “people are very confused ... they are not getting the right information ... They do not know what to do or even where to go to get information ...” In a parallel survey of 52 National Red Cross and Red Crescent Societies, 90% confirmed these concerns, reporting that poor availability of translated information and limited awareness of how and where to access vaccines were key barriers to migrants staying healthy and safe.⁴ This lack of reliable information fuelled vaccine hesitancy and heightened health risks – exacerbating migrants' vulnerabilities.⁵

Beyond public health emergencies, misinformation continues to endanger migrants. Research with 16,000 migrants across 15 countries found that access to accurate, practical information – such as contact details for services or legal guidance – was crucial to reducing exploitation and ensuring safer journeys. Without it, many migrants are forced to rely on unverified sources of information. As one migrant in Australia highlighted: “... we suffer from dodgy [unreliable] information provided to us by migration agents, for example”. Misinformation also fosters mistrust in humanitarian organizations, especially when they are not perceived as independent, but as aligned with authorities. This perception deters migrants from seeking essential support out of fear of arrest, detention or deportation.⁶

Research by the Lab on missing migrant women and children, based on discussions with over 800 participants across 17 countries, further reinforces how misinformation and distrust contribute to increased vulnerability and loss of life during migration.⁷ Migrants cited limited access to accurate information about how to access essential services and route-specific risks as key factors in exposure to threats, including human trafficking. Many migrants reported receiving false or misleading information before departure, often online or from smugglers – which shaped unrealistic expectations and led to harm. As one migrant in Africa noted, “My journey was just misinformation... at times you might ask questions and still get wrong answers.” Another in the Americas said, “The information did not help me at all because it was all a lie, the guide scammed us...” Social media, while helpful in some cases for staying connected, was often described as a “double-edged sword” – a source of both connection and harmful information.

Misinformation and disinformation about migrants also impact their safety, dignity and well-being by shaping public opinion and policy. As highlighted by the Kaldor Centre for International Refugee Law, misinformation about migrants is increasing globally, undermining social cohesion and causing harm.⁸ In countries where migrants are portrayed as threats, public discourse tends to support more restrictive migration policies and reduced access to essential services, increasing vulnerability.⁹

Existing research, including studies by the Lab, underscores the critical role of information – and misinformation – in shaping migrants' trust in and access to humanitarian assistance and protection. To address this, communication strategies must ensure that migrants receive clear, relevant and trustworthy information across all stages of their journey – in countries of origin, transit, destination and return. This information must be delivered in

ways that reflect migrants' diverse needs, including language, gender, age and access to technology. Importantly, migrants in the Lab's research identified family and friends who had previously migrated as their most trusted source of information. This highlights the value of engaging migrant communities and local networks in designing and delivering life-saving information.

Nicole Hoagland

Senior Adviser for Global Policy & Engagement

Red Cross Red Crescent Global Migration Lab

Magdalena Arias Cubas

Research Lead

Red Cross Red Crescent Global Migration Lab

Contributor Insight 3.2

UNHCR information integrity

In 2023, the UN High Commissioner for Refugees (UNHCR) launched a two-year ECHO-funded project to enhance UNHCR and humanitarian partners' understanding of and responses to misinformation, disinformation and hate speech (MDH) on digital platforms. The project included research, development of tools and guidance, advocacy and three field-based pilots. One of these pilots focused on the Rohingya. Its objectives were to:

- 1 Improve understanding of how MDH develops, spreads and impacts the protection of Rohingya.
- 2 Develop an action plan and tools to help UNHCR and partners respond to and mitigate the impacts of MDH targeting Rohingya.
- 3 Develop a strategy and recommendations for short- and medium-term work, ensuring collaboration with relevant actors and communities.

The pilot aimed to strengthen responses to MDH affecting displaced populations and humanitarian actors. Activities included country-level assessments to identify trends, protection impacts, and key partnerships and advocacy entry points. A regional multi-stakeholder meeting and follow-up webinar brought together UN agencies, civil society, digital rights organizations, journalists, academics and the private sector to share lessons learned and elaborate potential responses.

The pilot tested a range of responses, including:

- direct advocacy with technology companies to address harmful content and impersonation
- capacity-building workshops and crisis communications training incorporating MDH and social media risk
- country-tailored social media listening methodologies
- development of guidance and recommendations for specific incidents and general operational responses.

Regional collaboration and advocacy efforts were key to raising awareness and promoting coordinated responses to MDH targeting Rohingya.

The pilot also benefited from partnerships at global and regional levels, including those strengthened by the Global Refugee Forum Multistakeholder Pledge on Digital Protection. Guided by an advisory group of refugee experts, this initiative now includes 25 pledges from governments, the private sector, civil society and international organizations.

The pilot itself established new partnerships and built a baseline understanding of MDH trends to inform future analysis. Refugee-led initiatives have demonstrated the value of community-driven responses.

Key lessons from the pilot show that MDH is complex and constantly evolving, requiring flexible and adaptable responses. MDH campaigns can be coordinated or emerge organically. They may be driven by political agendas, financial gain or identity-based narratives. Local context matters – tailored responses informed by real-time monitoring are essential. Linking online and offline insights adds value, but sustained monitoring is resource intensive and most useful if undertaken with the aim of informing programmes and/or tailored advocacy. Integrating MDH strategies into emergency preparedness is increasingly important and more must be done to include forcibly displaced and stateless persons in designing and delivering actions to respond to online information integrity risks.

Gisella Lomax

Senior Advisor, Information Integrity

UNHCR

Contributor Insight 3.3

What drives trust at the community level?

Across contexts – from the Sahel to Syria – local actors consistently enjoy higher trust and lower hostility in online spaces compared to their international counterparts. In Burkina Faso, Mali and Niger, local NGOs received virtually no negative comments across several months of monitoring, even when addressing sensitive topics such as reproductive health or gender-based violence. This trust is often rooted in perceived cultural alignment, local presence and visibility of known individuals, which often shields them from broader anti-aid or anti-foreign narratives.¹⁰

However, trust alone does not guarantee reach. In the Sahel, local NGOs frequently operate in low-engagement digital environments, limiting their ability to shape public discourse. Posts by local organizations often go unnoticed unless amplified by media or influencers. In some cases, external visibility without narrative control can backfire. For example, when a Burkinabè local NGO initiative on female genital mutilation was featured by a third-party media outlet, it **faced backlash** that had not occurred when the messaging came directly from the organization itself. Conversely, when female empowerment is promoted by trusted local figures, it can generate positive engagement. For example, Facebook posts by a Burkinabè artist and UNDP Goodwill Ambassador celebrating a women's vocational

training initiative were **widely praised and avoided the criticism** often directed at content associated with aid agencies.

In 2024, the Goodwill Ambassador posted content supporting UNDP's role and received only positive comments, in sharp contrast to the general scepticism surrounding UN entities. This demonstrates that trusted messengers can act as a **buffer against harmful narratives**. Yet, influencer engagement must be strategic and context aware. In June 2025, **UNICEF faced backlash** following the appointment of a new foreign representative in Burkina Faso, with 85% of user comments expressing suspicion, often invoking sovereignty and anti-imperialist narratives. The stark difference in public sentiment toward local leaders versus institutional figures reflected complex dynamics and it is not always clear how much of the reaction is organic or coordinated.

In Syria, the contrast was similarly pronounced. In the monitored posts, local responders like the White Helmets attracted significantly more positive engagement than UN agencies, which were often criticized for perceived regime bias or bureaucratic distance. Community trust was strongly tied to **transparency, visible impact and narrative ownership on social media**.

To strengthen digital resilience, aid actors should invest in amplifying credible local voices, support community content creators and co-produce messaging with those already seen as legitimate. When humanitarian communication is grounded in relationships and proximity, it becomes far more effective in navigating contested digital spaces.

Christina Wille

Director

Insecurity Insight

Clara de Solages

Researcher

Insecurity Insight

3.2 When words harm: The International Red Cross and Red Crescent Movement's call to action

In its Appeal to States included in the resolution on neutral, independent, impartial humanitarian action¹¹ at the 2024 Council of Delegates,¹² the International Red Cross and Red Crescent Movement called on governments to take all appropriate measures to prevent, stop and remedy “any abuse, pressure, misinformation, disinformation and dehumanizing rhetoric”, particularly when spread by social media. The appeal emphasized the serious harm such content causes to the physical, psychological and reputational well-being of people in vulnerable situations, as well as to the staff and volunteers of the Movement.

By explicitly recognizing harmful information – whether in the form of disinformation, misinformation or dehumanizing rhetoric – as a global issue with profound humanitarian consequences, the Movement marked a pivotal moment. It was the first time that these issues had been directly addressed in its history, dating back to the inaugural international conference in 1867. While earlier themes had touched on information, they

focused largely on coordination and modalities for information sharing¹³ rather than on the humanitarian dangers of harmful information.

No specific resolutions have previously addressed propaganda, despite its widespread use in wartime. Under international humanitarian law, the use of propaganda, misinformation or disinformation during armed conflict is not per se prohibited, though certain forms may violate specific rules.¹⁴

The 2024 resolution further expressed concern over the spread of malicious ICT activities intended to cause, instigate or amplify harm to civilians and other protected persons and objects in armed conflict. It acknowledged that such activities may directly harm humanitarian organizations by:

- enabling data breaches and disinformation that target them
- undermining trust and threatening the safety of personnel, premises and assets
- threatening humanitarian access and the ability to deliver life-saving assistance.¹⁵

These threats underscore a critical shift: the Movement has now placed – at its highest platform – harmful information and malicious digital activity firmly within the realm of humanitarian concern, recognizing them as urgent threats to lives, dignity and principled humanitarian action.

Three months after the Movement's statutory meetings, the World Economic Forum released its *Global Risks Report 2025*,¹⁶ based on the Global Risks Perception Survey 2024–2025. This describes a year marked by escalating conflicts, extreme weather events amplified by climate change, widespread societal and political polarization and continued technological advances accelerating the spread of false and misleading information. For the second consecutive year, misinformation and disinformation were ranked as the top global risk projected for 2027.¹⁷ The report underscores how false or misleading content is complicating geopolitics – shaping voter behaviour and distorting realities in conflict zones. The growing use of digital platforms and surge in AI-generated content are fuelling the spread of divisive misinformation and disinformation. As polarization deepens, algorithmic bias – and its role in amplifying misleading content – may become more pervasive. The report further warns that the vulnerabilities linked to citizens' online activities are deepening alongside societal and political divisions, trends which risk fundamentally eroding individuals' trust in information and institutions.¹⁸

3.3 The evolving nexus between cyberattacks and harmful information

The growing intersection between cyberattacks and harmful information campaigns has amplified the impact of both. Coordinated information operations often accompany

cyber incidents: shaping narratives, sowing confusion and manipulating public perceptions. The objective is not only to influence specific outcomes but also to erode trust in institutions. This means that cybersecurity must go beyond encryption, firewalls and digital hygiene to encompass information resilience – the capacity to withstand, detect and respond to narrative manipulation.

As discussed in [Chapter 1, on page 29](#), the digital and cyber domains are largely shaped by two distinct perspectives:

- **Content:** Information, narratives, platforms and applications. Content influences hearts and minds and is typically governed through the lens of values, rights and public trust. Key concerns include control of the information space, foreign influence, censorship and threats to freedom of expression.
- **Cables:** The physical infrastructure, hardware and networks that enable connectivity. This perspective is dominated by sovereignty, security and strategic power, with debates centred on who owns, builds and operates the infrastructure. Key concerns include surveillance, cybersecurity, strategic autonomy and economic dependence.

These dual perspectives lead to very different governance approaches and interpretations of sovereignty, control and threat in the context of international security. The tension between them has surfaced in discussions (such as those at the UN Open-Ended Working Group on ICTs¹⁹) where some states emphasize the need to promote respect for the non-binding norms of responsible state behaviour and foster international cooperation, while others stress the importance of preserving state sovereignty, minimizing external dependencies and controlling critical ICT infrastructure. Understanding these divergent perspectives is essential. They shape the broader information environment in which foreign influence operations, harmful information campaigns and cyber incidents occur – contexts that increasingly determine whether humanitarian organizations can operate safely and the space for principled humanitarian action. States have expressed concern that malicious ICT activities targeting international and humanitarian organizations undermine their safety, independence and public trust.²⁰

Harmful information campaigns – including influence operations and other forms of foreign information manipulation and interference²¹ – are carried out by both domestic and foreign actors using increasingly sophisticated strategies to shape public opinion. These campaigns often deploy content that appears locally produced and organic. Common tactics include:

- coordinated inauthentic behaviour
- cyberattacks
- fake websites and social media accounts
- algorithmic amplification of diverse narratives
- impersonation of trusted entities
- engagement manipulation through trolls, propagandists, hashtags and bots

- spread of divisive content across platforms.

These tools are routinely used to erode trust, deepen polarization and influence political processes, including elections and broader national security priorities.

Meta defines coordinated inauthentic behaviour as “coordinated efforts to manipulate public debate for a strategic goal, in which fake accounts are central to the operation. In each case, people coordinate with one another and use fake accounts to mislead others about who they are and what they are doing.”²² Meta, with over 3 billion users globally (including on Facebook, Instagram, Messenger, WhatsApp and Threads), employs a combination of automated systems and manual review to detect and remove accounts and pages of coordinated inauthentic behaviour networks.

What makes an influence operation malign? Scholarly and policy research examines factors such as:

- 1 Transparency of origin – Who is behind the operation?
- 2 Content – What is the quality of the content? What activities are being conducted? How is the content being distributed?
- 3 Calls to action – Who is the target audience and to what end?

By applying such criteria, analysts and policymakers can better distinguish between influence activities that are part of ordinary public discourse and those that undermine public trust, polarize societies or exploit informational ecosystems for strategic ends.

3.3.1

Frameworks for understanding and responding to harmful information

Some strategic approaches to adversarial information operations aim to pre-emptively weaken perceived adversaries by shaping global narratives and influencing public perception. One widely cited framework is the ‘4 D model’,²³ which outlines four core tactics:

- *dismiss* the allegations or the critic
- *distort* the facts or narrative
- *distract* attention away from the main issue
- *dismay* the audience through intimidation or fear-mongering.

These tactics are often executed through tools such as trolls, sock puppet accounts (see [Annex I: Glossary, on page 353](#)) and coordinated disinformation campaigns. By exploiting trust within users’ peer-to-peer networks, disinformation operatives encourage individuals to share false or manipulated content – giving it the appearance of credibility as it spreads organically. This social endorsement effect makes falsehoods appear more persuasive when they come from someone within a user’s own community.

The original 4 D model has been extended and incorporated in the DISARM Framework, which adds a fifth tactic – divide.²⁴ *Divide* seeks to exploit pre-existing social, political

or cultural tensions or divisions within or between communities. Two core principles underpin many of these tactics:

- **Believability:** effective falsehoods contain a grain of truth, build on existing biases or familiar narratives, subtly reinforcing what the audience already believes.
- **Extension:** damaging falsehoods are designed to persist, spreading across time and platforms. Even efforts at denial unintentionally amplify its visibility.

Social media algorithms exacerbate these dynamics by rewarding content that provokes strong reactions, especially outrage. But virality is incompatible with complexity: as content spreads, it strips away nuance and context, leaving behind simplified, polarizing messages that fuel division and distrust.²⁵

Frameworks such as the ABCDE Framework²⁶ provide a structured method for analysing influence operations, breaking them down into five key dimensions:

- **Actors** – Who is behind the operation? This includes originators, sources and content amplifiers.
- **Behaviours** – What tactics, techniques and procedures are being used?
- **Content** – What narratives are being pushed? Includes narrative themes, audience targeting and messaging strategies.
- **Degree** – How extensive is the operation? This measures reach, scale and amplification of content or narratives.
- **Effect** – What is the impact? This evaluates the resulting harm, influence or disruption caused.

Together, such frameworks can help practitioners understand not just how influence operations work, but also who is behind them, what they aim to achieve and why they matter in the broader information environment.

Attribution – the ability to identify the actor or entity responsible for creating or spreading harmful information – is one of the most difficult yet critical challenges in today's information environment. Harmful narratives often appear organic, circulating through local voices, anonymous accounts or recycled content, but may in fact be amplified or initiated by coordinated networks, political actors or commercial interests. Sophisticated tactics such as impersonation of trusted organizations, use of inauthentic accounts or AI-generated content further obscure origins, making attribution complex and contested.

Similar to cyber operations, public attribution of a malicious actor in harmful information operations carries real risks: it can escalate tensions and endanger staff. Attribution therefore requires a cautious, evidence-based approach that combines technical analysis, cross-sector collaboration and community insights. Importantly, the focus should be less on public naming and shaming, and more on understanding patterns, intent and impact to inform proportionate and principled responses.

3.4 Addressing harmful information, strengthening early warning and response

The spread of harmful information before, during and after crises is not new but the rapid circulation of unverified content on social media has made its effects far more acute, particularly when it interferes with crisis response. Harmful narratives often undermine trust in whether authorities or other actors are accurately representing the crisis, mislead communities about aid distribution, and distort the perceived severity of events.

In September 2024, Hurricane Helene (a powerful Atlantic hurricane) made landfall in the US, causing catastrophic damage and the reported loss of 251 lives. In the aftermath, harmful information spread rapidly online. False claims alleged that government disaster relief funds were being redirected to ‘house illegal immigrants’ or to unrelated international crises. Other rumours falsely stated that federal aid was capped at \$750 per person, required repayment or could result in the forfeiting of recipients’ homes. These narratives exploited public fear, uncertainty and prejudice, eroding trust, disrupting aid efforts and forcing authorities to divert time and resources into countering it.²⁷

Contributor Insight 3.4

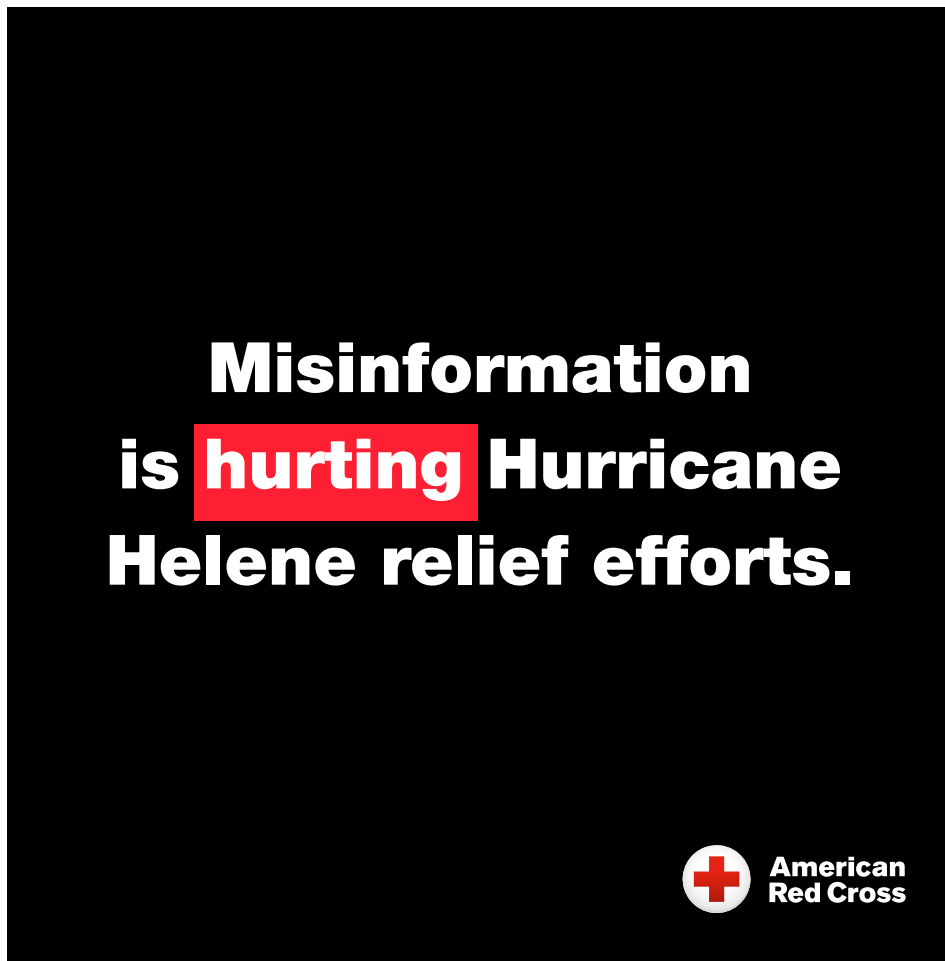


Misinformation surge during Hurricane Helene

In autumn 2024, the American Red Cross moved swiftly to prepare for and respond to Hurricane Helene as it brought devastation to communities across the south-eastern US and drew global attention.

On-the-ground response efforts were complicated by a surge in online misinformation. False claims circulated alleging the American Red Cross was confiscating donated items, taking over local relief efforts, or absent from affected communities – all untrue. For many people already unsure about where to seek help, these inaccurate claims only deepened uncertainty. To combat confusion, the National Society issued a direct public rebuttal, using a simple social media graphic and a clear statement on its website to set the record straight – providing clarity without amplifying false narratives.

Fig 3.1

Social media graphic used to rebut false claims, 2024

Underlying the communications response was a dual commitment: safeguarding the organization's reputation through calm and credible messaging, and shaping the public narrative through consistent storytelling that reaffirmed the humanitarian mission.

Social media listening team members tracked sentiment in real time, ensuring the response remained measured and emotionally intelligent. Strategic storytelling reinforced Red Cross operational efforts through media interviews and live updates from affected communities. Rather than stepping back, the American Red Cross leaned into its humanitarian mission.

Even in a hostile online climate, public support remained strong. Donations, volunteer interest and trust indicators held steady – or improved – reflecting the strength of an emotionally anchored, mission-driven supporter base. Years of trusted messaging and credibility helped buffer the organization against viral falsehoods.

The crisis emphasized the need to prepare not only for physical disasters but also for digital misinformation. Like logistical planning before a storm, the National Society continues to monitor online trends and social mentions, address emerging misinformation, deploy proactive storytelling and provide internal guidance to employees and volunteers

for awareness and alignment. These efforts aim to safeguard reputation and maintain public trust, especially among younger, digitally native audiences.

This challenge isn't unique to any one organization. In times of crisis, heightened emotions often fuel distortion. What Hurricane Helene revealed was not just the fragility of public perception, but also the need for intentional, value-driven reputation management.

Helene reaffirmed that rapid, compassionate communication matters. Communities look to the Red Cross not only for help in a crisis, but also for truth when confusion spreads. As misinformation evolves, our guiding principle remains the same: truth must always move faster than doubt or distortion.

Jodi Sheedy

Senior Director, Communications Strategy

American Red Cross

While disasters often cannot be prevented, robust early warning systems and effective emergency response are essential to minimizing impact. To be effective, warnings must reach at-risk populations rapidly, in the appropriate language and format, and with attention to diverse needs and contexts. Crucially, they must come from trusted sources so that people believe and act on them. Achieving this requires that early warning systems be co-designed with the communities they serve, in close collaboration with the local actors who play a central role in communicating and delivering these messages.²⁸



For me personally, what needs to be done to reduce fake news and strengthen preparedness for danger is to encourage people on a daily basis, without waiting for danger to arise, by informing them about what to do when a cyclone hits. And also to organize training for as many people as possible.”

Community member, Madagascar

Contributor Insight 3.5

Trusted warnings, timely action: Addressing a missing link in early warning systems

The Global Disaster Preparedness Center led a **research initiative across 15 locally led projects** in 14 countries,²⁹ examining the accessibility and actionability of early warning systems for 'last mile' communities – those hardest to reach and most vulnerable due to systemic social and structural challenges.

A consistent finding across contexts – from Brazil to Nepal, Albania to Malawi – was the central role of trust. The research showed that even scientifically accurate forecasts often fail to trigger early action when messages are mistrusted, sources are doubted or the information is confusing or inaccessible. This initiative is notable for centring community voices and documenting not only the barriers to effective early warning, but also practical strategies for building trust and improving the inclusivity and accessibility of early warning systems for underserved groups.

Key insights

Mistrust and misinformation did not emerge as abstract risks but as some of the most significant barriers to translating early warning into early action. In Brazil's coastal communities, **a lack of clear differentiation between official alerts and misinformation** created confusion and undermined confidence in the warnings. Without recognisable formats or clear source attribution, people struggled to verify which messages were credible and what guidance to follow.

Mistrust was particularly strong in areas where communities had repeatedly experienced 'false alarms' – warnings not followed by actual hazards. In Odisha, India, this **contributed to a 'cry wolf' effect**, leading many residents to ignore future alerts. Religious fatalism also played a role in eroding trust, with some individuals placing greater faith in divine protection than in institutional warnings. **Another study from India** highlighted how frequent alerts forced people who rely on daily labour wages to choose between heeding early action warnings or losing income, making trust in alerts especially critical when livelihoods are at stake.

Another common source of mistrust was the use of unclear or overly technical language. Studies from **Eswatini, Albania** and **Nepal** showed that early warnings often relied on jargon, unfamiliar terminology or were not translated into local languages. When messages are difficult to understand or feel irrelevant to local realities, they become inaccessible – and trust quickly breaks down.

Trust is a critical factor in whether communities act on early warnings. Addressing gaps in communication and strengthening confidence in official alert systems is essential to ensure that people respond to accurate information, especially when harmful or misleading messages circulate alongside official alerts.

Lessons and solutions

The research confirms that trust is foundational to effective early warning. Several strategies emerged to strengthen trust:

- **Leverage trusted messengers:** People are more likely to act on alerts delivered by individuals or organizations they know, such as community volunteers, religious leaders, local organizations or friends and family members. Familiarity builds credibility, highlighting the need to involve and train local networks to improve both reach and trust.
- **Make messages clear and locally relevant:** Alerts should use plain, non-technical language and be delivered in the local languages or dialects of the target community. Including visuals and culturally familiar references can help people better understand the messages, evaluate their risk and respond appropriately.

- **Ensure official information is easily recognisable:** Consistent formats, clear source attribution and delivery through trusted channels help communities trust and distinguish official alerts from misinformation. Strengthening digital literacy and working with local media can further reduce confusion and improve trust in risk communications.
- **Establish two-way communication:** Communities value feedback loops that allow them to ask questions, report issues and shape future messages. A participatory approach strengthens both trust and responsiveness to early warning systems. When combined with ongoing awareness-raising efforts and community education, it builds understanding of local risks and how early warning systems work – making people more likely to recognize and act on alerts when it matters most.

Vladislav Kavaleuski

Program Manager, Knowledge Management

IFRC, Global Disaster Preparedness Center

Contributor Insight 3.6

Strengthening trust, inclusion and community engagement in early warning, early action

In eastern Nepal, formal early warning dissemination channels have struggled to reach remote, disaster-prone communities. As a result, warning messages have not been well-understood, trusted or acted on. To address this gap, the Nepal Red Cross Society, with support from the Finnish Red Cross, has been implementing the Reducing Disaster Impact through Inclusive Preparedness and Anticipatory Action (REDI) project in 15 communities, reaching nearly 65,000 people since 2023.

The project has introduced an inclusive multichannel communication strategy to bridge the information divide. People-centred approaches and diverse communication channels help ensure that everyone receives warning messages in time to react. Protocols have been developed for rapid communication by local government authorities and stakeholders, and sirens are being installed in strategic locations. Nepal Red Cross Society volunteers conduct door-to-door visits and organize information sessions to reach individuals who may be inaccessible via SMS alerts or social media, or who simply prefer face-to-face communication.

Several lessons have emerged on how to enhance early warning systems and build trust. Locally led, inclusive systems are more effective because people see themselves reflected in them. Local knowledge and familiar communication practices provide a strong foundation. During the last monsoon season, for example, upstream communities monitored rainfall and alerted downstream communities to take immediate action, saving lives.

Real-life experiences and community drills have improved awareness of evacuation procedures, safe spaces and individual responsibilities. They have also fostered trust that no one, especially those in vulnerable situations, will be left behind. Addressing urgent needs, such as strengthening irrigation channels and constructing retaining walls, was crucial in establishing initial trust within communities. Ongoing awareness raising on weather forecasts, their limitations and the function of early warning systems remains crucial.

Importantly, people's mindsets have shifted from reactive to proactive, enabling earlier action based on trusted, localized information and growing confidence in their ability to respond. When a community is cohesive and empowered to drive its own development, people become active agents of resilience. This empowerment spreads across the community, making it safer and more resilient for everyone.

Sagar Shrestha

Director, Disaster
Management Department

Nepal Red Cross Society

Sushma Shrestha

Country Manager for
Nepal and Bhutan

Finnish Red Cross

Paula Uski

Senior Specialist
on Climate Change
and Disaster Risk
Management

Finnish Red Cross

Mari Koistinen

Senior Specialist on
Protection, Gender
and Inclusion (PGI)

Finnish Red Cross

Ika Trijsburg from the University of Melbourne, in her research on disinformation in disasters in cities, highlights that local government is particularly well placed to counter disinformation due to its close proximity to communities and central role in shaping daily life. Its nearness to residents gives local government a nuanced understanding of local issues and grievances, allowing them to design interventions that are locally embedded and context specific. As the level of government most directly responsible for policy decisions and service delivery, it holds a unique position of influence. Moreover, city authorities are often perceived as the most trusted level of government, a critical advantage in the fight against disinformation. Trust is essential for ensuring that factual information is received, accepted and acted on. Strengthening this trust at the local level enhances community resilience against false or misleading narratives, especially during crises (Trijsburg, 2023).³⁰

In framing preventive strategies for disaster-related harmful information, Trijsburg highlights the importance of understanding community risk factors and deploying coordinated strategies that focus on:

- **Pre-emption:** Disinformation typically exploits existing political and societal tensions, encouraging people to buy into false narratives.
- **Prebunking:** Disaster-related disinformation usually follows known narratives such as climate denialism, climate delayism, government distrust and prejudice. Prebunking involves proactively introducing information before disasters occur, helping communities recognize and refute false content.
- **Debunking:** When misinformation emerges, timely and simple corrective messaging – grounded in clear, accessible facts – can effectively counter it.



Prebunking

The practice of anticipating false or misleading claims before they spread and preparing people with accurate information, critical thinking tools or warnings in advance.



Debunking

A reactive strategy that addresses harmful information after it has begun to circulate. Its effectiveness depends on being timely, clear and delivered by trusted messengers. Effective debunking not only identifies a claim as false but also explains why it is false.

Contributor Insight 3.7

Using social listening to guide prebunking strategies in Scotland

This initiative focused on anticipating misinformation before it spreads by equipping Scottish health officials with the skills to apply social listening insights to their vaccine communication strategies. Through World Health Organization Europe (WHO/Europe)-led training, participants learned how to identify early warning signs, such as rising concerns, shifts in tone and emotionally charged language, and use these insights to design messages that build critical thinking and resistance before misleading claims take root.

Unlike debunking, which is reactive, prebunking is preventive. It prepares audiences to recognize potential misconceptions and question misleading claims before they take hold. In the Scotland training, the focus was on identifying early signals of hesitancy and understanding psychological factors, such as confirmation bias, that make certain narratives more persuasive than others. This proactive, audience-specific messaging was especially important in protecting public trust as human papillomavirus (HPV) vaccine uptake was beginning to decline.

WHO/Europe worked with Scotland's Ministry of Health to train health officials and physicians in the fundamentals of using social listening to inform their outreach. The training helped participants understand what makes misinformation persuasive and how to proactively address vulnerabilities. As a result, participants were better equipped to craft nuanced, evidence-aligned messages that pre-empted doubt among hesitant groups.

Prebunking works best when rooted in real-time insights. By strengthening institutions' capacity to anticipate and address risks early, this initiative showed how listening-led approaches can build trust, especially where health guidance must evolve with

public concerns. It offers a practical, scalable model for embedding resilience in vaccine communication systems.

Nancy Claxton

Regional Training Officer
in Risk Communication,
Community Engagement
and Infodemic
Management

WHO/Europe

Cristiana Salvi

Regional Technical Advisor
for Community Resilience
and Protection

WHO/Europe

Paco Pangalangan

Infodemic Management
Consultant

WHO/Europe

Contributor Insight 3.8

Training medical students in Moldova to debunk vaccine misinformation

Medical students and public health specialists in Moldova were brought together to build practical skills for addressing false health claims already circulating in communities. What made it stand out was its real-world application: trained students went door-to-door to speak directly with community members, correcting vaccine myths and sharing accurate information through open conversational dialogue.

This was a clear example of reactive yet responsible correction. Effective debunking is not just about identifying falsehoods, it's about how you respond. The training focused on clear, evidence-led communication that avoided amplifying myths, instead starting and ending with facts. Participants learned to explain why specific claims were false in ways that maintained trust and reduced confusion. This hands-on, myth-busting approach supported vaccine uptake during a time of multiple, overlapping crises.

The WHO Country Office in Moldova, in collaboration with WHO/Europe, organized the training for 30 participants including medical students, communication professionals and public health experts. After the training, participants engaged in community outreach, addressing people's concerns directly and offering guidance tailored to local narratives and needs.

Debunking is most effective when it is personal, contextual and trust-based. Training future frontline workers to respond effectively after falsehoods have spread builds long-term institutional capacity. The model is replicable – particularly in contexts where misinformation is already widespread and direct community engagement is possible.

Leonardo Palumbo

Community Engagement
Technical Officer

WHO/Europe

Paco Pangalangan

Infodemic Management
Consultant

WHO/Europe

Cristiana Salvi

Regional Technical Advisor
for Community Resilience
and Protection

WHO/Europe

In Eswatini, as in many other contexts, entrenched rumours and misinformation, spread both through word of mouth and social media, posed a major challenge to the COVID-19 response. Delays in mobilizing partners and scaling up interventions created a critical gap, during which rumours gained traction, eroding public trust and fuelling vaccine hesitancy, especially among younger populations. To address this, the structured coordination of a Risk Communication and Community Engagement (RCCE) pillar created a platform for cross-sector collaboration and more effective response strategies. Central to this effort was a focus on dynamic community listening – ensuring that local concerns were actively heard, analysed and fed into decision-making processes. This community-driven approach enhanced the credibility of interventions and built greater trust in public health messaging. Identified priorities included keeping these listening systems active, adaptive and inclusive, while continuing efforts to address practical barriers to vaccine access to ensure that accurate information is matched by equitable service delivery.³¹



It needed a number of teams who moved home to home sensitizing people just like the way they do these mass immunization sessions. Because when you move home to home, you have access to meet most of the people in their homes. So you can talk to them in the comfort of their compounds. So this actually helps a lot.”

Community member, Uganda

3.5 Risk communication and community engagement

RCCE³² is a cornerstone of the IFRC’s approach to managing public health crises and disasters. Developed jointly with WHO and UNICEF through the Collective Service,³³ RCCE combines clear, timely communication with meaningful engagement and listening strategies to build trust, counter harmful information and support effective action. It involves:

- **timely, accurate and actionable information** delivered through trusted community channels to help people understand risks and take informed decisions
- **community feedback systems** that enable local actors to gather, analyse and respond to rumours, questions, concerns and misinformation
- **two-way dialogue** to foster local ownership and ensure that responses reflect diverse needs and experiences, particularly from groups that are marginalized
- **Data for Action frameworks**³⁴ that integrate community insights into operational decision-making and support adaptive, responsive programming

- **collective coordination**, aligning and supporting national and local actors in their efforts to meaningfully engage with communities.

This approach has proven especially effective in fragile and complex contexts, where trust and community participation are essential. It reinforces the understanding that communication is not a one-way flow of information, but a dynamic exchange grounded in respect, transparency and shared responsibility.

Contributor Insight 3.9

WHO: Building resilient emergency systems (part 1 of 2)

Risk communication, community engagement and infodemic management (RCCE-IM) are central to effective public health emergency management. Mandated by the International Health Regulations, RCCE-IM is a strategic pillar of preparedness and response, influencing how societies navigate crises and protect their most vulnerable populations.

Since 2020, the WHO/Europe has responded to a series of complex emergencies, including the COVID-19 pandemic, disasters and armed conflicts such as the war in Ukraine, the 2023 earthquake in Türkiye and the mpox outbreak in Central Africa. In every crisis, people's actions and decisions, shaped by trust and access to information, have been central to achieving improved health outcomes. Public health measures only succeed when communities trust the guidance and the messengers who work with them to co-develop and co-deliver interventions. RCCE-IM helps bridge this gap by meaningfully engaging diverse communities, with a particular focus on those most at risk at each step of preparedness and response. A key development has been the inclusion of infodemic management as a core component of RCCE, monitoring health narratives to identify and counter false information, thereby helping to ensure that people can make informed decisions to protect their health in an emergency.

Core lessons learned

Through numerous health responses, WHO/Europe has identified key lessons on how RCCE-IM improves emergency response and saves lives – insights which can inform capacity-building efforts, including:

- 1 **Embedding and sustaining RCCE-IM:** Effective RCCE-IM must be institutionally embedded across all phases of emergency management, from prevention and preparedness through to response and recovery. It is a necessary component in preparing for crises and supports response when there is sustained investment in infrastructure, human resources, training and systems by health authorities. Experience during COVID-19 and earlier crises showed that countries with established, well-resourced RCCE-IM teams could act faster, owing to pre-existing public trust.

- 2 Multisectoral and community engagement:** No single sector holds all the answers in a crisis. RCCE-IM delivers the greatest impact when it is coordinated across sectors including health, education, civil protection and emergency services – and across all levels of government administration. Collaboration with community members and local stakeholders is also key. Trusted figures include faith leaders, youth advocates, family doctors, community health workers, community-based groups, civil society organizations (CSOs) and educators. All play an important role. When meaningfully engaged, they can significantly increase reach, credibility and acceptability of messaging, because it is co-developed with and for the communities they serve. Involving a diverse range of stakeholders ensures that communication efforts are relevant for at-risk populations and marginalized groups, thus increasing the likelihood of acceptance of official guidance.
- 3 Evidence-led, technical approach:** RCCE-IM is an applied discipline that sits at the intersection of public health, risk communication and behavioural and social sciences. Successful interventions are grounded in:

 - ongoing community and social listening to understand evolving concerns and beliefs of different demographics and groups
 - strategies that bridge risk assessment and risk perception
 - operational research, behavioural insights and social science evidence.

By using real-time data and feedback, risk communication strategies can be rapidly adapted to audience needs, changing public health guidance, emerging misinformation and disinformation threats, and barriers to action. Integrating scientific rigour, monitoring and evaluation into RCCE-IM is essential to ensure relevance and effectiveness.

- 1 Transparency and communicating uncertainty:** Trust between health authorities, stakeholders and those it aims to serve is essential. A key determinant is early and transparent communication, even if all information is not yet known. The dynamic, fast-changing nature of emergencies means that public health advice will evolve as new evidence emerges. Clearly communicating this uncertainty and updating guidance in a timely manner is crucial to maintaining credibility.
- 2 Two-way communication and social listening:** People are not passive recipients of information, especially in a crisis. They experience anxiety and hold beliefs and attitudes that influence their decisions. Understanding what people think, feel and need is essential to craft messages and responses that resonate and are effective. Mechanisms for real-time feedback (hotlines, social media, community dialogue – both online and offline) allow for responsive risk communication and rapid false information management. Effective social listening enables authorities to detect false or misleading information early. It enables correction of falsehoods to prebunk information before it can take root and address the issues that matter most to communities.
- 3 Managing misinformation and disinformation:** The scale and speed with which rumours, myths and false information circulate, especially through

digital platforms and social media, can rapidly fuel confusion, sow distrust and undermine public health efforts. Infodemic management focuses on the proactive dissemination of accurate, timely public health information rather than only correcting misinformation and disinformation once detected. Multidisciplinary teams which include health authorities, other sectors, local authorities, CSOs, community leaders and key stakeholders must work together to identify and address information gaps. Through regular, tailored messaging and advocacy, these actors can help protect the public from misinformation and reinforce clear, actionable guidance.

Proactive infodemic management involves:

- routine monitoring for viral falsehoods and knowledge gaps
- debunking false information that poses a risk to public health
- prebunking – pre-emptively addressing likely untrue and/or damaging rumours, and collaborating with trusted influencers and communication channels to disseminate public health advice and updates.

Nancy Claxton

Regional Training Officer
in Risk Communication,
Community Engagement
and Infodemic
Management

WHO/Europe

Leonardo Palumbo

Community Engagement
Technical Officer

WHO/Europe

Paco Pangalangan

Infodemic Management
Consultant

WHO/Europe

Cristiana Salvi

Regional Technical
Advisor for Community
Resilience and Protection

WHO/Europe

Contributor Insight 3.10



Rumours and misinformation in epidemics: Insights from the French Red Cross Foundation

Rumours and misinformation can have serious consequences during epidemics. The spread of unverified, false or inaccurate information about the causes, transmission and prevention of diseases, even when not intentional, can undermine public health messaging and practices – hindering public health response efforts and leading to poorer health outcomes. Over the past decade, the French Red Cross Foundation has supported more than 30 research projects on access to healthcare and epidemic response, both in France and internationally. These include studies on the Ebola virus disease outbreak in Guinea, the plague in Madagascar and COVID-19 in France, the Comoros and Senegal. These studies have generated valuable insights into some of the links between rumours, distrust and resistance to public health guidance provided by humanitarian actors.

The circulation of false narratives, such as claims that viruses are political conspiracies or that preventive measures spread diseases, fuels suspicion and drives public distrust

during epidemics. Such harmful narratives can become obstacles to the implementation of public health measures by humanitarian actors and lead to serious outcomes – ranging from refusals of disinfection or vaccination to delayed seeking of treatment, reduced reporting of cases or even attacks on humanitarian personnel. Each of these consequences contributes to the continued spread of diseases. In this context, research shows that locally led and culturally adapted community awareness programmes are key for (re) building trust in epidemics. These programmes are most effective when run by trusted community members and volunteers, co-developed with communities and aligned with local cultural practices.

Drawing on this experience, the French Red Cross Foundation reasserts the importance of both quantitative and qualitative social science research to counter harmful information, support trust in the humanitarian sector and strengthen epidemic preparedness and response programmes. Far from being a luxury, research that is robust, participatory and community based appears essential to detect emerging narratives, inform context-specific responses and ultimately save lives.

French Red Cross Foundation³⁵

Concluding remarks: Navigating humanitarian action in a post-trust information era

Humanitarian action now unfolds within a highly contested information environment, characterized by Barclay as a ‘post-trust culture’,³⁶ where truth plays a diminished role in shaping human decision-making. In this environment, technology amplifies and instrumentalizes information, allowing harmful narratives to spread quickly, especially where reliable information is scarce.

The threat landscape is complex and diverse, involving state agencies, proxy groups, non-state actors, hacktivists, criminal networks, bots, click farms, coordinated collectives and individuals. Such actors exploit polarization, fuel mistrust and disrupt humanitarian response. Harmful information is no longer just a communications challenge, it is a crisis affecting humanitarian access and action, as well as safety of staff and volunteers. Addressing it requires a shift in mindset and a recognition that trust, truth and information access are foundational to humanitarian impact.

Misinformation and disinformation remain among the top risks in the World Economic Forum’s *Global Risks Perception Survey*, underscoring the continuing scale and sophistication of false content. This trend intersects with political and societal polarization, algorithmic amplification and deepening digital divides. Without deliberate efforts and cooperation to counter human and systemic bias, algorithmic models risk producing harmful and unjust outcomes, particularly in communities living in fragile and marginalized situations.³⁷

To respond effectively, humanitarian organizations must prioritize three actions:

- 1 Equip communities and volunteers with accurate, trusted information and preparedness tools that strengthen local resilience and help counter the influence of harmful narratives.
- 2 Invest in community-led strategies that elevate local voices, feedback and leadership, particularly where digital inclusion and information access remain limited.
- 3 Advance global cooperation and policy coherence, including ethical AI governance, platform accountability and protection frameworks for humanitarian information systems.

Humanitarian action must be grounded not only in principles, but also in a nuanced understanding of today's information dynamics. Navigating this environment requires collective resolve, cross-sectoral innovation and a renewed commitment to trust, inclusion and integrity in the digital age.

[Chapter 6, on page 211](#) explores inoculation theory and other key strategies for responding to harmful information.

Asks, aims and recommendations

Asks

Strengthen global cooperation on platform accountability, AI governance and the protection of humanitarian information by embedding human rights, preparedness, accountability and collaboration into legal and policy frameworks. Invest in community-led resilience, preparedness and feedback systems that empower all community members, ensure access to information and protect vulnerable populations – applying a clear rights-based approach in practice.

Aims

Prevent harmful information from undermining humanitarian access and action.

Strengthen community resilience: equip communities and volunteers with trusted information, preparedness tools and locally grounded strategies to counter harmful narratives.

Advance policy coherence by embedding humanitarian perspectives in digital and AI governance.

Recommendations

States and policy-makers

- Champion the protection of humanitarian information within legal and policy frameworks for AI governance and platform accountability.
- Counter harmful narratives that stigmatize communities and/or principled humanitarian organizations, while avoiding the politicization of humanitarian action.
- Reaffirm the importance of safe, unhindered humanitarian access

and promote understanding of humanitarian mandates, including the auxiliary role of National Societies as independent yet recognized partners to public authorities.

- Engage with humanitarian actors to ensure timely, accurate and locally relevant communication strategies that support humanitarian action.
- Ensure national information laws and policies comply with international law.

Technology platforms

- Provide locally adapted tools, including multilingual translation, culturally relevant fact-checking and accessibility features.

- Amplify verified humanitarian information in crises by engaging with humanitarian and community actors.

- Monitor how humanitarian narratives are reshaped locally and mitigate harmful reinterpretations.
- Adapt and/or cool algorithms during crises to reduce amplification of harmful narratives.
- Strengthen partnerships with humanitarian organizations and community media to ensure reliable, timely and life-saving messages reach affected populations.

Humanitarian actors

- Co-create inclusive, locally relevant messages with communities, journalists and trusted influencers for contextual, cultural and linguistic relevance.
- Monitor perceptions of humanitarian action and adapt real-time messaging across online and offline channels.
- Build resilience through training, capacity-building and embedding harmful information management in humanitarian diplomacy, risk and behaviour change.
- Document and share insights and lessons learned on combating harmful information to support continuous adaptation and strengthen collective resilience.

Community and local leaders

- Contextualize and disseminate verified humanitarian information in ways that reflect community priorities.
- Engage communities directly through peer-to-peer networks, dialogue and locally led initiatives to address harmful information.
- Identify trust gaps and communicate concerns directly to humanitarian actors and authorities.

Endnotes

- 1 The Red Cross Red Crescent (RCRC) Global Migration Lab, hosted by the Australian Red Cross and Kenya Red Cross Society, conducts research on topics of critical importance in providing humanitarian assistance and protecting migrants in vulnerable situations to inform the International Red Cross and Red Crescent Movement's operational programming and humanitarian diplomacy.
- 2 RCRC Global Migration Lab. *Locked down and left out? Why access to basic services for migrants is critical to our COVID-19 response and recovery.* (2021) www.redcross.org.au/globalassets/cms/documents/migration-services/en-exec-sum-rcrc-global-migration-lab-locked-down-left-out-covid19.pdf; RCRC Global Migration Lab. *Sight Unseen: A vision for effective access to COVID-19 vaccines for migrants.* (2021) www.redcross.org.au/globalassets/cms/documents/migration-services/rcrc-gml-sight-unseen-covid19-vaccines-final.pdf; RCRC Global Migration Lab. *Migrants' Perspectives: Building Trust in Humanitarian Action.* (2022) www.redcross.org.au/globalassets/cms/global-migration-lab/gml-migrants_buildtrust_english.pdf; Red Cross Red Crescent Global Migration Lab and ICRC Central Tracing Agency. *Towards Safer Journeys: Migrant women and children's experiences of separation, going missing or dying.* (2025) www.redcross.org.au/globalassets/cms/global-migration-lab/towards-safer-journeys_summary-report_eng.pdf
- 3 Ibid (2021)
- 4 RCRC Global Migration Lab. *Sight Unseen: A vision for effective access to COVID-19 vaccines for migrants.* (2021) www.redcross.org.au/globalassets/cms/documents/migration-services/rcrc-gml-sight-unseen-covid19-vaccines-final.pdf
- 5 See, for example, Loomba, S., de Figueiredo, A., Piatek, S.J. et al. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour.* 2021;5, 337–348. <https://doi.org/10.1038/s41562-021-01056-1>; Knights, F., Carter, J., Deal, A. et al. *Impact of COVID-19 on Migrants' Access to Primary Care: A National Qualitative Study.* medRxiv. 2021. DOI:10.1101/2021.01.12.21249692; Deal A, Haywood, S.E., Huda, M. et al. *Strategies and action points to ensure equitable uptake of COVID-19 vaccinations: a national qualitative interview study to explore the views of undocumented migrants, asylum seekers, and refugees.* *J Migr Health* 2021;4:100050. doi: 10.1016/j.jmh.2021.100050; ABC News. 'Misinformation about COVID vaccines is putting Australia's diverse communities at risk, experts say.' 4 March 2021. www.abc.net.au/news/2021-03-04/covid-19-vaccine-misinformation-cald-communities/13186936; ABC News. 'Multilingual women are countering vaccine hesitancy in Victoria's culturally diverse communities.' 16 May 2021. www.abc.net.au/news/2021-05-16/workers-hired-to-counter-vaccine-hesitancy-migrant-communities/100141280
- 6 RCRC Global Migration Lab. *Migrants' Perspectives: Building Trust in Humanitarian Action* (2022) www.redcross.org.au/globalassets/cms/global-migration-lab/gml-migrants_buildtrust_english.pdf
- 7 Red Cross Red Crescent Global Migration Lab and ICRC Central Tracing Agency. *Towards Safer Journeys: migrant women and children's experiences of separation, going missing or dying – summary report.* (2025) www.redcross.org.au/globalassets/cms/global-migration-lab/towards-safer-journeys_summary-report_eng.pdf
- 8 Kaldor Centre for International Refugee Law. *Countering Misinformation about Refugees and Migrants.* (2025) www.unsw.edu.au/content/dam/pdfs/law/kaldor/resource/2025-03-countering-misinformation-refugees.pdf
- 9 Joint Research Centre. *Public discourse on migration shaped by misinformation and conspiracy theories.* (2025) https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/public-discourse-migration-shaped-misinformation-and-conspiracy-theories-2025-06-05_en
- 10 Insecurity Insight. *The Shrinking Humanitarian Space on Social Media: Insights from Burkina Faso, Mali and Niger.* (2025) <https://insecurityinsight.org/wp-content/uploads/2025/09/Learning-from-Social-Media-Narratives-in-the-Sahel.pdf>
- 11 Council of Delegates of the International Red Cross and Red Crescent Movement, Resolution CD/24/R-NIIHA, "Neutral, Independent and Impartial Humanitarian Action," including the Appeal to States, adopted 2024. Available at: https://rcrcconference.org/app/uploads/2024/09/CoD24_8DR-Draft-Res-NIIHA-EN.pdf.
- 12 The Council of Delegates of the International Red Cross and Red Crescent Movement includes the ICRC, IFRC and 191 National Red Cross and Red Crescent Societies. <https://rcrcconference.org/council-of-delegate/>
- 13 See, for example, the IFRC. *Principles and Rules for Red Cross and Red Crescent Disaster Relief.* (1986) p.4. Information quality is thematised, with a focus on cooperation between relief actors, in Disaster Relief Commission of the Red Cross and Red Crescent Movement. *Report of the Disaster Relief Commission.* (1991) p.6; IFRC. *International Disaster Response Laws.* (2004) www.ifrc.org/sites/default/files/2021-07/Principles_Rules_Red_Cross_Red_Crescent_Humanitarian_Assistance_EN.pdf
- 14 See ICRC. *Harmful Information: Misinformation, Disinformation and Hate Speech in Armed Conflict and Other Situations of Violence: ICRC Initial Findings and Perspectives on Adapting Protection Approaches.* (2021) p.11 www.icrc.org/en/publication/4556-harmful-information-misinformation-disinformation-and-hate-speech-armed-conflict
- 15 34th International Conference of the Red Cross and Red Crescent. Resolution 2: Protecting civilians and other protected persons and objects against the potential human cost of ICT activities during armed conflict. (2024). 34IC/24/R2
- 16 World Economic Forum. *Global Risks Report 2025.* (2025) https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf
- 17 Ibid. In answer to the question: "Please estimate the likely impact (severity) of the following risks over a 2-year and 10-year period."
- 18 Ibid, p.34–35
- 19 The first 'Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security' was established under UN General Assembly Resolution 73/27 (2018) and held its mandate from 2019 to 2021. The second group (Open-Ended Working Group on Security of and in the Use of Information and Communications Technologies 2021–2025) was established under UNGA Resolution 75/240 (2020) and was mandated to the end of 2025.
- 20 UN Open-Ended Working Group on Security of and in the Use of Information and Communications Technologies (OEWG), *Summary Report of the Chair, 2023*, para. 21. States also raised concerns over the growing use of malicious ICT-enabled information campaigns targeting other states – including those in transition or post-conflict – which undermine trust, threaten peace and security, and may directly or indirectly harm individuals. OEWG. Letter from the OEWG Chair. https://docs-library.unoda.org/Open-Ended_Working_Group_on_Information_and_Communication_Technologies_-_2021/Letter_from_OEWG_Chair_10_July_2025.pdf. para. 22
- 21 The European External Action Service (EEAS) defines foreign information manipulation and interference as "a pattern of behaviour that threatens or has the potential to negatively impact values, procedures and political processes," often involving deceptive tactics in the information space. See EEAS. *Information Integrity and Countering Foreign Information Manipulation & Interference (FIMI).* (2023) www.eeas.europa.eu/eeas/information-integrity-and-countering-foreign-information-manipulation-interference-fimi_en
- 22 Meta, Transparency Center — Transparency Reports and Threat Reporting, <https://transparency.meta.com/>
- 23 Nimmo, B. 4 D's of Disinformation (CyberwarCon), <https://www.cyberwarcon.com/ben-nimmo>

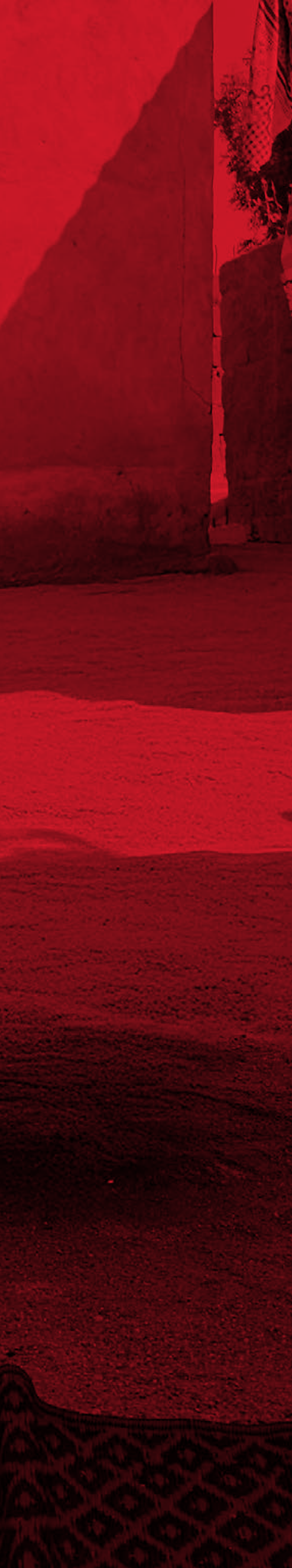
- 24 See DISARM Foundation. DISARM Framework. www.disarm.foundation/framework, <https://disarmframework.herokuapp.com>
- 25 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) pp. 206, 208–209
- 26 Pamment, J. The EU's Role in Fighting Disinformation: An EU Disinformation Framework. Future Threats, Future Solutions; No. 2. Lund University Publications. <https://lup.lub.lu.se/record/9d229c4e-5705-4ac0-b81f-c0f41579392a>
- 27 Craig, J. 'I saw the Hurricane Helene response up close. This is how disaster relief actually works.' Vox, 17 November 2024. www.vox.com/future-perfect/384734/hurricane-helene-asheville-response-fema-volunteers-climate-change
- 28 IFRC. *The Cost of Doing Nothing: The Humanitarian Price of Climate Change and How it can be Avoided* (2019) p.24 www.ifrc.org/sites/default/files/2021-07/2019-IFRC-CODN-EN.pdf
- 29 Albania, Bangladesh, Brazil, Eswatini, Ethiopia, India, Indonesia, Kenya, Malawi, Namibia, Nepal, Nigeria, the Philippines and Viet Nam.
- 30 Trijsburg, I. 'Disinformation Thrives on Division in Our Cities.' Pursuit, University of Melbourne. 31 August 2023. <https://pursuit.unimelb.edu.au/articles/disinformation-thrives-on-division-in-our-cities>.
- 31 IFRC Community Engagement Hub. National-Level Risk Communication and Community Engagement Coordination in COVID-19: Case study from Eswatini. (2022) https://communityengagementhub.org/wp-content/uploads/sites/2/2022/05/20220525_SolidarityFund_CaseStudies_Eswatini.pdf
- 32 IFRC. Risk Communication and Community Engagement (RCCE) Action Plan Guidance COVID-19 Preparedness and Response. (2020) www.who.int/publications/i/item/risk-communication-and-community-engagement-%28rcce%29-action-plan-guidance
- 33 WHO. 'The Collective Service: supporting RCCE in emergencies.' www.who.int/news/item/07-12-2021-the-collective-service-supporting-rcce-in-emergencies
- 34 WHO. The Collective Service. *Data Handbook for Risk Communication and Community Engagement (RCCE)*. (2022) www.rcce-collective.net/resource/rcce-data-for-action-handbook
- 35 Founded in 2010, the French Red Cross Foundation is an entity dedicated to humanitarian and social research. It reflects the French Red Cross' commitment to mobilizing scientific knowledge, ethical reflection and social innovation to support its actions. The Foundation fosters the production of scientific knowledge through academic partnerships and postdoctoral fellowships awarded to independent social scientists. It is an active member of RC3, the research consortium of the International Red Cross and Red Crescent Movement.
- 36 Barclay, DA. *Disinformation: The Nature of Facts and Lies in the Post-Trust Era*. (2022) p.xv
- 37 World Economic Forum. *Global Risks Report 2025*. (2025) https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf





Chapter 4

From context to consequence: Humanitarian sector voices on the impact of harmful information





Chapter 4

From context to consequence: Humanitarian sector voices on the impact of harmful information



Contents

	Introduction: Harmful information and operational realities	139
4.1	Narratives of harm	142
4.2	Harms and impacts of harmful information on humanitarian action	143
4.3	<i>Do no harm</i> in a harmful information age	150
4.4	Vulnerabilities and amplifiers	152
4.5	From fragmentation to focus: Avoiding overwhelm and building humanitarian capacity	153
4.6	When viruses go viral: Building resilience against harmful information in a pandemic	157
	Concluding remarks: Navigating a hostile information landscape	158
	Endnotes	162

Introduction: Harmful information and operational realities

We are living through a moment in history as consequential as the invention of the printing press: one that is fundamentally altering how information is produced, distributed and trusted. As with the printing revolution, today's digital technologies are redistributing power, accelerating the spread of ideas and challenging traditional gatekeepers of knowledge. They have also enabled the rapid proliferation of harmful information and narratives. The sheer volume and velocity of contemporary information flows, combined with unprecedented levels of internet access and connectivity, are having profound societal impacts. Information has become simultaneously a commodity, an asset and a risk.

This transformation directly affects humanitarian action, both the sector itself and the populations it serves. The widespread adoption of digital tools and platforms, unfolding at a pace and scale unimaginable just a decade ago, is reshaping how humanitarians operate and how communities experience and interpret information in crises settings.

In this new landscape, an urgent question arises: has information itself become an impediment to humanitarian response? This chapter draws on the perspectives of humanitarian practitioners to offer critical insights into how harmful information manifests in operational contexts – shaping decisions, straining relationships with communities and exposing challenges and risks that increasingly undermine humanitarian action.

◆ A connected world, a complicated reality

By 2024, around 5.5 billion people (68% of the global population) were online, according to the International Telecommunication Union (ITU). The number of websites has now surpassed 1.2 billion¹ and globally over 5.24 billion people are using social media,² a rise largely driven by increased mobile broadband access in emerging economies and developing countries.

This shift is reflected in platform demographics: Facebook, once concentrated in high-income countries, now counts nine of its ten largest user bases in middle-income countries. Among the top 20 user bases, 14 are in middle-income countries, including 9 in lower-middle-income ones. This geographical and socioeconomic transition has not only expanded access to information, but also its flow and character.

TikTok exemplifies this evolution. As of 2025, the platform has around 1.6 billion monthly active users.³ Its rise, also fuelled by mobile access, follows a similar pattern: nine of its top ten user bases are in middle-income countries. The platform is especially popular among younger audiences, with around 40% of users aged 18–24 years. TikTok, like other social media platforms, began as a space for creative entertainment, but has now evolved into a powerful arena for shaping narratives, mobilizing opinion and spreading harmful information.

Despite these sweeping changes, the digital divide remains a stark global reality. According to the ITU, 2.6 billion people – disproportionately women, older adults and

people living in the least developed countries – remain offline. This gap continues to reinforce global inequalities in access to information, services and opportunities.

Since the advent of the mobile phone, the gap between humanitarian responders and crisis-affected communities has steadily narrowed. Technological innovations over the last two decades have also revealed shortcomings in how humanitarian organizations listen, gather information and ensure accountability to affected populations – including through genuine feedback loops. Today, humanitarian actors can no longer choose whether to engage with communities, only how to do so. Yet in the digital age, a fundamental challenge lies in determining who is actually voicing their opinions. In digital spaces where identity, authenticity and influence are increasingly blurred, it is often unclear whose voices are being heard. Those affected by a crisis encompass a broad array of constituents from local communities to diaspora populations and distant individuals or networks with vested interests – all shaping the narrative.

At the same time the information ecosystem has become polluted, at times deeply toxic. Everyone now has a platform, including those with little understanding of the context or with deliberate intent to mislead or harm. This pollution may be organic or orchestrated, raising urgent questions about credibility, accountability and voice in the digital humanitarian environment. In today's noisy and fragmented information environment, it is harder than ever to determine whether humanitarian actors are truly listening to – and able to hear – the people they aim to serve. The risks of misreading the information landscape have never been higher and the consequences of getting it wrong more severe.



It's not just AI-generated images – they also reused old background photos from the Nepal earthquake. Some people hadn't been to the disaster site at all, but with today's traffic model, online views equal money. Many people are profit-driven, so they take old photos, feed them into GPT, add some current text and produce fake images."

Community member, China

Local, national and international institutions – especially those operating in fragile contexts or in politically unstable environments – are increasingly vulnerable to false allegations and harmful narratives. Humanitarian organizations, in particular, have faced a rising tide of harmful information flows in recent years, mirroring broader global trends. Such dynamics threaten not only the credibility and safety of humanitarian workers, but also the legitimacy of institutions, the trust of communities and ultimately the effectiveness of humanitarian action. On social media, inaccurate or misleading content circulates, often without verification, context or oversight, creating fertile ground for harm. In fragile institutional settings where trust may already be weak, such narratives and immediacy can deepen internal divisions and exacerbate polarization.

A recent example involved the creation of Facebook pages with names suggesting affiliation with a component of the International Red Cross and Red Crescent Movement. Managed anonymously, these pages regularly published content purporting to reveal insider information, including allegations of corruption, leadership dysfunction or internal conflict. While appearing to serve as whistleblowing platforms, their posts

were often based on partial or decontextualized information or documents. By using the Red Cross or Red Crescent emblem and adopting language that echoed humanitarian values, these pages cultivated an image of credibility and legitimacy – misleading both the public and members of the National Society and ultimately damaging the National Society’s reputation.

Such content – particularly video – is significantly more likely to mislead and go viral than text or images, amplifying their impact and reach. A notable example is the so-called ‘cash-in-trunks’ video,⁴ which has resurfaced repeatedly since 2011 in various contexts, despite consistent and unequivocal statements from the ICRC refuting its authenticity.

Humanitarian actors themselves have become both targets and, at times, inadvertent vectors of harmful information. Staff and volunteers may unknowingly share unverified content through informal networks, contributing to a climate of suspicion and discontent. Meanwhile, individuals or groups operating through unofficial channels may misuse or exploit the Red Cross and Red Crescent emblems, as well as the principled language of humanitarianism to pose as credible voices, misleading audiences and extending the reach of harmful narratives.

Contributor Insight 4.1

Role of the National Observatory and strategies to address misuse of the emblem

To monitor, map and address misuse of the Red Cross emblem across Italy, the Italian Red Cross established a National Observatory on the Protection of the Emblem in 2018. This body is responsible for carrying out advocacy efforts and corrective actions for cases of misuse that have national impact, as well as coordinating the responses of local branches. Italian Red Cross staff and volunteers, as well as any concerned citizen, can report cases of emblem misuse through a dedicated app.

In today’s social media age, the Observatory has seen an increasing number of reports of emblem misuse in social media content. These online misuses – often highly visible – pose a risk of spreading false information about the regulated use of the emblems. This not only undermines trust in Red Cross personnel and volunteers wearing the emblem in humanitarian service, but may weaken the emblem’s protective power in times of armed conflict.

In response, beginning in 2025, the Observatory prepared an ‘advocacy message’ to be sent from the Italian Red Cross social media accounts to users who have posted content involving emblem misuse. The message provides clear, accessible information on the correct use of the emblem, the importance of respecting it, and the humanitarian implications of misuse. The aim is to encourage users to remove the content and to raise

awareness – helping to reduce the spread of potentially harmful or misleading information related to the emblem.

Tommaso Natoli

Head of Humanitarian Diplomacy
and International Law

Italian Red Cross

Giulia Marcucci

Humanitarian Diplomacy Officer

Italian Red Cross

4.1 Narratives of harm

Humanitarian action is fundamentally local. Yet today, the narratives that shape perceptions of humanitarian action are increasingly influenced – and in some cases deliberately manipulated – by sophisticated, well-resourced actors, including private ‘disinformation-for-hire’ firms, state and non-state actors and proxy groups. Their reach and influence extend far beyond the traditional use of propaganda in times of peace or armed conflict. In today’s crises and armed conflicts, shaping the information space has become a strategic element, one that extends well beyond the context or battlefield. The goals to deflect, deceive, denigrate, distract, dismay and divide mark a profound transformation in the information environment of humanitarian action.

No longer confined to isolated local rumours or misperceptions, harmful narratives are now part of a dynamic, global and participatory information ecosystem, where content is produced, shaped and manipulated in real time. As humanitarian scholar Hugo Slim observes, individuals across the world are not merely witnesses to conflict, but are often actively engaging in shaping and/or amplifying narratives – sometimes without fully grasping the consequences. “Millions of people are using their social media accounts to visualize war, record its events, describe their experience and campaign around it. Many observers of war in countries far away have now become real-time spectators who encounter war as remote consumers and partisan camp followers on Facebook, Twitter, Instagram and TikTok... Watching war and entering its information space by forwarding, liking and commenting on war is now routine.”⁵

Efforts to safeguard humanitarian space in the digital era must also consider emerging legal and policy guidance. For example, the Tallinn Manual *on the International Law Applicable to Cyber Operations*⁶ explores how existing international law applies to cyber activities. While the manual is not legally binding, it is widely regarded as an authoritative interpretation of how international legal norms govern state and non-state conduct in cyberspace, including during armed conflict. It affirms that international humanitarian law continues to apply in the digital domain, including protection for humanitarian personnel (Rule 69); that states must not interfere with impartial humanitarian assistance to civilians in an armed conflict, whether through cyber means or other methods (Rule 131), nor misuse protected emblems (Rule 70); and that the principle of distinction between civilians and combatants and between civilian objects and military objectives remains applicable (Rule 32).

In June 2021, a group of international legal experts released the *Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities*.⁷ The statement provides an interpretation of how existing international law applies to the conduct of information operations by state and non-state actors,

including those involving disinformation, misinformation, hate speech and other harmful speech acts. It affirms that states remain bound by existing obligations which continue to govern conduct in the digital information space, such as sovereignty, non-intervention, human rights and international humanitarian law.

4.2 Harms and impacts of harmful information on humanitarian action

Harmful information is not an abstract threat: it has direct, tangible consequences for humanitarian operations, often in ways that undermine trust, disrupt access and endanger both staff and affected communities. From false accusations to targeted campaigns, humanitarian actors across the globe are increasingly confronted with the operational fallout of distorted narratives. These impacts are being felt in armed conflicts, during public health emergencies, in disaster responses and in the everyday functioning of humanitarian organizations, whether local or international. The following cases and examples, drawn from a range of humanitarian organizations, illustrate how harmful information has strained community relationships, exacerbated institutional fragility, hindered access to vulnerable populations and even put responses and lives at risk.

Contributor Insight 4.2



How misinformation and disinformation affect the Rohingya population in Bangladesh, with testimony from recent interviews

Patterns of misinformation and disinformation on social media have significantly affected displaced people from Myanmar hosted in camps in Bangladesh (especially in Cox's Bazar) in multiple ways. The content of this harmful information has ranged from severe online hate speech (that played a central role in the persecution and violence against the Muslim ethnic minority group), to rumours about aid cuts and prejudicial accusations about people taking opportunities away from host community Bangladeshi people living in the Cox's Bazar District.

In early 2025, BBC Media Action interviewed eight young men between the ages of 18 and 34 living in a cross-section of different camps in Ukha, Cox's Bazar. Some of the reported ways that harmful, false or misleading information affects the displaced people are noted below. These refer not only to harmful information on social media, but discriminatory, prejudicial and damaging information in local and national media in Bangladesh.

Examples cited by those interviewed include:

- **Recycled or miscaptioned footage** used to stoke anger or false narratives: videos of campfires in Cox's Bazar being reposted as if showing communal attacks elsewhere (e.g., Tripura, India); and video clips of overcrowded boats framed as "Rohingya boats heading to Indonesia," despite actually being footage of domestic ferries in Bangladesh.
- **Health misinformation:** persistent COVID-19 and vaccine rumours inside the camps, e.g., "vaccines are deadly/make you infertile," "COVID is a hoax," and information about misleading cures, often spread on Facebook, WhatsApp or by word-of-mouth, compounded by the 2019–2020 internet restrictions.
- **Hate speech and scapegoating:** posts dehumanizing or branding them as criminals, "ungrateful" and disease carriers thought to have spread into regional networks, shaping attitudes towards the displaced in Bangladesh.
- **Exaggerated security and militancy claims:** Social media posts that inflate or distort reports of crime or militancy, portraying the camps as controlled by armed groups and fuelling fear and hostility among host communities.
- **Rumours around high-profile events:** Disinformation spikes around incidents – fires, aid cuts, evacuations or repatriation announcements, such as viral claims about "foreign troops" in Cox's Bazar that local fact-checkers later debunked.

Misinformation and disinformation travel in and around the camps through commonly used platforms: Facebook, WhatsApp/IMO app, YouTube/TikTok, as well as offline phone-to-phone sharing (Bluetooth/SHAREit). Despite humanitarian organizations' efforts to set up rumour-tracking initiatives in Cox's Bazar to counter these dynamics, these have been deprioritized in successive funding cuts since 2021.

Melissa Everleigh

Senior Advisor

BBC Media Action

Contributor Insight 4.3



New Zealand Red Cross: Observations from the 2022 North Island severe weather response

In 2022 and 2023, Aotearoa New Zealand was significantly impacted by severe weather events, including Cyclone Gabrielle, which left widespread devastation across many North Island communities. A national state of emergency was declared and New Zealand Red Cross provided support alongside many other organizations. This included mobilizing

specialist teams across the country and launching the New Zealand Disaster Fund, enabling generous Kiwis to support those impacted.

In the initial months of the response, the New Zealand Red Cross became the target of a significant volume of misinformation and disinformation, including false accusations of fraud, fund misuse and failure to deliver critical services. Most of this was spread via social media and in some cases was inflamed by elements of the mainstream media and other not-for-profit entities. Attacks were often directed at both the organization and individual leaders – something we were neither expecting nor fully prepared for.

In response, a communication and engagement plan was formulated, underpinned by calm, nuanced, consistent messaging to share the facts. Multiple trusted communication channels were used over an extended time period. Choosing not to enter into ill-informed 'arguments and reactive debates' was important for the New Zealand Red Cross. It was hard and painful work, requiring time, effort and patience – but it was worth it. Although public sentiment was initially negatively impacted, today positive public sentiment towards the National Society is even stronger than it was prior to 2022. Key lessons are:

- **Education is key:** Don't wait for an emergency, help people understand what you do and why in advance.
- **Form alliances:** Collaborate with others, including media, to amplify accurate and corrective messaging.
- **Engage with communities:** Foster trust through regular engagement before, during and after emergencies. Connected followers are more likely to question or report false information.
- **Correct calmly:** Respond with professionalism and fact-based information, citing verifiable sources. Avoid inflammatory or reactive language.
- **Consider the individual:** Misinformation and disinformation can hurt. Ensure your internal communications are as strong as your external communications, providing support and reassurance to your teams.

In an age where a single viral post can undo years of goodwill, digital integrity must be treated as mission-critical. This is our new normal. By prioritizing transparency, community engagement and strategic communication, an organization can safeguard both its people and reputation – while contributing to a healthier, more informed information environment.

Shane Chisholm

General Manager, Engagement and Enterprise

New Zealand Red Cross

Sarah Stuart-Black

Secretary General

New Zealand Red Cross

Contributor Insight 4.4

Maintaining trust through crises and harmful information

The Canadian Red Cross has built a strong baseline understanding of trust through annual, comprehensive survey-based research. Using this framework, pulse surveys (a short, frequent questionnaire) can be deployed following significant emergency responses or reputational challenges to assess positive or negative impacts on organizational trust. Comparing results to the baseline provides evidence-based insights into what truly impacts organizational trust. Pulse surveys can also include targeted questions to measure the impact of harmful information – assessing both belief in false narratives and their influence on behaviour and outcomes.



Jasper wildfires, 2024

In 2024, wildfires devastated the small mountain town of Jasper, Alberta, forcing a full evacuation of the entire community and destroying approximately 30% of homes. The Canadian Red Cross responded immediately, providing emergency shelter, financial assistance and a range of personal support services to evacuees.

An emergency appeal for donations was launched, with provincial and national governments pledging to match donations. Operational and communications strategies focused on strong community presence, visible engagement and rapid use of public donations for relief operations.

During the response, a coordinated disinformation campaign emerged, falsely claiming that the Canadian Red Cross was absent locally and that funds would not benefit the community in an effort to diminish public support. The Canadian Red Cross challenged and corrected the disinformation in real time and highlighted its successful emergency relief operation and the impact on the community. Following the successful transition from relief to recovery, pulse surveys measured trust levels and the impact of the disinformation campaign.

Results showed increased levels of trust in the Canadian Red Cross throughout the area, which was consistent with previous responses where visibility and effective delivery of assistance drove higher trust.

However, belief in the core claims of the disinformation narrative remained significant, underscoring the need for proactive strategies to address harmful narratives even when overall trust remains strong.



Hurricane Fiona, 2022

Hurricane Fiona made landfall in eastern Canada in 2022 with devastating impact across four provinces, becoming one of the costliest disasters in Canadian history. Each province managed its own emergency response, with the Canadian Red Cross enlisted to support emergency relief operations and administer local government-funded financial aid programmes – each with different eligibility criteria and levels of support.

The lack of coordinated public communications between local governments created significant confusion and frustration for communities seeking aid. This was exacerbated by operational challenges within the Canadian Red Cross in scaling programmes rapidly, with much of the resulting criticism and false perceptions directed at the National Society as the most visible ‘face’ of the relief efforts.

In response, the Canadian Red Cross adopted a proactive public communication approach: making leadership available to media, coordinating messages that took responsibility for improving access and engaging extensively on digital platforms to correct misinformation and guide people toward trusted sources of information and practical ways to access support.

Following the relief operations, pulse surveys assessed both organizational trust and perceived effectiveness, with tailored questions to gauge the impact of the misinformation. Findings showed that many people viewed the overall response of the Canadian Red Cross as slow or mismanaged – perceptions closely aligned with the misinformation narratives directed at the organization during the response. However, the same respondents also believed that assistance ultimately reached people in need. Overall, trust levels in the Canadian Red Cross increased across the region, highlighting the resilience of trust when transparency, accountability and follow-through are maintained, despite harmful information.

Nathan Huculak

Head of Communications

Canadian Red Cross

Contributor Insight 4.5



Kenya, 2024 Gen Z protests

During the 2024 protests against the Finance Bill, a single tweet from a Kenyan influencer falsely alleged that Kenya Red Cross Society vehicles were being used to transport politicians. This misinformation directly endangered aid workers and vehicles, leading to attacks that disrupted life-saving first aid operations.

The Kenya Red Cross Society responded swiftly and strategically, issuing a series of public messages across social media to mitigate the impact of the false claims.

This included:

- firmly refuting the allegations, clarifying that its vehicles were not being used for political purposes
- condemning the attacks on humanitarian personnel and assets, underscoring the need to respect humanitarian space
- reaffirming our commitment to the fundamental principles, particularly neutrality
- launching a public awareness campaign using hashtags such as #NotATarget, reinforcing the message that humanitarian workers must be protected and respected in all circumstances.

The Secretary General of the Kenya Red Cross Society, Dr Ahmed Idris, emphasized: “Humanitarian action depends on trust. When harmful information circulates, it puts lives at risk and undermines the neutrality that allows us to reach people in crisis. Combating misinformation is not optional, it is a priority for protecting communities and ensuring they get the help they deserve.”

Safia Verjee

Executive Director, International Center for Humanitarian Affairs

Kenya Red Cross Society

Contributor Insight 4.6



Negative reactions to a campaign

At the very beginning of the conflict in Ukraine, the Bulgarian Red Cross faced a wave of harmful information circulating on social media. The negative reactions were triggered by the organization’s policy to accept only new items during its public donation campaign for people affected by the conflict arriving in Bulgaria. Criticism focused on why used goods were not accepted, and risked undermining public trust. The Bulgarian Red Cross monitored the situation closely and used every opportunity to explain its approach – emphasizing the urgent needs of the people affected, the importance of preserving their dignity and the operational reasons for accepting only new items or financial donations. While the response helped to manage the situation, the experience revealed crisis communications as an area for further improvement.

This incident occurred in the broader context of a positive perception survey and market study undertaken by the Bulgarian Red Cross to strengthen domestic fundraising and National Society capacity development, with a view to long-term sustainability. The results showed strong public and business support, with the highest scores for “Helps on the most significant topics”, “Makes a visible contribution to its cause” and “Operates transparently, with good accountability.” The Bulgarian Red Cross’ causes and campaigns enjoy high visibility and recognition, along with strong trust among large businesses and it holds

a dominant position over other charitable organizations nationally – an advantage that should be preserved.

Dr Sofia Stoimenova
Secretary in Chief
Bulgarian Red Cross

Contributor Insight 4.7



Turkish Red Crescent's national blood supply system: Ensuring safety, transparency and sustainability

Türkiye's national blood supply system – built entirely on voluntary, non-remunerated and regular blood donations – is managed by the Turkish Red Crescent under the authorization and supervision of the Ministry of Health. Each day, blood donations are collected at approximately 300 locations, supported by 18 regional blood centres, 69 blood donation centres as well as mobile teams.

All donated blood is tested in accordance with the highest national and international standards, screened for infectious diseases, separated into components and stored appropriately. This system is designed not only to meet immediate hospital needs but also to anticipate predictable demand through a carefully planned stock management model. This ensures timely supply to hospitals nationwide, including during emergencies, with delivery based solely on official requests.

Blood donation in Türkiye is strictly voluntary. The costs of processing and preparing blood components are fully covered by the Social Security Institution (*Sosyal Güvenlik Kurumu*). In Türkiye, as in other countries across Europe blood and blood products cannot be traded commercially and requesting payment from patients is illegal. The Turkish Red Crescent has never regarded blood as a commodity, and any violations should be promptly reported to health authorities.

Despite this transparent, ethical and humanitarian-based system, misleading and false claims such as “donated blood is being sold” occasionally circulate online. Such disinformation undermines donors' trust, discourages participation and jeopardizes the sustainability of safe blood supply. These harmful narratives often spread more rapidly and reach broader audiences than official clarifications, particularly when amplified by incomplete information on social media or sensationalist reporting. Addressing misinformation is therefore essential to preserving blood donor confidence and ensuring system sustainability.

To counter this, the Turkish Red Crescent shares timely, verified updates via its [blood services website](#), [Kanver](#) and social media channels. Its communication strategy is firmly grounded in transparency, traceability and accountability, ensuring that the public remains accurately informed and meaningfully engaged.

Safeguarding public trust extends beyond the technical accuracy of information – it is the very foundation of a voluntary blood system and of humanitarian action. By reinforcing this trust, the Turkish Red Crescent guarantees a safe, reliable blood supply, while reinforcing the spirit of social solidarity.

Dr Selami Kılıç

General Director of Blood Services

Turkish Red Crescent

Harmful information has tangible consequences, both online and offline. These impacts can be direct or indirect, depending on how clearly the harm can be linked to a specific incident. The severity of harm is shaped by the scope, scale, duration and magnitude of the incident – to borrow language from cybersecurity – as well as the resilience of the targeted individual or group affected. Often, the harms are multiple, compounding and interconnected. Drawing on the examples and materials gathered for this report, a framing of harms can support more effective response strategies and inform policy-making.

4.3 ***Do no harm in a harmful information age***

The notion of ‘do no harm’⁸ remains a critical ethical lens for humanitarian actors, ensuring that interventions do not inadvertently contribute to suffering or exacerbate existing vulnerabilities. In today’s information-saturated environment, do no harm extends beyond physical interventions or material assistance to include how information is collected, managed, communicated and used. In the digital age, it requires treating information with the same level of care and responsibility as any other aspect of humanitarian programming. It means anticipating risks, minimizing unintended consequences and prioritizing the dignity, safety and agency of affected people in every communication or information-related decision. Building **information resilience** within humanitarian action involves several key commitments:

- Assessing the unintended consequences of how information is produced, shared and communicated.
- Evaluating how messaging and data practices interact with local political, social and cultural dynamics.
- Designing proactive risk mitigation strategies to reduce exposure to harmful information and misuse of data.
- Embedding digital and data responsibility at all stages of humanitarian planning and response.

To operationalize this, humanitarian actors must strengthen their understanding of six interrelated concepts:

- 1 Navigating information responsibly:** Ensuring that organizational communications do not mislead, inflame tensions or stigmatize individuals or communities. Avoiding the amplification of unverified or emotionally charged content that could incite fear, mistrust or violence.
- 2 Protecting digital identities:** Safeguarding personal data, location details and digital traces of vulnerable populations. Recognizing the risks of surveillance, digital profiling and exploitation.
- 3 Countering harmful narratives:** Proactively addressing harmful information that delegitimizes humanitarian action or endangers affected populations and humanitarian actors. Engaging in context-sensitive strategic communication that builds trust without oversimplifying or sensationalizing complex crises.
- 4 Understanding the information ecosystem:** Assessing who creates, controls and consumes information and who is excluded. Recognizing power dynamics within local and global information landscapes and their impact on perceptions and trust.
- 5 Ensuring digital inclusion:** Acknowledging the digital divide, particularly in terms of who is represented, heard and visible online and who is not. Avoiding assumptions based solely on digital visibility while ensuring that the voices of groups that are offline or marginalized are not neglected.
- 6 Strengthening information resilience:** Supporting communities and partners to navigate, verify and produce reliable information. Fostering information literacy not just as a defensive tool, but as a participatory one that empowers individuals to shape their own narrative.

Incorporating these concepts into humanitarian practice is essential to uphold ethical standards and ensure that information is managed in line with the humanitarian imperative: to protect and assist without causing further harm.

Contributor Insight 4.8



Human–human contact in the age of misinformation

In the digital age, the production and spread of misinformation, malinformation and disinformation pose significant threats to the effectiveness of humanitarian response,⁹ particularly in the realm of mental health and psychosocial support (MHPSS). While digital platforms dominate the information landscape due to their reach and immediacy, they often lack the relational depth required to build trust – an essential foundation for effective MHPSS.¹⁰

Human-to-human information sharing, especially through in-person outreach, remains critical in disaster contexts. The Albanian Red Cross response to the 2019 earthquake exemplifies this approach: volunteers rapidly mobilized to go door to door in hard-to-reach areas, providing accurate information and psychosocial support.¹¹ This relational approach not only ensured timely and trusted information delivery but also fostered emotional reassurance and community cohesion¹² – key protective factors for mental health.¹³

Such examples highlight the unique value of the International Red Cross and Red Crescent Movement model of volunteerism, in which local actors – neighbours, family members and friends – serve as embedded, trusted messengers within their communities.¹⁴ In an era where digital misinformation erodes public trust and exacerbates psychological distress, the relational credibility of local volunteers is indispensable.

To safeguard the integrity and accessibility of MHPSS services, humanitarian actors must invest in hybrid communication strategies that combine digital tools with relational, community-based outreach. Research underlines that trust, cultural relevance and interpersonal connection are foundational to effective MHPSS – qualities that cannot be replicated by algorithms alone.¹⁵

Yasin Duman

Research Specialist

Red Cross Red Crescent Movement MHPSS Hub

Joshua Lee

Technical Advisor

Red Cross Red Crescent Movement MHPSS Hub

4.4 Vulnerabilities and amplifiers

“

Why do you think some people refuse to evacuate or secure their homes? People refuse to move because they are afraid their belongings will be stolen, then also because of a lack of trust in the news they hear, fearing it is all lies, and especially because they do not listen to what others say.”

Community member, Madagascar

Across humanitarian contexts, harmful information does not emerge in a vacuum. Practitioners consistently point to a combination of structural vulnerabilities and systemic gaps that amplify the risk and impact of harmful information – both for affected populations and for humanitarian responders themselves.

A commonly cited vulnerability is low digital literacy, particularly among groups that are marginalized. Many people lack the skills or resources to critically assess the trustworthiness of the information they receive, especially on social media and messaging apps, which are often their primary sources of news. This is compounded by weak or absent independent local media, leaving a vacuum easily filled by politically motivated or misleading narratives.

Poorly coordinated information management within and across agencies is also highlighted. In rapidly evolving crises, inconsistent or delayed communication from trusted

actors can create confusion and leave space for rumours to flourish. The political exploitation of humanitarian narratives – where disinformation is intentionally used to discredit aid actors or manipulate perceptions of aid delivery – is a growing threat.

Many sector voices are also candid about what is missing in the humanitarian information ecosystem. There is currently no shared protocol or framework to manage rumours and counter disinformation that spans across organizations. Training on digital threats, media literacy and narrative analysis remains insufficient, particularly for frontline staff and volunteers. Furthermore, engagement with technology platforms, local media outlets and academic institutions is limited and often ad hoc, despite the critical role these actors play in shaping information environments.

4.5 From fragmentation to focus: Avoiding overwhelm and building humanitarian capacity

In the face of an unprecedented flood of information – fragmented, fast-moving and often harmful – humanitarian actors are finding ways to cut through the noise, avoid overwhelm and focus on building the capacity, trust and resilience needed to navigate an increasingly volatile information environment.

Contributor Insight 4.9



Building resilience from within

The Slovak Red Cross views the threat of harmful information as both urgent and deeply relevant, particularly in our region, where recent disinformation campaigns have significantly contributed to societal polarization. In response, we chose to begin by strengthening resilience to harmful information within our own organization – focusing on employees and volunteers who are often on the front lines of this challenge. To support this goal, we developed an educational e-book that combines theoretical insights, clear definitions of key terms and practical guidance for responding to specific situations. It includes advice on when and where to seek help, as well as strategies for managing disinformation-related stress. Written in accessible, plain language, the publication is designed for a wide range of age groups and educational backgrounds and pays particular attention to the regional context.

Prior to this initiative, we conducted an online survey among 450 volunteers across Slovakia, with one quarter responding. The results were striking: 55% reported being frequently or occasionally targeted by strangers with accusations rooted in disinformation. Of these incidents, 40% were directed specifically at the Slovak Red Cross. Additionally,

74% of respondents expressed a strong interest in receiving training on this topic from the Slovak Red Cross headquarters.

Harmful information is an omnipresent and often invisible threat – both online and offline. Its danger lies not only in its capacity to erode public trust in our organization and undermine our humanitarian mission, but also in its personal toll. It can affect mental health, reduce motivation to continue humanitarian work and even increase vulnerability to conspiracy theories.

We collaborated with a Slovak expert in critical thinking to develop appropriate responses to common disinformation-related scenarios encountered by our staff and volunteers in the field. A communications specialist from the Slovak Red Cross ensured that the messaging aligned with our organizational voice and values, and helped define the necessary theoretical scope. The e-book includes links to relevant Slovak-language resources, such as websites focused on navigating harmful information in the regional context. It also provides contacts for mental health support and legal advice, indicating when individuals should seek professional assistance. Thanks to its practical structure and modular content, the resource can be easily adapted to other regional or national contexts.

Zuzana Vongrejová

Communications Specialist

Slovak Red Cross

Contributor Insight 4.10

Crisis communication preparedness at South Sudan Red Cross

Contingency planning

The South Sudan Red Cross has established a Crisis Communication Plan to anticipate and respond to the spread of harmful information during emergencies. This plan sets out clear objectives, key messages, target audiences and communication channels to be used in a crisis. To operationalize the plan, the National Society formed a Crisis Communication Committee comprising the Secretary-General, Communications Manager, Safety and Field Coordinator, Partnerships Coordinator and representatives from the Emergency Operations Centre. The Communications Manager conducts regular environmental scanning to detect any mention of South Sudan Red Cross in public discourse and flags negative narratives for immediate action. The Safety and Field Coordinator monitors the safety and security of staff and volunteers, while the Partnerships Coordinator ensures effective coordination with partners. The Emergency Operations Centre includes departmental managers and three regional coordinators who facilitate two-way communication between headquarters and branches.



Crisis communications in practice

When harmful narratives target the South Sudan Red Cross, the Crisis Communication Committee convenes immediately. Roles are activated and a clear internal communication line is established to ensure coordinated messaging.

The National Society has faced several such challenges. Recently, a commander of a militia group affiliated with the government wore a South Sudan Red Cross jacket bearing the National Society logo and in a recorded video described their operations in the Upper Nile region. The video went viral, generating strong reactions. A few years earlier, the South Sudan Red Cross office in Torit was attacked by armed youth accusing the National Society of employing foreigners; several staff members were beaten and injured. In both cases, it acted swiftly. An emergency meeting of the crisis communication teams was convened, roles were assigned and internal communication lines were established to ensure communication was well-coordinated and channelled to the affected branch. The Communications Manager prepared reactive and defensive media lines, which the Secretary-General used when speaking with the media. Staff and volunteers received a briefing from the Head of Branch.

In the emblem misuse case, the South Sudan Red Cross chose not to engage with the media, as the post had not attracted significant media attention. Instead, internal communication and partner engagement were prioritized to manage reputational risk and maintain trust proactively.

Pascal Ladu

Communications Manager

South Sudan Red Cross

The IFRC's community-based surveillance (CBS) system plays an important role in improving how humanitarian actors listen to and engage with communities. By enabling trained volunteers and local health actors to report unusual health events or concerns directly from within communities, CBS provides real-time insights – often before such signals are visible through formal health systems. Crucially, CBS captures both clinical and non-clinical signals, including community fears, resistance to health interventions and circulating rumours. These early signals help identify emerging misinformation, mistrust or behavioural shifts that could undermine humanitarian efforts. Because CBS is rooted in local engagement and implemented by trusted community members, it does more than enhance surveillance: it also fosters credibility and trust while reinforcing social cohesion. This participatory approach supports two-way communication between communities and humanitarian actors and enables adaptive responses based on community feedback. By feeding directly into risk communication and community engagement strategies, CBS helps close the loop between listening and action. In so doing, it demonstrates that meaningful participation can both strengthen operational effectiveness and build trust in humanitarian interventions.



Bridging the gap: Using technology to help community volunteers counter misinformation, build trust and reduce disease outbreaks

Disease outbreaks often spread fear and misinformation faster than the illness itself. When trust in, or access to, health information is low, people delay seeking help – sometimes with fatal consequences. In Fiji and Comoros, trained Red Cross and Red Crescent community volunteers using the Nyss® platform are changing that.

In Fiji, leptospirosis – a bacterial disease – is endemic. Limited access to health information and widespread misinformation have left communities complacent, with some people dying at home without even seeking treatment, which further fuels fear among communities. The Fiji Red Cross Society trained volunteers in community-based surveillance (CBS), using the Nyss® digital platform (developed by the Norwegian Red Cross and partners) to report suspected leptospirosis cases via SMS. In return, the volunteers receive accurate Ministry of Health-approved health information to share locally with community members, countering rumours, promoting prevention and encouraging timely care seeking. Health workers report increased care seeking at health centres, more recoveries, fewer severe cases and greater community trust.

“The Nyss® reply messages share the correct information to the community: they understand after that.” (Team Leader, Fiji Red Cross Society, Ba District, Fiji.)

In Comoros, cholera reappeared in early 2024 after 17 years. With limited local experience in managing cholera, rumours spread quickly. The IFRC supported the Comoros Red Crescent by deploying a Public Health Emergency Response Unit, training 200 volunteers and 10 supervisors in CBS across 79 villages in Anjouan, the outbreak’s epicentre. Comoros Red Crescent volunteers sent nearly 2,900 alerts via Nyss®, receiving tailored Ministry of Health-approved messages in French, helping communities act quickly and get patients to treatment centres earlier.

In both countries, CBS has been crucial – empowering local volunteers to connect communities to health systems, counter misinformation, stop diseases spreading, build trust and save lives.

Anine Kongelf

Senior Analyst

Australian Red Cross

Abbey Byrne

Senior Specialist

Australian Red Cross

Tonje Tingberg

Senior Adviser

Australian Red Cross

Lea Franconeri

Senior Adviser

Norwegian Red Cross

4.6 When viruses go viral: Building resilience against harmful information in a pandemic

The COVID-19 pandemic created a perfect storm for harmful information to thrive. Two key factors contributed to this: widespread fear, uncertainty and doubt; and the unprecedented ability for information – alongside misinformation and disinformation – to be rapidly created and spread via the internet, mobile networks and social media platforms. The World Health Organization (WHO) observed that, alongside the spread of the COVID-19 virus itself, the world was experiencing an ‘infodemic’ – an overabundance of information, both accurate and inaccurate, circulating in digital and physical spaces during an epidemic or pandemic.¹⁶ This flood of information confused audiences, undermined trust in credible sources and made it more difficult for people to adopt protective behaviours.

To respond, WHO launched several initiatives. The Early AI-Supported Response with Social Listening (EARS) platform provided real-time insights into online discussions about COVID-19, enabling health authorities to better understand public concerns and deliver accurate, timely information. The WHO Information Network for Epidemics (EPI-WIN) helped make scientific information more accessible and offered online training to support health workers and communities.¹⁷ WHO also worked to amplify evidence-based public health messaging and guide users toward trustworthy sources. It encouraged the public to report misinformation and collaborated with major social media platforms to prioritize science-based COVID-19 content in search results.

The UN’s Pause campaign (‘PledgetoPause’)¹⁸ encouraged people to take a moment to verify the accuracy of any information before sharing it. A study¹⁹ conducted by the Massachusetts Institute of Technology in the UK and the US found that this simple act of pausing to reflect on the origin, credibility, relevance and accuracy of content before sharing it on phones, computers or social media significantly reduced the likelihood of sharing misinformation. Participants exposed to the campaign were notably less likely to share false headlines. The campaign was based on a powerful premise: that interrupting, even for a few seconds, the impulse to share something – an urge often driven by emotions such as excitement, anger, sadness, elation or even altruism – can create space for more critical thinking, better judgement and less amplification of false or misleading claims.

The 2021 UN Open-ended Working Group on Information and Communications Technologies (ICTs) noted that “the COVID-19 pandemic demonstrated the risks and consequences of malicious activities that seek to exploit vulnerabilities in times when societies are under enormous strain”.²⁰ In doing so, the group drew a clear connection between the COVID-19 infodemic and the malicious use of ICTs. This underscores how pandemics can create fertile ground for both cyber threats and harmful information. This same report also highlighted that: “While agreeing on the need to safeguard all critical infrastructure and critical information infrastructure supporting essential services to the public, along with endeavouring to ensure the general availability and integrity of the Internet, States further concluded that the COVID-19 pandemic had accentuated the importance of protecting healthcare infrastructure including medical services and

facilities through the norms addressing critical infrastructure, such as those affirmed by consensus through UN General Assembly resolution 70/237.”²¹

The global response to the COVID-19 infodemic set an important precedent. It demonstrated what is possible when there is a concerted, coordinated high-level effort to counter harmful information and underscored the critical importance of international solidarity in doing so. The May 2025 WHO Pandemic Agreement acknowledges “the importance of building trust and ensuring the timely sharing of information to prevent misinformation, disinformation and stigmatization”, and explicitly affirms the importance of effective communication and public awareness efforts. It emphasizes the importance of strengthening pandemic literacy and science- and evidence-based information, risk communication and community-level engagement.²² Notably, the Pandemic Agreement itself became the target of systematic disinformation. Its negotiation process was complicated by false claims related to national sovereignty, pandemic countermeasures, intellectual property rights and freedom of speech²³ – demonstrating how international frameworks can become entangled in harmful narratives.

Concluding remarks: Navigating a hostile information landscape

Virtual spaces are not neutral arenas for dialogue. Digital platforms have become battlegrounds where trust and truth are deliberately undermined and where individuals, communities and even nations are subjected to harmful information attacks. In this dense, volatile space, the boundaries between truth and falsehood blur. Humanitarian narratives must contend with an overwhelming flood of emotional, politicized and sometimes instrumentalized content.

A decade ago, there was a sense that the “information flows cannot be controlled nor do they need to be”.²⁴ This perspective reflected an earlier digital era, when social media platforms were emerging and user-generated content largely reflected the goodwill of individuals, rather than the calculated harmful intent of malicious actors. At the time, it was a progressive call for greater participation, transparency and accountability. But today’s landscape is profoundly different. The question is no longer whether information should be regulated, but what should be addressed, how, by whom and with what safeguards – particularly in terms of protecting fundamental rights and freedoms. Striking this balance has become one of the central challenges of the current era.

At its core, humanitarian action is about meeting the needs of people in crises, not securing visibility or reputation of the organizations delivering it. Yet, when the spotlight shifts from communities and onto organizations – through unmerited criticism or targeted harmful information – their safety and ability to operate effectively is under threat. Increasingly, harmful narratives shape public and societal perceptions of humanitarian action, constrain access, erode trust and shrink the operational space for principled humanitarian action. While legitimate scrutiny is vital for accountability and improvement, it must be distinguished from deliberate distortion or harmful information designed to delegitimize and obstruct humanitarian efforts. The consequences, as highlighted throughout this report, are stark: diminished access, reputational damage,

emotional strain and heightened security threats to staff and volunteers, operational delays and loss of public trust.

These dynamics underscore the urgent need for more coordinated and systemic responses. Addressing harmful information requires sector-wide collaboration and investment, and proactive engagement with those who shape and influence the information environment – not least governments and technology companies (see [Chapter 5, on page 167](#) for a detailed examination of their role).

Above all, this landscape raises urgent questions that no single organization can resolve alone; the challenges transcend the mandate or influence of any individual organization. The dilemmas demand collective reflection and shared approaches:

- 1 **Neutrality versus harmful narratives:** When dehumanizing or discriminatory narratives spread – sometimes by state or political actors – should humanitarians remain silent to preserve neutrality and access, or speak out to prevent harm?
- 2 **Debunking versus risking access:** Should falsehoods be publicly challenged, even if doing so risks undermining neutrality or operational access? How can humanitarian actors respond without being drawn into politicized or polarized debates?
- 3 **Integrity versus polarization:** In fragmented and hostile information environments, how can humanitarian actors protect the integrity of their messaging and maintain trust without being co-opted, misrepresented or silenced?
- 4 **Collective action versus fragmented responses:** What partnerships, mechanisms and international norms are needed to protect humanitarian organizations from information attacks? Should the sector advocate collectively for stronger platform policies and governance frameworks to safeguard humanitarian space?
- 5 **Speak out versus silence:** When harmful narratives systematically target groups in vulnerable situations – such as migrants – and are driven by state or political interests, what space remains for humanitarians to respond? How can they challenge such narratives without compromising neutrality? At what point is silence perceived as tacit approval of discriminatory or dehumanizing narratives, and when does silence risk becoming complicity?

Together, these dilemmas underscore the urgent need for the humanitarian sector to find collective strategies that preserve principled humanitarian action while protecting at-risk groups and safeguarding trust in an increasingly hostile information environment.

Asks, aims and recommendations

Asks

Treat harmful information as an operational risk by embedding preparedness, accountability and transparency into all phases of humanitarian action, from early warning to recovery – and measure its human, social and operational harms as rigorously as physical impacts.

Aims

Safeguard humanitarian access and action by recognizing harmful information as a systemic risk.

Protect operations, staff and community trust: Integrate harmful information analysis and response into all phases of humanitarian action.

Build resilience by resourcing responses and systematically measuring harms to inform evidence-based action.

Recommendations

States and policy-makers

- Build societal resilience through public information literacy and awareness initiatives.
- Integrate harmful information risks into disaster management and crisis response frameworks.
- Support mechanisms to identify and address harmful information incidents that endanger civilians or disrupt humanitarian operations, ensuring the mandates of principled humanitarian actors are respected.
- Support international cooperation and uphold cyber norms to protect humanitarian action and critical services.
- Establish flexible funding lines for harmful information response, with priority for building local resilience and community-led rumour tracking.

Technology platforms

- Adapt or “cool” algorithms during humanitarian crises to reduce amplification of harmful narratives.
- Implement crisis protocols that temporarily re-weight algorithms in favour of safety and reliability over optimizing engagement.
- Prioritize authoritative information from trusted humanitarian, health and local actors for the delivery of timely and life-saving guidance.
- Collaborate with principled humanitarian actors to design context-specific, transparent interventions on platforms

that respect users' human rights and protect the safety of crisis-affected populations.

- Ensure independent oversight to balance freedom of expression with the need to prevent harm to civilians, humanitarian actors and operations.

Humanitarian actors

- Integrate harmful information monitoring into programmes, assessments and early warning systems.
- Document, analyse and share actionable insights and lessons learned from harmful information incidents and their operational impacts to inform real-time adaptation and long-term resilience, without impeding principled humanitarian action.
- Strengthen staff and volunteer capacity to identify, document and respond safely to harmful narratives. Strengthen internal communication systems to ensure staff and volunteers are

informed, coordinated and able to respond safely and consistently to harmful narratives while upholding humanitarian principles.

- Invest in multilingual monitoring, two-way communication, rapid verification systems and partnerships with trusted local moderators.
- Facilitate regional research hubs to strengthen monitoring, measurement and evidence-based assessment of harms.
- Strengthen collaboration and training for humanitarian staff and volunteers on digital threats, information and media literacy, and narrative analysis.

Community and local leaders

- Co-create rumour-tracking and feedback systems to detect harmful information early.
- Verify and amplify accurate information through trusted local channels.

- Document harmful information incidents that affect safety, services or access.
- Foster dialogue and peer-to-peer engagement that reduces stigma, fear and division.

Endnotes

- 1 Netcraft. Web Server Survey (2025) www.netcraft.com/blog/april-2025-web-server-survey
- 2 Backlinko. Social Network Usage & Growth Statistics (2025) <https://backlinko.com/social-media-users>
- 3 Iqbal, M. TikTok Revenue and Usage Statistics (2025) www.businessofapps.com/data/tik-tok-statistic
- 4 ICRC. 'False allegations: ICRC condemns video showing cash in trunks.' Article. 18 January 2018. www.icrc.org/en/document/false-allegations-icrc-condemns-video-cash-trunks
- 5 Slim, H. *Solferino 21: Warfare, Civilians and Humanitarian Work in the 21st Century*. (2022), p.55.
- 6 Schmitt, MN. (ed.) *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. (Cambridge University Press, 2017) <https://doi.org/10.1017/9781316822524>
- 7 Oxford Institute for Ethics, Law and Armed Conflict. *The Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities*. 2 June 2021. www.elac.ox.ac.uk/the-oxford-process/the-statements-overview/the-oxford-statement-on-the-regulation-of-information-operations-and-activities
- 8 Anderson, M.B. *Do No Harm: How Aid Can Support Peace or War*. (1999)
- 9 Borges do Nascimento, I.J., Pizarro, A.B., Almeida, J.M. et al. Infodemics and health misinformation: a systematic review of reviews. *Bulletin of the WHO* 2022;100(9), 544–561. <https://doi.org/10.2471/BLT.21.287654>; Hilberts, S., Evers, S., Govers, M. et al. The impact of misinformation on social media in the context of natural disasters. *European Journal of Public Health* 2024;34. <https://doi.org/10.1093/eurpub/ckae144.245>
- 10 Brun, C., and Horst, C. Towards a Conceptualisation of Relational Humanitarianism. *Journal of Humanitarian Affairs*, 2023;5(1), 62–72. <https://doi.org/10.7227/JHA.103>; Yu, J., Bekerian, D.A., and Osback, C. Navigating the Digital Landscape: Challenges and Barriers to Effective Information Use on the Internet. *Encyclopedia* 2024;4(4), 1665–1680. <https://doi.org/10.3390/encyclopedia4040109>; IASC. IASC Reference Group on Mental Health and Psychosocial Support in Emergency Settings. (2007)
- 11 Venghaus, G. Albania: Final Evaluation of the 2019 Albania Earthquake Emergency Appeal. (2021) <https://reliefweb.int/report/albania/albania-final-evaluation-2019-albania-earthquake-emergency-appeal-january-february>; Red Cross EU Office. News. Enhanced post-earthquake support to Albania. 21 February 2020. <https://redcross.eu/latest-news/enhanced-post-earthquake-support-to-albania>
- 12 Red Cross EU Office. News. Enhanced post-earthquake support to Albania. 21 February 2020. <https://redcross.eu/latest-news/enhanced-post-earthquake-support-to-albania>
- 13 IASC. IASC Reference Group on Mental Health and Psychosocial Support in Emergency Settings. (2007); Hetherington, E., McDonald, S., Wu, M. et al. Risk and Protective Factors for Mental Health and Community Cohesion After the 2013 Calgary Flood. *Disaster Medicine and Public Health Preparedness* 2018;12(4), 470–477. doi:10.1017/dmp.2017.91
- 14 Lindal, L. *Community-based Psychosocial Support. Volunteer Manual*. IFRC Reference Centre for Psychosocial Support. (2023) <https://pscentre.org/wp-content/uploads/2023/03/Community-based-Mental-Health-and-Psychosocial-Support-Volunteer-Manual.pdf>
- 15 Buitrago, DCC., Rattner, M., James, LE. et al. Barriers and Facilitators to Implementing a Community-Based Psychosocial Support Intervention Conducted In-Person and Remotely: A Qualitative Study in Quibdó, Colombia. *Global Health, Science and Practice* 2024;12(1) <https://doi.org/10.9745/GHSP-D-23-00032>
- 16 WHO. Health security in a digitalized world during public health crises, *Weekly Epidemiological Record* 2024;99(04), 23–25. <https://iris.who.int/handle/10665/375830>
- 17 Verified is a UN initiative, in collaboration with Purpose, that aims to empower people around the world with science-based information during the COVID-19 response. Working with UN agencies, influencers, civil society, businesses and social media platforms, Verified creates and distributes trusted, accurate information and encourages consumers to change their media consumption practices to reduce and stop the spread of misinformation online. <https://shareverified.com/en>
- 18 UN Regional Information Centre for Western Europe. 'UN Secretary-General launches #PledgeToPause to fight misinformation.' News 21 October 2020, <https://unric.org/en/un-secretary-general-launches-pledgeto-pause-to-fight-misinformation>
- 19 UN. 'New MIT study says United Nations Pause Campaign slows spread of life-threatening misinformation.' Press release. 1 July 2021. www.un.org/sites/un2.un.org/files/press_release_pause_campaign_final_1_july.pdf
- 20 UN General Assembly. Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security Final Substantive Report (A/AC.290/2021/CRP.2) (10 March 2021), para. 4. The UN Open-Ended Working Group on Security of and in the Use of Information and Communications Technologies, established by the UN General Assembly (A/RES/75/240), is mandated to advance responsible state behaviour in cyberspace and strengthen the security and stability of the ICT environment.
- 21 Ibid, para. 26
- 22 WHO. WHO Pandemic Agreement. (2025) Article 16. https://apps.who.int/gb/ebwha/pdf_files/WHA78/A78_R1-en.pdf
- 23 Finch, A., Klock, KA, Gostin, LO, et al. *Safeguarding the Pandemic Agreement from Disinformation*. Think Global Health. (2024) www.thinkglobalhealth.org/article/safeguarding-pandemic-agreement-disinformation
- 24 IFRC. *World Disasters Report 2013: Focus on Technology and the Future of Humanitarian Action* (2013), p.172 www.ifrc.org/sites/default/files/WDR-2013.pdf





Chapter 5

Navigating regulation, rights and societal resilience





CROCE ROSSA ITALIANA



OCEAN WIND

Chapter 5



Navigating regulation, rights and societal resilience

Contents

	Introduction: Information landscape and humanitarian contexts	169
5.1	Defining harmful information: A strategic and contextual challenge	170
5.2	The risks of information control in emergencies	171
5.3	Digital ceasefires and harmful information	173
5.4	Sovereignty in cyberspace	175
5.5	Media as a pillar of societal resilience	175
5.6	A threat to humanitarian action and to humanity itself	186
5.7	Red Cross and Red Crescent Appeal to States	189
5.8	UN action on AI and information integrity	194
5.9	Content moderation and the power of platforms	194
5.10	From self-regulation to state oversight: The evolving governance of online content	196
5.11	Framing a response: Supply and demand solutions to disinformation	197
5.12	Civic trust and societal resilience	200
	Concluding remarks: A collective responsibility for preserving principled humanitarian action	202
	Endnotes	206

Introduction: Information landscape and humanitarian contexts

The spread of false information – whether overt or covert – poses a serious threat to the humanitarian sector, undermining its work and endangering the populations it aims to serve. While not new, digitalization has transformed the scale, speed and complexity of harmful information. This shift is unfolding amid growing distrust in institutions, making its effects even more corrosive. Today’s information environment is crowded and layered, with multiple forms of harmful content coexisting and reinforcing one another. Harmful information rarely exists in isolation; it amplifies other risks such as geopolitical tensions and environmental crises. The World Economic Forum’s *Global Risks Perception Survey 2024–2025*¹ warns that vulnerabilities linked to online activity are deepening alongside widening societal and political divisions, eroding public trust in information and institutions.

Humanitarian contexts are inherently complex, particularly during disasters, armed conflict and other emergencies. In such situations, the state plays a central role in shaping the information environment – serving as the primary source of official communication, issuing alerts, coordinating messaging and engaging with both domestic and international responders. Timely, accurate and trusted information is essential: it saves lives, fosters trust and supports public order. However, malicious actors often seek to control or distort these narratives. Countering such efforts requires not only credible channels of communication but also organizational and societal resilience.

An important actor navigating this complexity at the national level is the National Red Cross and Red Crescent Society. National Societies hold a unique position as **auxiliaries to public authorities in the humanitarian field** – a status that allows them to support and engage with the state while maintaining independence and adherence to humanitarian principles.

The status of National Red Cross and Red Crescent Societies is not arbitrary; it is formally recognized by states that are signatories to the Geneva Conventions. These states have committed – through both the UN and the International Conferences of the Red Cross and Red Crescent – to uphold and respect the Movement’s fundamental principles. These principles – humanity, impartiality, neutrality, independence, voluntary service, unity and universality – are essential to ensuring that National Societies can operate effectively and free from political interference. In practice, this means that National Societies must be able to deliver humanitarian assistance independently, impartially and neutrally, even in complex or politically sensitive environments. Respect for these principles is not only a matter of good practice; it is a legal and moral obligation for states that have endorsed the Geneva Conventions and the commitments made in international humanitarian forums.

Traditionally, this auxiliary role has encompassed disaster preparedness and response, health services, support to vulnerable populations and the promotion of humanitarian principles and international humanitarian law. In practice, it also demands trusted public awareness, communication of information and meaningful community engagement – essential to effective humanitarian action. National Societies contribute in multiple ways: communicating early warnings and public health messages; coordinating with authorities to ensure consistent, timely and localized information; supporting risk

communication and community feedback mechanisms; and increasingly working to counter harmful information and rumours that may undermine humanitarian responses.

Given the complexity of today's information environment, there is a need for greater recognition of – and support for – the independence of National Societies in fulfilling these roles. The 2024 Council of Delegates called on each National Society to strengthen dialogue with public authorities to reinforce their independent action and decision-making, including by anchoring their auxiliary role in domestic law in line with Movement standards and past International Conference resolutions.²

5.1 Defining harmful information: A strategic and contextual challenge

Determining what constitutes harmful information is essential for any entity seeking to develop a strategic response aligned with risk management. The impact of harmful information depends on several variables including its scope (scale and severity), duration, magnitude of the incident and the resilience of the affected individual(s), organization(s) and broader context. Harm may result directly from the incident or indirectly, with distinctions often based on the degree of certainty that the information caused the outcome. A single incident can generate multiple forms of harm, classified as direct (primary) or indirect (secondary or tertiary). These may include physical, psychological, societal and deprivational harms. The effects of harmful information are often disruptive and multidimensional, manifesting across several categories simultaneously.

Governments interpret harmful information in diverse ways, shaped by their legal frameworks, political priorities and societal values. Some describe it as deliberately false or misleading content intended to deceive, manipulate or cause harm – whether to individuals, public institutions or national security. Others place greater emphasis on intent, such as its use to influence elections, incite violence or undermine public health measures.

Common definitions distinguish between:

- disinformation: deliberate falsehoods intended to cause harm
- misinformation: false or inaccurate information shared without harmful intent
- malinformation: genuine information used with the intent to cause harm.

In these framings, intent is the key differentiator.

One of the most contested areas in defining harmful information is the boundary between legitimate political dissent and incitement. In some contexts, this line is blurred, raising serious concerns about freedom of expression. Content may be labelled as dissenting, destabilizing or inciting violence, particularly during times of crisis or internal tension. The challenge is ensuring that measures designed to counter harmful information are not used as a pretext for silencing dissent. Clear safeguards are needed

to distinguish between speech that challenges authority and speech that genuinely threatens public order, safety and human dignity.



You know, ... I don't think the law has caught up with social media and being able to kind of hold people to account for the information that they've been putting up there. And I think, ... you know, we've got a right to kind of free speech. So ... there's a conflict there, isn't there? To a certain extent, you know, he can say anything he likes, but it could be false... And then there's the information that people have a right to know, what the truth is. So I think there's a difficult balance to get between the laws and free speech."

Community member, UK

5.2 The risks of information control in emergencies

While states have a legitimate responsibility to protect territorial integrity and public order, overly broad or opaque security measures can come at the expense of individual rights and community safety. In some contexts, authorities have withheld or delayed the release of critical information, shared incomplete or misleading narratives or used public messaging to shape perceptions during emergencies. In situations of armed conflict, information often becomes a strategic instrument – used to influence, mobilize or obscure.

Measures such as restricting, blocking, filtering, censoring or regulating content are sometimes introduced under the pretext of national security, public order or preservation of culture. Yet, when vaguely defined or applied without transparency, such measures can undermine access to essential information, restrict freedom of expression and compromise the integrity of the information space. Those with disproportionate control over media and platforms can use harmful information to discredit civil society and humanitarian organizations, associating them with malign or unlawful actors or blaming them for crises to justify repressive policies. This manipulation fuels discrimination, human rights abuses and social tensions.

At the core lies a critical tension: addressing harmful information effectively while safeguarding fundamental rights. Vague or overly broad definitions of harmful information risk misuse – suppressing journalism, silencing dissent or curtailing civil society. In some contexts, legislation intended to combat disinformation has been used to detain journalists, restrict reporting or target individuals for expressing political opinions. Such measures can erode public trust and deepen fear and polarization. Effective responses to harmful information must therefore be grounded in legality, necessity and proportionality.

A body of UN human rights instruments, resolutions and guidance underscores the importance of protecting human rights online, including freedom of expression and

access to information. These instruments provide essential guidance in navigating the tensions between addressing harmful information and safeguarding fundamental rights. For example, Article 19 of the Universal Declaration of Human Rights (1948) states that “Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.”³

One notable multilateral initiative is the Freedom Online Coalition, a group of 42 states committed to advancing internet freedom. The coalition has repeatedly urged governments to refrain from sponsoring disinformation, including campaigns that undermine humanitarian principles or incite violence. It also emphasizes the importance of maintaining a global, free, open, secure and interoperable internet and other digital communications services during times of armed conflict. The coalition has warned that internet shutdowns hinder access to life-saving information for crisis-affected populations, disrupt protection mechanisms and vital services and obstruct the delivery of humanitarian assistance.⁴ Importantly, this underscores not only the dangers of disinformation but also the risks posed by the absence of information – a critical concern in emergencies where timely communication can mean the difference between life and death.

Contributor Insight 5.1

Estonia: Upholding freedom while countering harmful information

Estonia ranks among the freest and most digitally advanced societies, with a culture that values openness and democratic resilience. According to Freedom House’s *Freedom on the Net* report, Estonia has one of the world’s most open online environments, with no state-imposed restrictions on expression. The 2025 *World Press Freedom Index* ranks Estonia second globally, emphasizing a robust legal and political framework that enables journalists to operate safely and independently.

Estonia distinguishes clearly between free media and hostile propaganda. We do not engage in counter-propaganda but instead enforce EU sanctions to limit harmful disinformation. This regulatory, rather than ideological, approach safeguards facts and democratic integrity. Estonia’s response is multifaceted, combining regulation, education and public engagement.

Domestically, harmful disinformation is treated as a national security and societal challenge requiring a whole-of-society response. Estonia promotes media education across all school levels, supports independent journalism, monitors hostile narratives in collaboration with the public sector, NGOs and volunteers, and maintains open communication with both media and citizens. The Government Office coordinates strategic communication across ministries, including psychological defence and crisis messaging. Roles are clearly defined: the Consumer Protection and Technical Regulatory Authority enforces the Digital Services Act, the Information System Authority and CERT-EE manage cyber incidents and the Internal Security Service addresses foreign influence.

Internationally, Estonia promotes both digital rights and media freedom. As the 2025 Chair of the Freedom Online Coalition, we focus on ensuring that emerging technologies, including AI, are governed in a rights-respecting way, while advancing digital inclusion and cross-regional dialogue. As Co-Chair of the Media Freedom Coalition from mid-2023 until mid-2025, Estonia prioritized journalist protection, supporting independent media and raising awareness of disinformation risks. Strong international cooperation underpins this approach, reflecting Estonia's commitment to human rights and resilient, open and secure information ecosystems.

Our lesson is clear: freedom must be protected as a core value, but it must also be paired with proportionate regulation, sanctions enforcement, strong cyber defences, education and societal engagement. Only this balance ensures the protection of freedom of expression in a democratic society, while remaining resilient against information manipulation.

Maarja Kask

Third Secretary, Department for International Organisations and Human Rights

Ministry of Foreign Affairs of Estonia

Silver Küngas

First Secretary, Department for International Organisations and Human Rights

Ministry of Foreign Affairs of Estonia

5.3 Digital ceasefires and harmful information

While traditional ceasefire agreements focus on halting the use of force and physical violence, there is growing recognition that the information space – including the spread of harmful information – can also threaten the sustainability of peace. Harmful information, such as hate speech, incitement to violence and disinformation, have been used to undermine ceasefires, discredit negotiation parties or provoke renewed hostilities. In some recent contexts, mediators and peacebuilders have begun to acknowledge the strategic role of information and have advocated for explicit clauses in ceasefire and peace agreements that explicitly reference the prohibition of hate speech, incitement and disinformation, as well as the importance of access to the internet and accurate information. However, such provisions remain rare, vague or insufficiently monitored.

Incorporating information-related commitments in ceasefire agreements, such as respecting media freedom, refraining from online incitement and establishing joint communication mechanisms, could strengthen trust between parties and help prevent a relapse into violence. This is particularly critical in contexts where digital platforms are actively used to mobilize support, reinforce polarization or spread disinformation.



Humanitarian impacts of digital pollution: The case of Libya

Access to information and the ability to communicate are central to the enjoyment of several human rights, especially in situations of crisis. Just like access to food, access to information is multidimensional – it cannot be satisfied simply by the existence of an information environment.

Internet shutdowns force people into fragmented or unreliable information bubbles, removing them from the ability to verify or even access potentially life-saving information. At the same time, the degradation of content in the information ecosystem can render access to information ineffective or even dangerous.

Unsurprisingly, these two phenomena are increasingly documented in combination, especially in the aftermath of disasters. Local activists have highlighted this interplay during the 2023 floods that ravaged north-eastern Libya, where these risks converged with deadly consequences.

As reported by Libyan journalists and human rights defenders, the mix of internet shutdowns, breaks in connectivity and an unchecked spread of misinformation and disinformation significantly increased the disaster's impact. It increased confusion, eroded credibility and trust in disaster response and delayed life-saving assistance. The result was a slower, less effective humanitarian response but, most tragically, a much higher number of casualties and greater damage to affected communities.⁵

Access Now and other digital rights organizations work to ensure that affected communities have unrestricted access to secure, reliable information, a longstanding challenge in the humanitarian sector where significant progress has been made over time.

However, these gains are now under threat. The normalization of internet shutdowns, combined with the shrinking of funding for local digital rights initiatives and journalism is threatening these gains. Meanwhile, the spread of so-called 'AI-slop' – low quality, AI-generated content – is further polluting the information ecosystem, pushing it beyond a critical threshold. As local initiatives struggle to stay afloat, hostile actors are stepping into the void driven by political or financial motives. In some cases, these actors offer funding under conditions incompatible with humanitarian principles.

As seen in Libya, local journalists, relief organizations and digital rights defenders continue to lead the response, stepping up when their communities need them most. But without sustained donor and tech sector support and stronger protection from the international community, the information environment in crisis settings will continue to worsen and in turn worsen the harmful impacts of disasters.

Giulio Coppi

Senior Humanitarian Officer

Access Now

Marwa Fatafta

Middle East and North Africa Policy
and Advocacy Director

Access Now

5.4 Sovereignty in cyberspace

Determining whether an influence operation in cyberspace is a breach of sovereignty is a complex issue under international law. It is closely tied to whether the act infringes on a state's territorial integrity or interferes with inherently governmental functions, such as delivering public services, conducting elections or collecting taxes.

Discussions at the UN Open-ended working group on responsible use of information and communications technologies (ICTs) have examined these issues in the context of cyber incidents.⁶ Key considerations include whether an incident damages or limits the functionality of infrastructure or related equipment; alters or deletes data; interferes with inherently governmental functions; and whether a state has sought to influence, disrupt or delay democratic processes in another state which could include through propaganda, disinformation or covert actions.⁷ Thus, whether such acts amount to a violation of sovereignty depends on their nature and repercussions: Does the incident violate territorial integrity? Does it interfere with or usurp an inherently governmental function?

Researchers such as Pamment emphasize the importance of assessing *foreignness* in influence operations, specifically, whether they have connections to foreign states, citizens or interests aiming to influence public opinion in another state.⁸

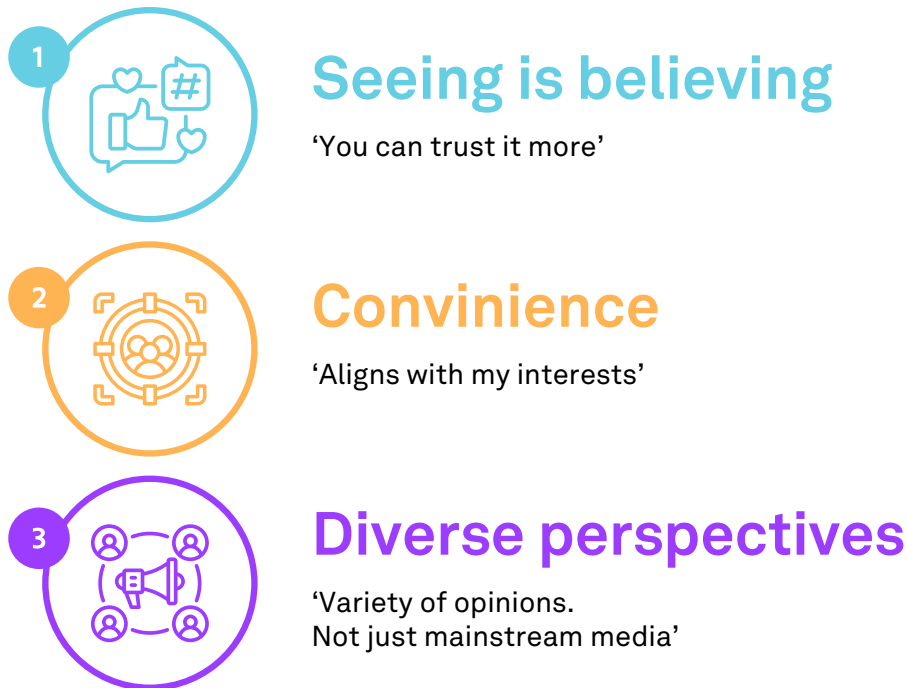
5.5 Media as a pillar of societal resilience

The level of support – or lack of – that states and other stakeholders show for the presence of humanitarian organizations on their territory can significantly influence how they are perceived. When an organization is portrayed as biased, politicized or pursuing its own agenda, its legitimacy is undermined, particularly in contexts where there is limited public knowledge or familiarity with its mandate or work. In such environments, perceptions matter deeply. Humanitarian organizations must be seen as principled, competent and effective. Where public awareness is low and/or where polarization is high, people are more susceptible to harmful information, making them more vulnerable to misconceptions and distrust when forming their opinions.

Researchers Humprecht, Esser and Van Aelst identify five key risk factors that influence a society's vulnerability or resilience to disinformation.⁹ These are: (1) high levels of societal polarization; (2) low trust in news media; (3) highly distributed media landscapes (offering more entry points for disinformation); (4) large media markets, where attention-based revenue models incentivize sensational or false content – what the Global Disinformation Index describes as “the loudest voices getting the most attention”; and (5) high levels of social media use, which are consistently correlated with greater susceptibility to disinformation.

According to the Reuters Institute *Digital News Report 2024*, audiences are drawn to video and other content on social and video platforms for three main reasons: the perceived authenticity of unfiltered, user-generated content, which appears more *trustworthy*; a preference for making up one's own mind without editorial framing; and growing mistrust in traditional media.

Fig 5.1 Motivations for using social video



Source: Reuters Institute for the Study of Journalism¹⁰

Many users, especially younger audiences, tend to trust bystander or first-person footage more than traditional news sources, perceiving it as less filtered, less biased and/or less politically manipulated. Concerns about misinformation often centre less on entirely false content and more on seeing opinions and agendas that they may disagree with, perceived bias or superficial, unsubstantiated journalism.¹¹

These dynamics underscore the importance of visibility, transparency and trust-building in humanitarian communication – especially in environments where disinformation can fill gaps left by limited public understanding.

Contributor Insight 5.3

Audiences increasingly rely on social media for news despite concerns about information quality

Social media and video networks have become increasingly central to how people access news in recent years, while other forms of news consumption have dwindled. Data from

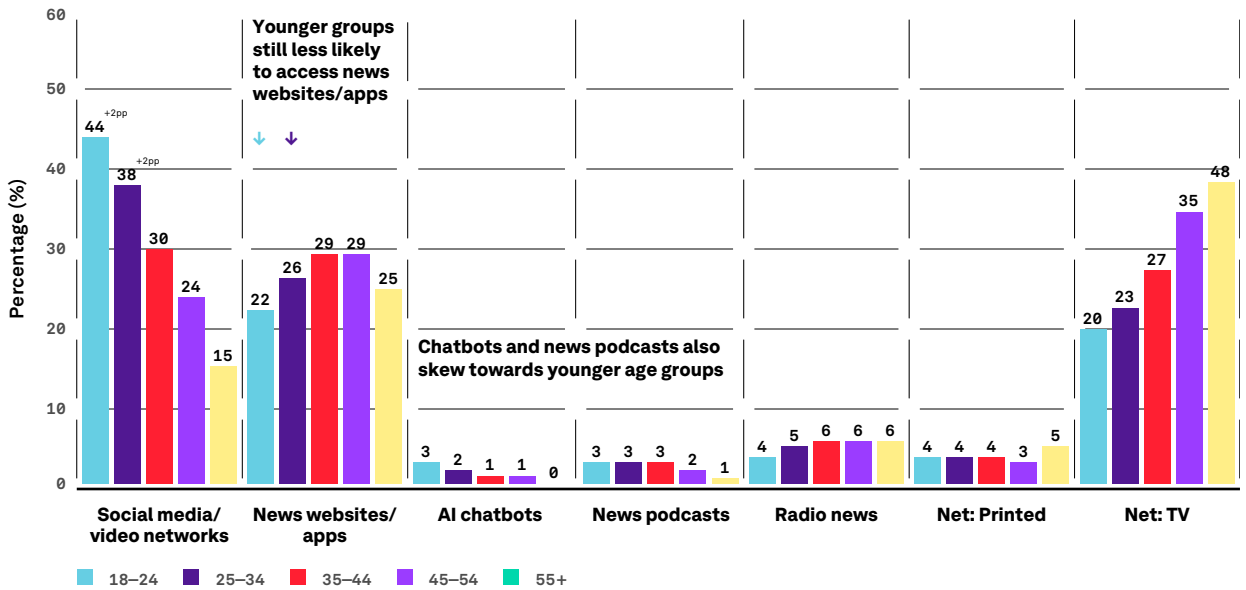
the Reuters Institute *Digital News Report 2025* show that social media overtook television as a source of news in the US for the first time, and in many Latin American, Asian and African countries, social media has been the main source of news for some time.

Generational shifts are driving much of this change. While older adults (55+) continue to rely on traditional media for news, everyone else is now effectively digital-first. Within that online space, younger groups have become less likely to go directly to a news website or app – and more likely to consume news via social media or video networks. Across the 48 countries surveyed, 44% of under-25s and 38% of those aged 25–34 prefer accessing news through platforms, while just a quarter of each group prefer going directly to news websites. Respondents cite convenience and relevance of the news they see through platforms as key reasons – it is often encountered while doing other online activities.

Fig 5.2

Proportion of each age group that say each are their ‘main source’ of news, 2025

Aggregate data from 48 countries



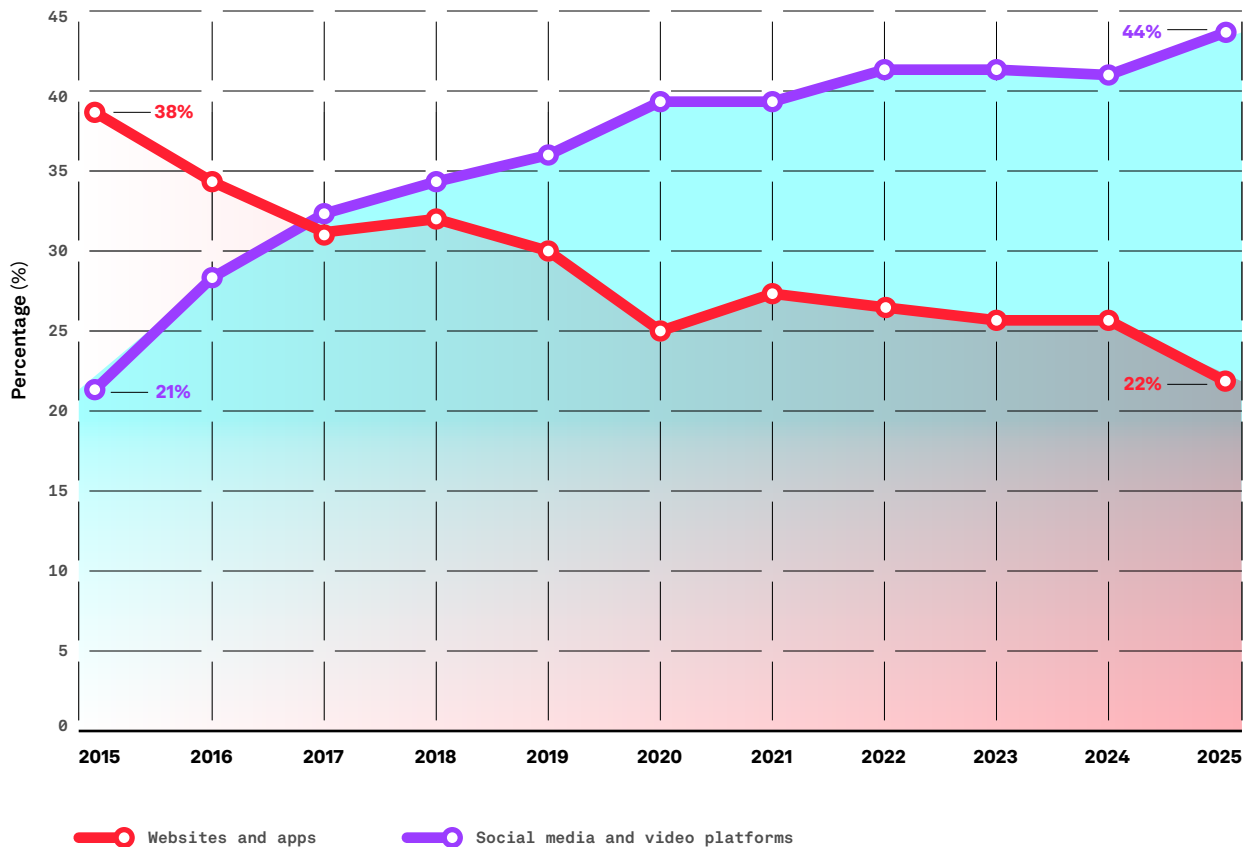
Q Question: You say you’ve used these sources of news in the last week, which would you say is your main source of news?

B Base: Respondents who used a source of news in the last week: 18–24 = 9,807, 25–34 = 15,722, 35–44 = 16,354, 45–54 = 15,804, 55+ = 33,449.

Fig 5.3

Proportion of under-25s who prefer to access news via websites or social networks, 2015–2025

Aggregate data from 48 countries

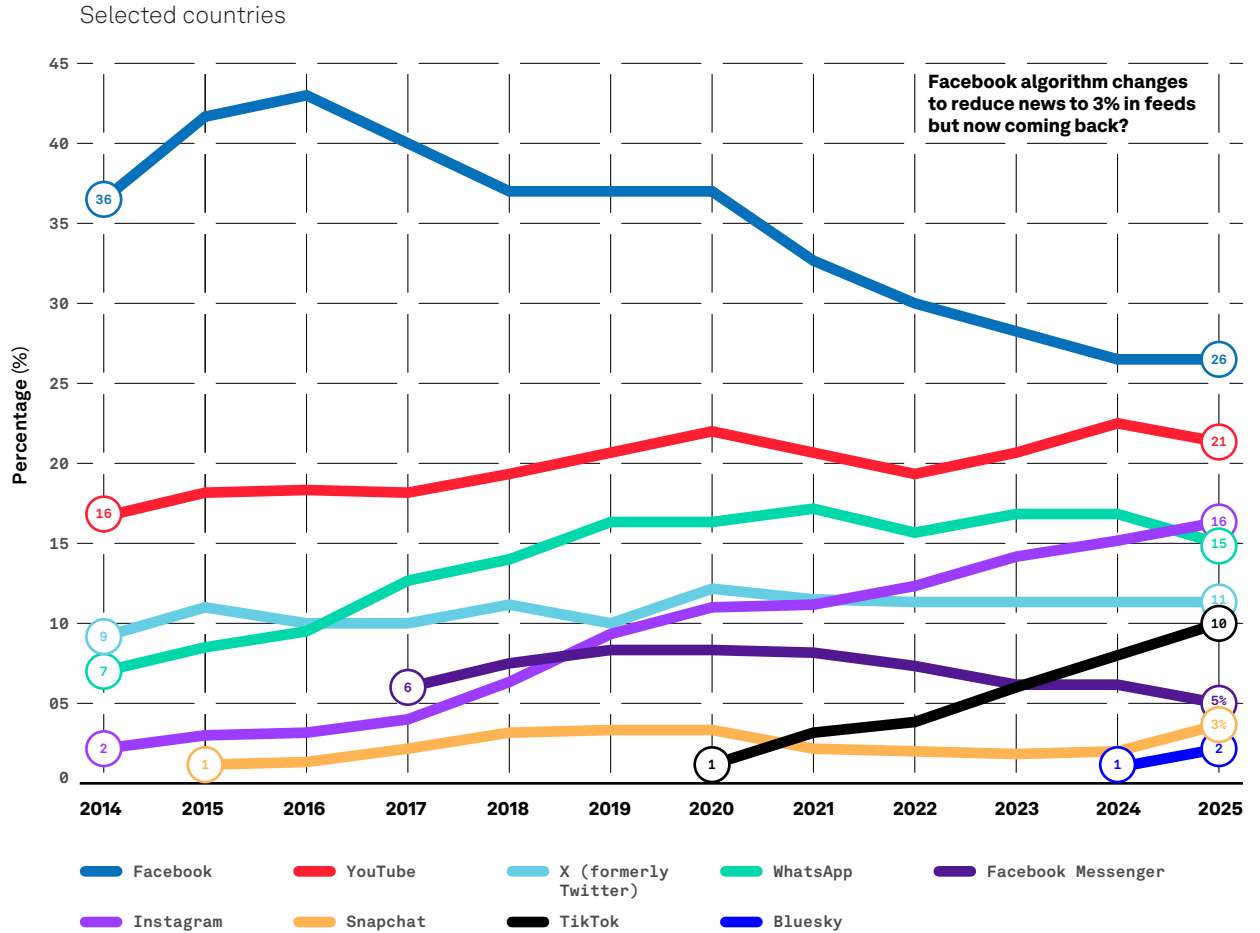


Q **Question:** You say you've used these sources of news in the last week, which would you say is your main source of news?

B **Base:** 18–24 = ranging from 5,598 in 2016 to 9,807 in 2025. Note: Number of markets surveyed in 2016 (26), 2017 (36), 2018 (37), 2019 (38), 2020 (40), 2021 (46), 2022 (46), 2023 (47), 2024 (48), 2025 (48)

The specific platforms used for news have also changed over time. While Facebook remains the main social network for news, its use has declined considerably since 2016. Meanwhile, video-based platforms such as YouTube, Instagram and TikTok have grown in popularity, resulting in more fragmented attention. Across 12 countries tracked since 2014, six networks now have a weekly reach of 10% or more of the population, compared to only two platforms a decade ago.

Fig 5.4 Proportion of respondents who used each as a source of news in the past week, 2014–2025



Q **Question:** Which, if any, of the following have you used for finding, reading, watching, sharing or discussing news in the last week?

B **Base:** Total sample in each country-year in UK, US, Germany, France, Spain, Italy, Denmark, Finland, Australia, Japan (2014–2025), Brazil and Ireland (2015–2025) around 2000. Note: We did not ask about Bluesky in France, Italy, Finland, Denmark, Japan and Canada (2024) and in France, Italy, Denmark and Japan (2025).

Across these crowded digital platforms, news organizations must compete for attention with a range of other voices, many of whom are often more effective at attracting audiences through partisan commentary, engaging story-telling and relatable personalities often perceived as more authentic.

Our data show that traditional news media particularly struggle to retain audience attention on video platforms, where algorithmic recommendations – rather than, e.g., the suggestions of friends – play a dominant role. While audiences still tend to pay more attention to traditional news media and journalists on Facebook, online creators and personalities now outperform established news media on platforms like TikTok and Snapchat.

Even as audiences are embracing platform-based consumption for news, many express concern about the reliability of the information they find there. Facebook (the most used

platform for news) and TikTok (the fastest growing platform for news) are seen as the biggest problems when it comes to misinformation, with around half of global respondents perceiving each as a major threat. Although trust in news overall has declined in many countries, findings repeatedly show that traditional news media are *still* trusted much more than social media or search platforms when it comes to important news.

Across markets, 58% of respondents say they worry about distinguishing true from false news on the internet – a concern that is much higher in Africa and parts of Latin America. When people need to check information that they suspect is false, the most widely cited source is a ‘news brand I trust’ (38%), ahead of official sources, search engines and fact-checking websites. Even among those who say they would verify information via social media and search engines, it is ‘trusted news brands’ that people are most likely to be using. In most countries, these trusted sources tend to be the brands (and websites) of news organizations with a reputation for impartial news such as the BBC in the UK, ARD in Germany and NHK in Japan.

Fig 5.5

Top 3 news brands used to check whether something is true

Trusted brands

UK

Trusted brands



1. BBC News
2. The Guardian
3. Sky News

US

Trusted brands



1. CNN
2. Fox News
3. BBC News

Germany

Trusted brands



1. Tagesschau (ARD)
2. N-tv
3. ZDF/heute

Japan

Trusted brands



1. NHK
2. Yahoo! News
3. Yomiuri Shimbun

Australia

Trusted brands



1. ABC News
2. BBC News
3. 7 News

Q **Question:** *In the previous question you said you would tend to go to a news source you trust to check information. Which one?*

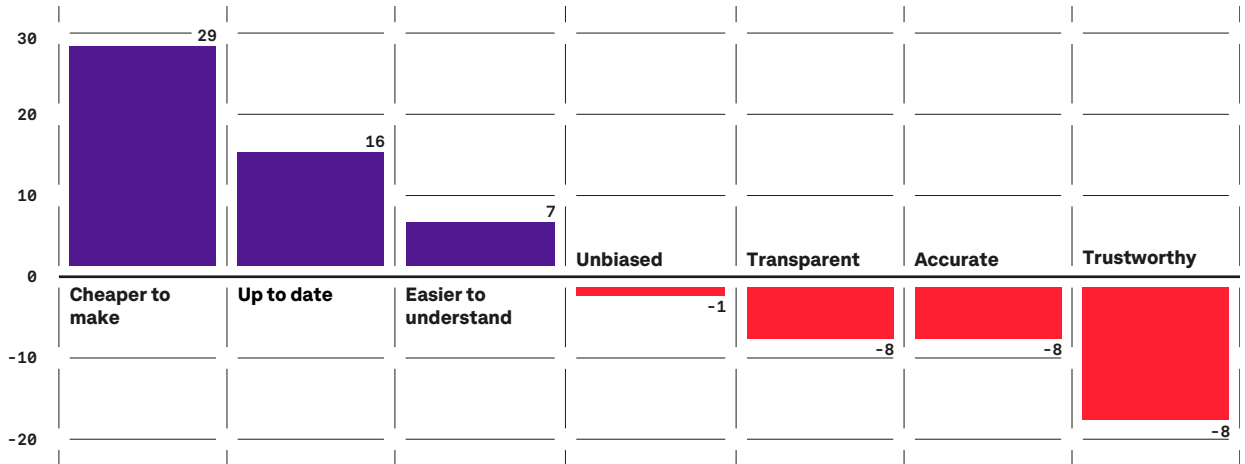
B **Base:** *All those that selected trusted brands in the UK = 867, US = 873, Germany = 811, Japan = 622, Australia = 786.*

Concerns about (hyper) personalization, polarization and misinformation and disinformation are likely to intensify with AI, as synthetic content floods the internet and more audiences use AI to access information. Among respondents, 7% already use an AI chatbot for news on a weekly basis – rising to 15% among those under-25. The integration of AI into widely used tools like search engines may drastically increase the uptake of these technologies, even if users are not always aware they are interacting with AI. Across countries, survey respondents believe AI will make news cheaper to produce (+29 net difference) and more up to date (+16), but less transparent (-8), less accurate (-8) and less trustworthy (-18). These concerns could increase the value of trusted news brands, including their websites and apps, which may serve as reliable anchors for audiences seeking information on contentious and important stories.

Fig 5.6

Net difference between proportion of respondents who think generative AI will make news more or less of each

Average score across 31 countries



Q **Question:** *In general, do you think that news produced mostly by AI, albeit with some human oversight, is likely to be more or less of each of the following, compared to news produced entirely by a human journalist?*

B **Base:** *Total sample across 31 countries = 54,638.*

The outlook for news organizations remains uncertain. Platform- and algorithm-driven access to news is now the dominant behaviour, especially among younger audiences, significantly weakening the traditional so-called ‘post and refer’ model, as platforms increasingly try to retain users within their own ecosystem. The rise of generative AI search interfaces is expected to further challenge efforts to draw audiences back to news websites. However, journalism continues to play an important role in the information environment, especially in times of crisis. To remain relevant, publishers will need to rethink how they reach audiences and develop more sustainable business models.

Looking ahead, a central challenge will be how to safeguard trustworthy information and journalism within an increasingly AI-driven, platform-dominated media environment. Beyond financial sustainability, key questions remain for policy-makers, researchers and civil society at large: How can media literacy be strengthened at scale? How should content provenance be communicated – through labelling or other mechanisms? And how can issues such as platform concentration, data licensing and copyright be effectively addressed in the AI era?

Amy Ross

Research Fellow

Reuters Institute for the Study of Journalism, University of Oxford

Nic Newman

Senior Research Associate

Reuters Institute for the Study of Journalism, University of Oxford

A resilient information ecosystem relies not only on strong institutions, but also on the vitality and independence of media actors. Supporting local journalism, fact-checking networks and open media environments is essential to building public trust and effectively countering harmful information. However, several evolving dynamics are reshaping the information landscape in ways that undermine these efforts:

- The rise of ‘local media deserts’ – areas with little or no access to reliable local news sources – or communities where residents face significantly reduced access to the news of local public discourse.¹² In these environments, people increasingly rely on social media as a primary source of information.
- Shifting media consumption habits are pushing conversations into private, less visible digital spaces – known as ‘bounded social media places’ – such as private messaging apps and closed groups. These spaces are often viewed as more trustworthy due to their known audiences, controlled visibility and real-time synchronicity by facilitating continuous conversations and personalized content sharing, often outside the influence of public-facing algorithms.¹³ However, their opacity makes it harder to monitor or challenge the spread of misinformation.

As highlighted in the World Economic Forum’s 2025 *Global Risks Report*, it is becoming increasingly difficult for the public to discern where to turn for trustworthy information. Across 47 countries, only 40% of people say they trust most news, and there are concerns about the risk of misinformation and disinformation over the next two years, especially in high-income countries. It ranks among the top 5 perceived risks in 13 countries including India, Germany and Canada, and in the top 10 in 30 additional countries.¹⁴

The media plays a dual role in today’s information landscape – acting both as a vital conduit for trusted information and, at times, a vector amplifying harmful narratives. This complexity is critical to understand in humanitarian contexts where media coverage can shape public perceptions of crises, humanitarian actors and the legitimacy and/or effectiveness of aid efforts. Not all media are the same and these distinctions matter:

- Independent journalism plays a watchdog role, conducting fact-checking, holding power to account and providing nuanced, evidence-based reporting.
- State-aligned media often reflect government interests, particularly during armed conflict or emergencies, shaping narratives that support national agendas.
- Commercial and tabloid outlets may prioritize sensationalism and speed over accuracy, driven by attention-based revenue models and algorithmic amplification.
- Digital and social media platforms can accelerate the spread of both accurate information and harmful narratives, amplifying reach through network effects, algorithms and user engagement.

Contributor Insight 5.4

Societal resilience: Bridging the information divide by distinguishing reliable from unreliable news sources

Since 2018, NewsGuard has deployed a team of journalists to rate and review the reliability of news sources across the open web, social media and content platforms, using transparent, apolitical journalistic criteria. The need for such labelling of news sources has grown more acute in recent years as the number of unreliable 'news' sources – often presenting themselves as regular news websites while disregarding basic journalistic standards – has surged.

These range from made-for-advertising websites to propaganda outlets spreading harmful false claims, and more recently, AI-generated 'slop' with no human oversight. For example, ahead of Germany's February 2025 snap federal elections, NewsGuard identified a network of 102 AI-generated German-language sites, spreading false claims with apparent authenticity. By August 2025, NewsGuard had documented over 1,270 such AI-generated sites.

Ratings indicate a source's overall trustworthiness and risk of publishing false content. Assessment and review criteria include whether errors are transparently corrected, whether false or egregiously misleading false claims are regularly published, and whether ownership and potential conflicts of interest are disclosed. Each rating is supported by a detailed 'Nutrition Label,' [NewsGuard's term for its ratings profile] explaining the evidence, citing examples of problematic content and including any response from the publisher. NewsGuard's ratings currently cover more than 36,000 online sources across nine countries (Australia, Austria, Canada, France, Germany, Italy, New Zealand, the UK and US), representing more than 95% of online engagement with news in these markets.

Chine Labbe

Senior Vice President, Partnerships, and
Managing Editor, Europe and Canada

NewsGuard Technologies

Meanwhile, media outlets themselves are under growing pressure. Trends documented in the *World Press Freedom Index*¹⁵ reveal shrinking press freedoms, alongside rising harassment, censorship and targeted information attacks. Many local and independent outlets face severe economic precarity, leading to newsroom closures and weakened on-the-ground reporting. In parallel, regulatory tools such as licensing requirements or anti-fake news laws have, in some cases, been used to suppress independent journalism, rather than protect the public from disinformation.¹⁶

In humanitarian settings, the media plays a powerful role in shaping narratives around aid operations. Politicized or misleading coverage can erode public trust in humanitarian actors, hinder access or even place humanitarian personnel, volunteers and affected populations at risk. Despite these challenges, there are also opportunities for constructive engagement. Partnerships with trusted local journalists can enhance the accuracy,

relevance and reach of communication on humanitarian crises. At the same time, media literacy initiatives can help communities distinguish fact from fiction, better understand the risk of harmful information and engage with media content more critically.

Given the diversity of today's media landscape, it is crucial to identify where harmful information is spreading – whether through television, radio, print, digital platforms or social media – and to understand how different communities access and trust information. This requires disaggregated analysis by age, gender, language, geography and other demographic factors, alongside close engagement with community leaders and local stakeholders. As media consumption habits continue to evolve, the need to promote high-quality journalism and empower individuals to navigate an increasingly complex and polarized media environment has never been greater. For the humanitarian sector, this means actively engaging to build understanding of what humanitarian action is and is not, and of the principles that underpin it.

Contributor Insight 5.5

BBC Media Action: A whole-of-society approach is called for

While humanitarian agencies often see the most damaging effects of harmful information during crises, the problem can pre-date the crisis, and sustainable solutions require a whole-of-society approach in which governments, civil society, independent media, technology companies, the private sector and communities each play a role. The humanitarian sector should be more proactive in supporting such approaches, but this demands a step-change. Agencies such as UNHCR, IFRC and others have an essential role in embedding information integrity strategies across their work – not just in communications departments, but as part of their operations, as well as in core humanitarian standards and response plans.

This includes building trusted relationships with local media and community communicators before crises, and embedding these efforts into coordination mechanisms with other societal actors. In practice, this means:

- working with governments to ensure crisis information policies protect rights and reach those most at risk
- partnering with civil society to strengthen community-led responses
- collaborating with and supporting independent media to uphold accuracy and trust
- engaging with technology companies to address harmful content and increase transparency
- demanding adequate legal and regulatory protections from harmful information

- equipping communities with the skills and resources to navigate an increasingly complex information environment.

It is vital that humanitarian actors avoid working in isolation and draw on the existing work and experience of others working on these issues. For example, the media development and democracy-support sectors have been grappling with information integrity issues for many years and have many lessons and best practices (and ongoing initiatives) to share. Linked to this, humanitarian actors should ensure they are leveraging and engaging with global efforts to build normative frameworks in support of information integrity. For example, the UN Global Principles for Information Integrity were launched in 2024, but implementation is a work in progress.

The rapid evolution of AI makes this engagement even more urgent. AI is already transforming the information environment – accelerating the creation and spread of false content, but also offering new tools for verification, translation and rapid information sharing. Decisions being made now about AI governance, design and regulation will profoundly influence how information flows in the next decade. The humanitarian sector must be at the table in these debates, advocating for safeguards that protect vulnerable populations and exploring solutions such as content provenance technologies (e.g., C2PA) to ensure crisis information is identifiable, verifiable and trusted.

Ultimately, harmful information is not a peripheral issue: it is a core risk to effective humanitarian action. Addressing it demands that agencies treat information integrity as part of critical infrastructure – planned for, invested in and maintained with the same seriousness as water, shelter and health systems. By contributing their expertise, partnerships and grounded understanding of at-risk communities to whole-of-society approaches, humanitarian actors can help shape an information ecosystem where truth has a fighting chance – and where communities are empowered to act on it when it matters most.

Alasdair Stuart
Head of Policy
BBC Media Action

Contributor Insight 5.6

When collaboration between media and humanitarians can save lives

In humanitarian crises, communication can be life-saving – or dangerously misleading. And while humanitarian organizations and local media share a commitment to minimizing the impacts of harmful information on vulnerable communities, their relationship is often fraught with tension – too often unaware of how complementary their ambitions truly are.

Humanitarian organizations are guided by neutrality, dignity and 'do no harm'. Local media seek to inform, be objective and hold power to account. These goals aren't oppositional – they're synergistic. But without dialogue, misunderstandings flourish. Journalists may misinterpret operational silence as evasiveness; humanitarian actors may view media

coverage as oversimplified or sensationalist. In the chaos of a crisis, both believe they're helping – yet mistrust and poor collaboration can inadvertently fuel misinformation, leaving affected communities to bear the cost.

When these sectors collaborate, the impact can be powerful. Humanitarian actors bring credible data and balanced contextual understanding; local media offer reach, cultural insight and familiar, trusted voices.

Community-based organizations often navigate this space more naturally. As part of the community, they tend to see local journalists not as outsiders, but as neighbours, former classmates and peers. While this common ground doesn't guarantee a flawless relationship, it does offer a foundation built on mutual lived experience. In contrast, when international agencies arrive, local media is frequently viewed as an unpredictable element, an unacceptable risk to be 'managed', rather than a partner to engage.

At Internews we frequently saw this dynamic and in response we developed the *Information and Risks: A Protection Approach to Information Ecosystems* toolkit. This was designed in collaboration with a global advisory board of protection agencies, the Global Protection Cluster and local media organizations. It provides tools for humanitarian workers, journalists and community organizations to jointly assess how poor information access can exacerbate protection risks. This tool can shine a light on the shared values, approaches and practical ways to collaborate across all disciplines to address harmful information in a crisis situation.

Building meaningful relationships with media requires more than stage-managed community visits. It requires time, mutual respect and an openness to sharing and acknowledging each other's contributions and constraints in a crisis to transform this relationship from transactional to strategic. Some distance must be maintained. Local media must maintain independence to fulfil their accountability role – which they cannot do if they become too entangled in humanitarian operations or funding streams. Humanitarians cannot share all information with the media. But there is a need for spaces of curated neutral ground, where actors in both sectors can engage authentically, without the pressure of a live microphone.

In an era when information can either save lives or spread harm, building trusted partnerships with non-traditional actors isn't optional. It's a necessary part of protecting those most at risk.

Irene Scott

Humanitarian consultant and former Humanitarian Director

Internews Network

5.6 A threat to humanitarian action and to humanity itself

Harmful information undermines trust, deepens social divisions and weakens the ability of institutions – humanitarian or otherwise – to respond effectively in times of crisis. Addressing it requires a whole-of-society response, not just a humanitarian one. It is

intrinsically linked to preserving humanitarian space and the urgent need to prevent the instrumentalization of humanitarian action and actors. Harmful narratives threaten the perceived legitimacy of humanitarian organizations, shaping public perceptions in ways that can directly hinder their ability to access and respond to populations in need. In some cases, this involves misrepresenting an organization's neutrality, intent or operations to incite hostility and disrupt response efforts.

Recognition of this threat is growing. In a significant step, the UN Security Council adopted Resolution 2730¹⁷ in May 2024, spearheaded by the Government of Switzerland and co-sponsored by 98 member states. The resolution explicitly condemns "disinformation, information manipulation, and incitement to violence" against humanitarian and UN personnel. It further raises concern over the increasing use of malicious information and communication technologies including data breaches and information operations that target humanitarian organizations, disrupt their relief operations, undermine trust and threaten the safety and security of their personnel, premises and assets. The resolution encourages member states and the UN system to take appropriate action to address the increasing threat of disinformation campaigns and misinformation that undermine trust in humanitarian organizations, put personnel at risk and hinder humanitarian activities. Importantly, the resolution reaffirms the need for all parties to armed conflict to preserve the ability of humanitarian organizations to act in a manner consistent with the principles of humanity, neutrality, impartiality and independence. This is vital for delivering assistance to persons in need and for ensuring their protection and safety, as well as that of humanitarian personnel.

Contributor Insight 5.7

UK's commitment to protecting humanitarian action from harmful information

Information integrity is critical in humanitarian crises: communities are vulnerable and need to know what they can trust. Access to accurate and timely information for people in the midst of armed conflict can mean the difference between life and death. Conversely, false narratives targeting humanitarian organizations pose a serious threat to perceptions of the neutrality of aid workers and relief operations, damaging community acceptance, restricting humanitarian access and increasing risks for aid workers. And we need truthful reporting of humanitarian crises globally: the absence thereof undermines public support for humanitarian action.

Yet disinformation in the humanitarian space is growing. The UK is tackling this concerning trend directly – through our seat in key multilateral forums, our diplomatic channels and our UK-funded aid. For instance, the UK co-sponsored UN Security Council Resolution 2730, explicitly condemning disinformation and encouraging member states and the UN system to take appropriate action to address the increasing threat of disinformation campaigns and misinformation that undermine trust in UN and humanitarian organizations and put humanitarian personnel at risk. The Political Declaration for the Protection of

Humanitarian Personnel, developed by a Ministerial group led by Australia and including Brazil, Colombia, Indonesia, Japan, Jordan, Sierra Leone, Switzerland and the UK, commits states to take practical action that counters misinformation and combats disinformation, information manipulation and hate speech targeting humanitarian organizations, personnel and activities. This includes efforts to de-politicize humanitarian action, including by building understanding with local authorities and the media, protecting the independence of journalists, raising awareness, and calling out actors that perpetuate disinformation and hate speech and working with technology companies to support these efforts.

As we continue to learn from experience and adapt to evolving threats, the UK is also investing in:

- independent journalism and fact-checking in fragile contexts
- civil society initiatives to counter harmful information and build resilience to it
- accountability mechanisms to promote community feedback and inclusive information ecosystems

We also invest in research, such as a recent project with Grand Challenges Canada, which highlighted the importance of working with communities to build trust, enhance resilience and empower local leadership, and of identifying technological innovations that are scalable and adaptable. These lessons are shaping our humanitarian responses.

The UK remains committed to working with partners to counter harmful information, protect humanitarian personnel and build resilient communities. Together, we can ensure that humanitarian action is guided by truth, trust and integrity.

Laure Beauflis

Director Humanitarian, Food Security and Resilience

Foreign, Commonwealth and Development Office, Government of the United Kingdom of Great Britain and Northern Ireland

The UN General Assembly's resolution on 'Countering disinformation for the promotion and protection of human rights and fundamental freedoms' calls on states to counter all forms of disinformation through policy-based measures, including public education, digital literacy and capacity-building initiatives.¹⁸ As highlighted in [Chapter 4, on page 157](#), the Pandemic Agreement is a landmark international treaty aimed at setting global standards for pandemic prevention, preparedness and response. It underscores the importance of trust, transparency and timely information sharing in effective pandemic communication.¹⁹

5.7 Red Cross and Red Crescent Appeal to States

Since the Statutory Meetings of the International Red Cross and Red Crescent Movement began in 1867, themes related to information have rarely been addressed.²⁰ However, 2024 marked a turning point with the adoption of an Appeal to States to take action against harmful information and dehumanizing rhetoric.



Appeal to States:

“We appeal to States to take all appropriate measures to prevent, stop and remedy any abuse, pressure, misinformation, disinformation and dehumanizing rhetoric, through social media or otherwise, that harms the physical, psychological or reputational wellbeing of people in vulnerable situations and the staff and volunteers of the Movement components serving them.”

The appeal highlights the increasing impediments faced by principled humanitarian actors in delivering protection and assistance to people in need including the spread of misinformation and disinformation that imperils humanitarian workers and people in their care.²¹ While previous resolutions have reflected the Movement’s longstanding commitments to provide quality information and combat discrimination, the 2024 appeal reflects a heightened sense of urgency. This shift underscores the importance of the information environment, while reinforcing the need to safeguard unhindered and safe access to people in need to carry out principled humanitarian action.

This resolution (appeal) recognizes that harmful information and dehumanizing rhetoric pose serious risks to people in humanitarian need and humanitarian personnel. Addressing this is thus integral to the protection of people and upholding humanitarian operations.

In disasters, crises and other emergencies, laws, policies and plans that promote the availability and integrity of information – and address the challenges of harmful information – are essential to enable and facilitate effective humanitarian action. Harmful information that obstructs humanitarian response, fuels panic and erodes trust runs counter to these objectives. By integrating measures to counter harmful information, states can uphold their responsibilities while protecting the space for neutral, impartial and independent humanitarian action.

Contributor Insight 5.8

Harmful information and international humanitarian law

While there are no specific provisions on harmful information in international humanitarian law (IHL), this body of international law does regulate the spread of certain forms of information. Specifically, IHL prohibits the encouragement of IHL violations, including war crimes, by states and parties to armed conflict, whether online or offline. It also prohibits “acts or threats of violence, the primary purpose of which is to spread terror among the civilian population”.²² This means that threatening violence is prohibited under IHL if the primary purpose or intent of such activities is to spread terror among the civilian population.

Furthermore, IHL prohibits inciting violence against medical services and humanitarian operations. In particular, spreading disinformation intended to obstruct or frustrate medical and humanitarian work is difficult to reconcile with IHL and may violate the obligation of parties to conflict to respect and protect humanitarian personnel and medical services. Moreover, states must not only abstain from such activities but also protect impartial humanitarian organizations from threats posed by other actors within their jurisdiction or control, including private persons and companies. Propaganda aimed at recruiting children, and publishing images of prisoners of war, are in almost all cases violations of IHL.²³

While not all harmful information during armed conflict falls within the scope of IHL, for over 160 years, there has been consensus that impartial humanitarian operations and the personnel involved therein must be respected and protected. Additionally, advocating hatred that constitutes incitement to hostility, discrimination or violence, and inciting genocide, may also violate international human rights law or other rules of international law.

Philippe Stoll

Lead on Movement Initiative on Harmful Information

International Committee of the Red Cross

Contributor Insight 5.9

Disaster law and harmful information

Legal and policy frameworks for disaster risk management (DRM) provide the foundation for preventing and mitigating risks, preparing for crises and ensuring effective response and recovery in societies when disasters strike.²⁴ To remain effective, these frameworks need to also address emerging challenges that undermine their objectives – such as the spread of harmful information.

Misinformation, disinformation and rumours can heighten vulnerability and disrupt every stage of the DRM continuum. Harmful narratives may undermine prevention and mitigation initiatives and distort public understanding of hazards. False narratives may discourage communities from taking preparedness measures or erode trust in early warning systems.²⁵ During response, they may delay life-saving action, undermine coordination among responders or incite hostility toward humanitarian actors.²⁶ In recovery, harmful information can fuel stigma, exacerbate inequalities and obstruct community resilience.

These growing challenges demand greater attention within legal and institutional arrangements for DRM. Addressing information risks requires legal and policy frameworks that: (1) recognize information integrity as a core component of DRM; (2) protect individuals by balancing rights with measures to mitigate harmful falsehoods; and (3) enhance coordination, accountability and public trust through mechanisms for information verification, community engagement and transparent communication.²⁷

Safeguarding information integrity is critical to ensuring that DRM efforts truly protect and empower communities. States should therefore take harmful information into account as a potential barrier to effective DRM and in their efforts to strengthen legal preparedness for disasters.

Isabelle Granger

Global Lead, Disaster Law and Auxiliary Role

IFRC

5.7.1

The evolving place of information in humanitarian standards

The 1994 Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief makes little reference to information, except in relation to the availability of data “pertinent to the implementation of effective disaster response.” Building on this, the 1997 Humanitarian Charter highlighted the “transparency of information and decision-making” as essential to effective and accountable humanitarian action. Subsequent frameworks have placed growing emphasis on information as central to dignity, participation and protection.

The Humanitarian Charter and Sphere Standards (1997–present) identify access to information as a core element of participation; the Core Humanitarian Standard on Quality and Accountability (2014) includes explicit commitments to information-sharing and communication with affected people; and both the ICRC Rules on Personal Data Protection and IFRC Policy on the Protection of Personal Data frame personal data and information management as integral to protection, dignity and legal compliance. In parallel, the CDAC Network Guidelines advanced the principle of ‘communication as aid,’ focusing on community engagement and misinformation management, while the International Civil Society Centre’s Solidarity Action Network has more recently turned attention to the impact of disinformation campaigns on civil society.

Contributor Insight 5.10

Personal data protection as a defence against harmful information

Personal data has become a highly exploitable commodity in disinformation campaigns, used to create, target and disseminate harmful content through AI-generated media such as deepfakes (see [Annex I: Glossary, on page 353](#)) or via large-scale unauthorized data harvesting on social platforms. Where specific regulation surrounding targeted disinformation or AI is absent, existing data protection laws – now in place in over 100 countries – can provide a foundational layer of defence. While limited in scope, these laws require consent, transparency and restrictions on data sharing, and offer individuals enforcement mechanisms.

As the humanitarian sector increasingly relies on the collection of personal data and digital tools to deliver assistance, organizations must take proactive measures to protect staff and beneficiary data. This includes minimizing data collection, enforcing adequate security measures, carefully selecting digital service providers and managing the risks of publicly shared content.

Disinformation campaigns targeting humanitarian operations, such as allegations of fraud or misuse of funds, may also trigger government or donor demands for greater data disclosure in the name of accountability. In such cases, organizations need counter-disinformation strategies that increase transparency without compromising humanitarian principles or individual privacy protections.

Manuela Cardoso

Consultant, Legal Department and Data Protection Office

IFRC, Geneva

Contributor Insight 5.11

Humanitarian standards in a changing information ecosystem

Government policies and regulations on digital platforms, artificial intelligence and data governance are not keeping pace with the rate of technological change and are fragmented across jurisdictions. In the US, for example, Section 230 of the Communications Act of 1934 provides broad immunity to service providers for content generated by users of their platforms. For humanitarian organizations, this creates a heightened responsibility to safeguard communities from harmful information while navigating a shifting and uneven regulatory environment.

The **Sphere Minimum Standards**, along with others in the **Humanitarian Standards Partnership (HSP)** portfolio, are built through consensus and grounded in international law, rights, evidence and practical experience. They distil vast knowledge into accessible guidance for frontline practitioners and others. Yet, in an era defined by fast-moving digital innovation and disinformation, clear gaps remain. New Connectivity As Aid standards are coming soon, but the HSP portfolio still lacks sufficient guidance on data stewardship, intellectual property and the use of AI. Addressing these areas is urgent, as harmful information increasingly undermines trust, accountability and the protection of crisis-affected people.

Sphere, as host of the HSP, recognizes the need for inclusive and consultative development of new standards to meet these gaps. Future standards should highlight how harmful information, accelerated by technology, disproportionately affects vulnerable populations and provide practical guidance to mitigate these risks. Strategic partnerships will be essential. Sphere encourages standards developers to collaborate with organizations committed to meaningful community engagement, such as **People First Impact Method (P-FIM)**, to ensure that community voices shape new standards from the outset. In a complex and fast-changing information ecosystem, humanitarian self-governance through widely owned and versatile standards will save lives. This is one way the sector can and must work together to build trust, uphold truth and strengthen resilience in the face of harmful information.

Tristan Hale

Director of Operations

Sphere Standards

Taken together, these instruments underscore that information is not peripheral to humanitarian action but fundamental to its effectiveness, accountability and legitimacy. Looking ahead, this recognition should also encompass harmful information, ensuring that standards explicitly address how it can undermine safety, dignity, humanitarian access and action response.

5.7.2

The tech sector: Platforms, power and responsibility

The explosive growth of online content, driven by new technologies and user-generated platforms, has made harmful information more prevalent and harder to detect, moderate and remove. What began as spaces for personal connection have evolved into powerful tools for influence, distortion and manipulation, shaped by the speed, scale and reach of digital information. Social media platforms are designed to reward engagement, not accuracy. What goes viral is not necessarily what is true. Built around attention-based algorithms, these platforms fuel contests over narrative and influence with real-world consequences. These platforms are designed for maximum engagement, where every like, share and comment release a hit of dopamine, reinforcing addictive use and deepening echo chambers.²⁸

The rise of AI has further concentrated power in the hands of a small number of companies that develop and control the systems including algorithms shaping what billions of people see, search for and share. Their influence extends beyond access – these systems increasingly affect how public opinion is formed, how issues are framed and which narratives gain prominence. Generative AI has drastically lowered the barriers

to creating and distributing false or misleading content, across text, video, audio and images. Distinguishing AI-generated from human-made content is increasingly difficult. While some material is simply the result of error or AI hallucination, other content is deliberately produced by threat actors – state actors, activist groups or individuals – aiming to mislead, influence or disrupt. This rapid, low-cost production and deployment of synthetic content presents significant challenges for information integrity.

5.8 UN action on AI and information integrity

The UN General Assembly's March 2024 Resolution on AI systems²⁹ highlights the growing risks to **information integrity**, access to information and human rights. It warns that the improper or malicious design, development, deployment and use of AI systems could undermine the Sustainable Development Goals (SDGs), deepen digital divides and reinforce structural inequalities and biases. The resolution raises concerns about the potential accidents and the risk of compounded threats from malicious actors. It encourages member states and stakeholders to develop effective and interoperable tools, reliable content authentication and provenance mechanisms – such as watermarking or labelling. These measures would enable users to identify information manipulation, determine the origins of and distinguish between authentic and AI-generated or manipulated digital content. It further encourages efforts to strengthen media and information literacy to strengthen societal resilience.³⁰

The June 2023 UN policy brief on **Information Integrity on Digital Platforms** was the result of consultations developed under the auspices of the Secretary-General as part of the Our Common Agenda and Summit of the Future initiatives. This seeks to provide a concerted global response to information threats in the information environment. The principles aim to guide member states, digital platforms and other stakeholders in fostering a more inclusive, transparent and safe digital space, particularly freedom of opinion, expression and access to information.³¹ In September 2025, the UN released the first in a new Issue Brief series entitled **From Principles to Practice: Strengthening Information Integrity**.

The UN's 2024 *Governing AI for Humanity* report³² warns that large parts of the world are excluded from international AI governance conversations, with only seven countries party to seven prominent non-UN AI initiatives and 118 countries not party to any.³³

5.9 Content moderation and the power of platforms

Content moderation has become one of the most powerful and contested functions exercised by global technology platforms. Companies behind social media and messaging services play a central role in shaping the information ecosystem, determining what

is amplified, removed or monetized. The model of social media platforms – functioning as intermediaries allowing user-generated content to spread at scale – has effectively shifted much of the responsibility for identifying and managing harmful information onto the private sector. This places content moderation and governance in the hands of technology companies rather than public institutions. Content moderation is carried out under each platform’s own terms of services and definitions, typically through a combination of human moderators and automated systems. Platforms may block or remove content either in response to legal requests or based on internal assessments. However, definitions of harm vary and enforcement is uneven.

This power is exercised with limited transparency, especially regarding how content is moderated and curated. Many governments have been reluctant to regulate legal but harmful information citing their obligations to uphold freedom of expression. At the same time, global technology platforms often comply with national content laws, even when those laws conflict with international human rights standards. This results in an uneven and fragmented landscape, where users’ access to content depends largely on their location and the legal environment of that jurisdiction. While some platforms publish transparency reports outlining how often governments request content removal, reporting practices vary widely and often lack the detail necessary for meaningful public accountability.

While content regulation can be lawful and necessary, especially to prevent harm, some states use it to control narratives, limit transparency and suppress dissent. This raises a deeper set of questions: Who decides what constitutes harmful information? Who is accountable? And how can we safeguard fundamental rights in a fragmented and global information space?

Under international law, for example, Article 19 of the International Covenant on Civil and Political Rights, freedom of expression may only be restricted when three conditions are met: the restriction must be provided by law (clearly defined in legislation), necessary (to achieve a legitimate aim) and proportionate (not excessive or overly broad). Yet many current censorship practices fail to meet these criteria, raising serious concerns about compliance with international standards.

The UN Guiding Principles on Business and Human Rights and related standards provide important guidance for technology companies. These principles call for heightened human rights due diligence and risk management to prevent and address adverse impacts, including from the spread of harmful information. This is particularly critical in contexts affected by humanitarian crises, where such information can exacerbate vulnerabilities and fuel harm. Companies are expected to invest in and implement robust measures to mitigate potential negative impacts of their activities on people’s safety and dignity. This includes ensuring that their policies, procedures and practices are consistent with international human rights standards and take into account international humanitarian law.

Complementing this, the UNESCO framework for regulating digital platforms is grounded in the recognition that information is a public good. As such, it calls for government action to protect and support the integrity, accessibility and equitable availability of information ecosystems.³⁴

Meta (formerly Facebook), for example, does not explicitly list humanitarian organizations as targets in its reporting on inauthentic behaviour. Yet, its findings on narrative manipulation and inauthentic behaviour in contexts of armed conflict have clear

implications for humanitarian actors, whose operations may be disrupted or delegitimized by such tactics.³⁵

Microsoft highlights that not all threats posed by AI originate within the systems themselves; many stem from the broader information ecosystems.³⁶ Two prominent categories stand out:

- **Impersonation:** AI-generated deepfakes (audio, video or images – see [Annex I: Glossary, on page 353](#)) increasingly enable impersonation of individuals, with serious risks including fraud, blackmail, coercion, defamation and information warfare.
- **Content production:** AI tools can be misused to produce harmful synthetic content at scale, including disinformation, spam, non-consensual intimate imagery and grooming scripts. These threats are often amplified versions of pre-existing problems, now rendered faster and more pervasive by generative AI.

Microsoft has also called on states to adopt clear norms and restrictions to curb harmful foreign influence operations, particularly those targeting crisis and emergency contexts, humanitarian and emergency response organizations, elections, marginalized communities and protected groups such as ethnic minorities and LGBTQ+ populations. It advocates for limits on the use of certain tools and techniques including synthetic media (e.g., deepfakes) and emphasizes that social media data from foreign citizens should not be exploited for influence operations.³⁷

5.10 From self-regulation to state oversight: The evolving governance of online content

Despite their language of *community*, digital and social media platforms are, above all, businesses. Their primary responsibility is to shareholders, not the public good – so the metrics that matter most are financial. This profit-driven logic shapes how platforms approach both problems and solutions. With an engineering-first mindset, they often frame even complex societal challenges as design problems solvable through technical fixes. Ultimately, platform terms of service – written by private companies – administer communities of unprecedented scale.³⁸ Yet fundamental dilemmas remain: Should these companies restrict the flow of information? What content should be restricted, how and by whose standards?

The US legal framework, particularly Section 230 of the Communications Decency Act, has played a central role in shaping this landscape.³⁹ It shields US platforms from liability for user-generated content while also protecting them when they moderate content in “good faith.” This approach effectively left much of the internet to self-regulate, with content moderation emerging as a corporate practice – not out of legal obligation, but as a strategy to pre-empt stronger regulation. In contrast, enforcement around intellectual

property – governed by the more stringent 1998 Digital Millennium Copyright Act has drawn much clearer lines.⁴⁰

The European Union's Digital Services Act, which came fully into force in 2024, establishes a comprehensive legal framework to regulate online services and address the spread of illegal and harmful content, including misinformation and disinformation, while safeguarding fundamental rights such as freedom of expression. Non-compliance can result in significant penalties, including fines of up to 6% of a company's global annual turnover. The Digital Services Act seeks to strike a balance between mitigating systemic societal risks associated with online content and protecting freedom of expression and information.⁴¹ Key provisions relevant to harmful information include:

- 1 Systemic risk management:** 'Very large online platforms' (with more than 45 million active users in the EU) and 'very large online search engines' are required to identify, analyse and mitigate systemic risks stemming from the design, functioning and use of their services. This includes risks posed by disinformation that could threaten democratic processes, public health or public security.
- 2 Transparency and accountability:** Platforms must disclose their content moderation policies, advertising practices and key parameters of algorithmic decision-making and provide users with accessible complaint-handling and appeal mechanisms.
- 3 Code of practice:** In February 2025, the strengthened EU Code of Practice on Disinformation was formally integrated into the Digital Services Act framework. Measures include demonetizing disinformation, strengthening fact-checking and improving access to reliable and authoritative information.
- 4 Enhanced user-reporting mechanisms:** Online platforms must provide accessible and effective systems for users to report illegal content.
- 5 Crisis-response mechanism:** In exceptional circumstances, such as threats to public security or public health, the European Commission may require platforms to take targeted measures to address the rapid dissemination of harmful information during crisis periods.⁴²

These developments build on earlier EU initiatives, including the European Commission's Code of Practice on Disinformation and the Strengthened Code of Practice on Disinformation.⁴³

5.11 Framing a response: Supply and demand solutions to disinformation

Professor Anya Schiffrin of Columbia University developed a 'taxonomy of solutions' – an analytical framework that categorizes various policy responses to online misinformation and disinformation into two broad approaches: supply side and demand side.

Contributor Insight 5.12

Understanding the universe of fixes for online misinformation and disinformation

Since 2016, scholars, policy-makers and practitioners have all been worried about the spread of misinformation and disinformation online. The COVID-19 pandemic and associated spread of vaccine resistance has made the problem even more urgent in the public health and humanitarian spheres. To better understand policy proposals, I developed an analytical framework and was the first to create a taxonomy of solutions which distinguishes between ‘supply-side’ and ‘demand-side’ proposals,⁴⁴ many of which have since been implemented around the world.

Demand-side solutions emphasize the role of the consumer, while **supply-side solutions** emphasize the supply of information, looking more to its producers, suppliers and purveyors. In the taxonomy, supply-side solutions fall into two subcategories: 1) suppressing poor-quality, dangerous or illegal information; and 2) creating and/or promoting high-quality information either online or by supporting journalism.

Demand-side solutions include efforts to teach media literacy, journalist efforts to engage audiences and verification efforts such as labelling and fact-checking. These solutions all emphasize audience demand for information and the role of individual choice. Media literacy attempts to build discernment skills among audiences so they can identify which sources to trust. Similarly, solutions involving community participation include efforts by journalists to build trust by seeking to bolster engagement with reliable, relevant material. Solutions involving verification mechanisms, such as fact-checking and labelling, provide a means of establishing what is objectively correct or ‘true.’ A recent synthesis of hundreds of academic studies carried out by the International Panel on the Information Environment⁴⁵ found that demand-side interventions such as media and information literacy, labelling and publishing corrections appeared effective in more than 10% of studies. However, these solutions are expensive and difficult to scale.⁴⁶ [Table 5.1, on page 199](#) shows the range and classifications of supply-side and demand-side solutions.

Supply-side solutions

While the US has largely followed the demand-side path, many other governments, including in Europe, have put less emphasis on consumer decisions and focused more on the supply of information available to consumers. Germany’s 2019 Network Enforcement Act (NetzDG) attempts to hold social media platforms responsible for combating online speech deemed illegal under domestic law. The European Digital Services Act 2022 harmonizes different national laws to address illegal content. Both aim to suppress false or potentially harmful misinformation or disinformation and are supply-side solutions. So is the use of AI to screen and filter information and legal action – defamation suits – against purveyors of falsehoods. Efforts to promote good-quality journalism, provide accurate information through platforms such as Google and YouTube, to support public broadcasters or fund local news are all attempts to boost the *supply* of high-quality information.

There are, of course, overlaps between supply-side and demand-side solutions. Middleware tools serve both sides by shaping what is available (supply) while helping audiences make informed choices (demand). These include NewsGuard’s rating systems on the credibility

of online sources and tools to counter misinformation, and the Journalism Trust Initiative which establishes indicators for the trustworthiness of journalism. Similarly, the international fact-checking movement protects the accuracy of the supply of information but also requires active consumer involvement.

Table 5.1

Taxonomy of solutions – supply side versus demand side

Demand-side solutions

Supply-side solutions

Media literacy training

Assisting audiences to better distinguish between what is true and false

Suppression of poor-quality information

Controlling information flows
Includes content suppression, downranking of content, removal of bots and de-platforming to restrict what information is shown

Promotion of a healthy information ecosystem

Increasing quality information
Attempts by YouTube to increase quality of content and by Google to highlight accurate information

Community participation

Includes journalists' efforts to establish trust in and engagement with high-quality information, as well as to develop citizen journalism

AI and content moderation

Using AI to distinguish between true, false and illegal information and supplementing with human content moderators where required

Advancing AI

Using AI to promote good-quality information, while recognizing current limitations

Fact-checking

Includes labelling and browser extensions that audiences can use

Regulation and laws

Hate speech laws existing in many parts of the world, such as NetzDG and the Digital Services Act

Support for high-quality journalism

Policies such as funding the BBC, giving subscription vouchers, tax credits and subsidies, and relying on innovation funds and philanthropic or donor support

Raising awareness

Reporting on the platforms for a stronger understanding of the effects of misinformation and disinformation

Defamation lawsuits

Demonstrating repercussions for actively spreading lies, such as the 2023 lawsuit U.S. Dominion v. Fox News Network about voting machines

Healthy journalism business model

Engaging with private and public sector adverts to ensure a healthy supply of advertising revenue can fund journalism

Source: Schiffrin, 2017

Dr Anya Schiffrin

Senior Lecturer in Practice

School of International and Public Affairs, Columbia University

5.12 Civic trust and societal resilience

“

So what I would want to suggest is: they should start with the ground level in schools. Educating people in schools, telling them the quality source of information. Once that is done, those children will keep that and they will grow with that kind of mentality. So the education part should continue using the Ministry of Education in schools that should know the correct source of information. Our community meetings like the church and other political gatherings should also be emphasized. But now the only problem is because of these issues, political issues, people want to gain support, they want to use whatever means. But what I would suggest is let the children be taught the correct things from the Ministry of Education first.”

Community member, Zambia

Some countries have adopted whole-of-society approaches to build societal resilience – to inoculate – against information threats. These include education programmes, public tracking of foreign disinformation, election protection measures, transparency for political advertising and campaign activities. In addition, some governments use legal measures, such as laws for the removal of illegal content.

Contributor Insight 5.13

Why Finland is a forerunner in media literacy

For many years, Finland has been regarded as a global leader in developing and promoting media literacy. The small northern European country has attracted international attention and frequent requests to share experiences behind its strong reputation. Some of the key elements of the Finnish approach to media literacy include:

- **Long traditions.** Media education has been part of Finland's democracy and education system for decades. References date back to the 1950s, with many organizations and practitioners accumulating extensive experience. This long-term perspective has ensured continuity in the national approach.
- **Policies and strategies.** Finland was an early adopter of national policies to promote media literacy. It has a dedicated national authority for media education with a statutory mandate to advance media literacy and strengthen resilience as part of comprehensive security. Media education is integrated into curricula across all levels of education, ensuring that children and young people develop these skills throughout their educational path.

- **Cross-sectoral approach.** Media education in Finland is diverse and widely implemented. Schools, public authorities, cultural institutions, NGOs and private sector actors all contribute to the promotion and development of media education.
- **Trust and stability.** Finland consistently ranks first in both the Media Literacy Index and the World Happiness Report. These results reflect a society with high levels of trust in authorities and the media. Citizens' strong critical media literacy skills and digital competences foster inclusion, strengthen stability and safeguard democracy.

Julia Alajärvi

Senior Adviser

National Audiovisual Institute of Finland

International research⁴⁷ highlights several systemic challenges facing civil society actors, which hamper their ability to assess information threats and develop effective responses. While a range of initiatives are emerging, many remain isolated, pointing to an urgent need for stronger coordination, knowledge-sharing and the importance of skill diffusion to prevent duplication and build collective capacity. A critical constraint is the limited access to social media data, which limits civil society's ability to assess the scale of disinformation and to evaluate the effectiveness of different interventions. Furthermore, engagement with technology platforms remains uneven: many organizations, particularly in the Global South, struggle to have their concerns heard and are frequently overlooked by platform decision-makers.

The OECD identifies five key determinants of public trust in government⁴⁸ that may provide some interesting insights for civil society in building institutional trust:

- responsiveness and reliability in delivering services and anticipating needs (reflecting competence)
- perceptions of integrity, openness and fairness (reflecting public values).

Academic literature further distinguishes between: trust in competence – the ability to deliver on expectations – and trust in intentions – the perception that actions are taken in good faith. These two dimensions are interdependent and influence how people evaluate institutions. Trust is ultimately built on both performance and principle – the perceived capability of institutions and the values guiding their actions. Transparency, fairness and accountability serve as mutually reinforcing attributes that strengthen both pillars and are essential for building societal resilience in the face of harmful information.

Concluding remarks: A collective responsibility for preserving principled humanitarian action

For the humanitarian sector, strengthening cross-sector engagement is essential – not only to ensure the flow of trustworthy information but also to identify and respond to harmful information. Disruption alone is not enough; responding effectively requires collective action, similar to the collaborative efforts developed for global health security during the COVID-19 pandemic.

Harmful information is borderless, adaptive and often directed at the people most at risk. Algorithmic systems amplify its spread, while global disparities in data governance, AI capacity and political and societal polarization intensify its reach. Effective responses require greater algorithmic transparency on how content is sorted, ranked, amplified and targeted, combined with approaches that protect privacy, dignity and authenticity.

Resilience must be built at every level – individual, institutional and societal – drawing on community trust, local knowledge, behavioural science, digital and information literacy. Evidence shows that simple interventions, such as digital prompts,⁴⁹ can reduce the sharing of false content, underscoring the value of pairing regulation with context-specific solutions. Governments, platforms, civil society and researchers must work together to test and scale what works. This underscores the importance of pairing regulation with behavioural science and practical interventions that are context specific. Governments, platforms, civil society and researchers must work together to test what works, for whom and in which context.⁵⁰

Humanitarian crises – from pandemics to armed conflict – create fertile ground for harmful information. Information related to the COVID-19 pandemic exposed both the speed of harmful information spread and the cost of delayed action. Meeting the challenges of identifying, preventing and mitigating harmful information requires a systems approach and urgent, cross-sectorial action. Fragmentation between humanitarian actors, digital rights advocates and the technology communities must be bridged. This includes creating space for emerging questions, fostering diverse and underrepresented perspectives, and prioritizing principles and accountability. Strong, forward-looking partnerships are essential, not only to shape governance frameworks, but to influence the design and deployment of tools used to navigate this rapidly evolving landscape.

Ultimately, states must act. As underscored in the Movement's Appeal to States, addressing harmful information is now central to principled humanitarian action. Trusted information may not be water, food or shelter but it is imperative to accessing all three, as well as to ensuring safety, dignity and autonomy.

Asks, aims and recommendations

Asks

Balance the protection of humanitarian space, the safety and dignity of affected populations and the integrity of humanitarian operations with freedom of expression by building rights-respecting, transparent, accountable and resilient information ecosystems.

Aims

Ensure laws and policies protect principled humanitarian action and safeguard trusted and reliable information.

Hold platforms accountable for measures that protect principled humanitarian action and implement transparent, rights-respecting crisis protocols that safeguard affected populations and humanitarian personnel.

Embed harmful information analysis into humanitarian operations to safeguard trust, access and principled humanitarian action.

Engage affected communities in information strategies and work with digital platforms to ensure timely, accurate and safe information flow.

Build trusted information ecosystems through digital and media literacy, local dialogue, initiatives that counter polarization and ongoing monitoring to evaluate and adapt interventions.

Recommendations

States and policy-makers

- Adopt rights-respecting regulations to counter harmful information while safeguarding humanitarian action and trusted information.
- Coordinate with humanitarian actors to ensure laws and regulations protect and respect principled humanitarian action.
- Integrate risks from harmful information that could affect principled humanitarian action into negotiations, frameworks and operational planning.
- Respect and support the independence of National Societies in their auxiliary role, ensuring they can operate without interference.

Regulators and technology companies

- In humanitarian crises, report transparently on harmful content removal, moderation practices and algorithmic adjustments to ensure trust, accountability and prevent harm to people in need and humanitarian personnel and volunteers.
- Collaborate with principled humanitarian actors to design localized mitigation tools such as fact-check bots, verified information hubs and multilingual content.
- Strengthen enforcement against coordinated harmful information campaigns targeting principled humanitarian organizations and the safety and dignity of affected communities, humanitarian personnel and volunteers.
- Develop and apply rights-respecting crisis-response protocols in collaboration with humanitarian actors and share relevant data with humanitarian organizations and research hubs to support trusted, principled action.

Humanitarian actors

- Embed trusted information as essential for principled humanitarian action and for combating harmful information into humanitarian standards and operational frameworks.
- Share evidence of harmful information trends with regulators and platforms in ways that respect data protection rules and support principled humanitarian action to mitigate harmful information, protect humanitarian action and safeguard affected populations, humanitarian personnel and volunteers.
- Partner with local journalists, fact-checkers and trusted content creators to amplify accurate, contextualized and life-saving information for populations in need.
- Strengthen community engagement by integrating media literacy, information resilience and feedback mechanisms into humanitarian programming, ensuring communities can access, understand and act on reliable information.

Communities and local leaders

- Promote community-based media and independent journalism to provide timely, accurate and accessible information.
- Lead digital and media literacy initiatives to build critical thinking and resilience to harmful information.
- Act as trusted intermediaries by disseminating verified information through local channels such as radio, schools and faith institutions.
- Facilitate dialogue to counter polarization, address stigma and prevent harmful information from escalating into violence.

Endnotes

- 1 World Economic Forum. *Global Risks Report 2025* (2025) p.34–35 https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf
- 2 Council of Delegates of the International Red Cross and Red Crescent Movement, Resolution 5: Call for respect and support for principled humanitarian action (2024). https://rcrcconference.org/app/uploads/2024/10/CoD24_R5-Res-NIIHA-EN.pdf
- 3 UN. Universal Declaration of Human Rights (adopted by UN General Assembly Resolution 217 A (III)). 10 December 1948. www.un.org/en/about-us/universal-declaration-of-human-rights
- 4 “Global, free, open, secure, and interoperable access to the Internet is essential during armed conflict, allowing people to request, access, or deliver aid and obtain or share vital information and early warnings. This access also enables humanitarian actors to coordinate operations and distribute information and aid and is equally crucial for post-conflict restoration and peacebuilding.” Freedom Online Coalition. Joint Statement on Protecting Human Rights Online and Preventing Internet Shutdowns in Times of Armed Conflict, adopted June 2025. <https://freedomonlinecoalition.com/joint-statement-on-protecting-human-rights-online-and-preventing-internet-shutdowns-in-times-of-armed-conflict>
- 5 Access Now. ‘Libya floods: people need reliable internet now.’ Press release. 22 September 2023. www.accessnow.org/press-release/libya-floods-internet
- 6 The UN Open-Ended Working Group on Security of and in the Use of Information and Communications Technologies (OEWG), established by the UN General Assembly (A/RES/75/240), is mandated to advance responsible state behaviour in cyberspace and strengthen the security and stability of the ICT environment.
- 7 A number of states have issued position papers on the application of international law in cyberspace, many of which are available through the UN Office for Disarmament Affairs (UNODA) repository and the UN OEWG documentation pages. <https://meetings.unoda.org/open-ended-working-group-on-information-and-communication-technologies-2021>
- 8 Pamment, J. Countering Information Influence Activities: *The State of the Art*. Riga: NATO Strategic Communications Centre of Excellence, (2018).
- 9 Humprecht, E., Esser, F. and Van Aelst, P. Resilience to Online Disinformation: A Framework for Cross-National Comparative Research, *The International Journal of Press/Politics*, 2020:25(3). See also Humprecht, E. ‘Why resilience to online disinformation varies between countries.’ *Democratic Audit* (2020) www.democraticaudit.com/2020/03/24/why-resilience-to-online-disinformation-varies-between-countries
- 10 Reuters Institute for the Study of Journalism. *Reuters Institute Digital News Report 2024* (2024) p.16 https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2024-06/RISJ_DNR_2024_Digital_v10%20lr.pdf
- 11 Newman, N. *Overview and Key Findings of the 2024 Digital News Report*. Reuters Institute for the Study of Journalism (2024) <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/dnr-executive-summary>
- 12 Torre, L., Ramos, G., Noronha, M. et al. Sourcing Local Information in News Deserts. *Media* 2024:5(3), 1228-1243 www.mdpi.com/2673-5172/5/3/78
- 13 Malhotra, P. “What You Post in the Group Stays in the Group”: Examining the Affordances of Bounded Social Media Places. *Social Media + Society* 2024:10(3) DOI:10.1177/20563051241285777
- 14 World Economic Forum. *Global Risks Report 2025* (2025), p.35. https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf
- 16 See, for example, Lim, G. and Bradshaw, S. *Chilling Legislation: Tracking the Impact of ‘Fake News’ Laws on Press Freedom Internationally*. Center for International Media Assistance. (2023) www.cima.ned.org/publication/chilling-legislation
- 17 UN Security Council, Resolution 2730 (2024) on the protection of humanitarian and UN personnel [https://docs.un.org/en/S/RES/2730\(2024\)](https://docs.un.org/en/S/RES/2730(2024))
- 18 UN General Assembly, Countering Disinformation for the Promotion and Protection of Human Rights and Fundamental Freedoms, A/RES/76/227 (2021) <https://undocs.org/A/RES/76/227>
- 19 World Health Organization. *WHO Pandemic Agreement* (Resolution WHA78.1), adopted 20 May 2025, 78th World Health Assembly. The agreement, along with the resolution, sets out the final text agreed by member states and establishes the process for negotiating and adopting the Annex on Pathogen Access and Benefit Sharing. https://apps.who.int/gb/ebwha/pdf_files/WHA78/A78_R1-en.pdf
- 20 The 2022 Council of Delegates Resolution 12: Safeguarding Humanitarian Data (CD/22/R12) acknowledged that cyber operations, data breaches and disinformation pose serious risks to the trust and functioning of impartial humanitarian organizations. The Movement’s Appeal for Respect for Neutral and Impartial Humanitarian Action highlighted how disinformation and misinformation jeopardize the safety of humanitarian workers, influence public perceptions and hinder humanitarian response. The Council of Delegates is one of the Statutory Meetings of the International Red Cross and Red Crescent Movement.
- 21 The spread of misinformation and disinformation is addressed in the Council of Delegates resolution ‘Call for respect and support for principled humanitarian action’ (CD/24/R5). The impact of disinformation on humanitarian organizations is included in the ICT resolution of the International Conference. International Conference of the Red Cross and Red Crescent, Resolution 2: ‘Protecting civilians and other protected persons and objects against the potential human cost of ICT activities during armed conflict’ (34IC/24/R2), adopted 2024. https://rcrcconference.org/app/uploads/2024/11/34IC_R2-ICT-EN.pdf
- 22 Protocol I additional to the Geneva Conventions, Article 51(2), (1977) and Article 13(2) of Protocol II (1977).
- 23 Rodenhäuser, T. and D’Cunha, S. Foghorns of war: IHL and information operations during armed conflict. ICRC Law and Policy Blog. (2023) <https://blogs.icrc.org/law-and-policy/2023/10/12/foghorns-of-war-ihl-and-information-operations-during-armed-conflict>
- 24 See the IFRC Disaster Law website for more information: <https://disasterlaw.ifrc.org/why-disaster-law>
- 25 See Richter, S. ‘True or false? Here are five ways to manage false information about risk.’ UN Office for Disaster Risk Reduction. Feature. 16 December 2022. www.undrr.org/news/true-or-false-here-are-five-ways-manage-false-information-about-risk.
- 26 See European Commission. *Information manipulation and misinformation: A threat for EU civil protection and humanitarian aid*. <https://civil-protection-humanitarian-aid.ec.europa.eu/resources-campaigns/information-manipulation-and-misinformation>. See also Prevention-Web. *Common Challenge: Misinformation and disinformation on disaster risk communication – Practical tips*. www.preventionweb.net/hubs/disaster-risk-communication-hub/do/misinformation-disinformation
- 27 Although the IFRC’s Disaster Risk Governance Guidelines (<https://disasterlaw.ifrc.org/DRMguidelines>) do not address harmful information in DRM directly, they offer useful guidance on relevant topics such as enhancing disaster risk knowledge and education, protecting the rule of law in disasters, and enhancing coordination across all elements of the DRM continuum.
- 28 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) pp.3, 16, 19, 62
- 29 UN General Assembly Resolution A/78/L.49: Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, adopted 21 March 2024.
- 30 Ibid, section g.

- 31 UN. *Information Integrity on Digital Platforms (Our Common Agenda Policy Brief 8)*, Executive Office of the Secretary-General. (2023) www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf, www.un.org/en/information-integrity
- 32 UN. *Governing AI for Humanity: Final Report*. (2024) p.29 www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf
The UN High Level Advisory Body on Artificial Intelligence (HLAB-AI) established by the UN Secretary General was composed of 32 experts and aims to align AI development with human rights and the SDGs. High-Level Advisory Body on Artificial Intelligence. UN Office of the Secretary-General's Envoy on Technology. www.un.org/digital-emerging-technologies/ai-advisory-body
- 33 Ibid. p.8. The report shows seven prominent non-UN AI initiatives. Seven countries are parties to all the sampled AI governance efforts, whereas 118 countries are parties to none (primarily in the global South). The sample comprised OECD AI Principles (2019), G20 AI Principles (2019), Council of Europe AI Convention drafting group (2022–2024), GPAI Ministerial Declaration (2022), G7 Ministers' Statement (2023), Bletchley Declaration (2023) and Seoul Ministerial Declaration (2024). Canada, France, Germany, Italy, Japan, UK and USA are parties to all sampled initiatives/instruments.
- 34 UNESCO. *Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information*. (2023) www.unesco.org/en/internet-trust/guidelines
- 35 Gleicher, N. 'Removing Coordinated Inauthentic Behavior.' Meta 8 July 2020. <https://about.fb.com/news/2020/07/removing-political-coordinated-inauthentic-behavior/>; Clegg, N. 'What We Saw on Our Platforms During 2024's Global Elections.' Meta. 3 December 2024. <https://about.fb.com/news/2024/12/2024-global-elections-meta-platforms>
- 36 Microsoft. *Microsoft Digital Defense Report 2024*. Section: AI Threats and Ecosystem-Level Risks. (2024) p.88 www.microsoft.com/en-us/security/security-insider/intelligence-reports/microsoft-digital-defense-report-2024
- 37 Ibid. Section: Recommendations on Limiting Foreign Influence Operations. p. 93 www.microsoft.com/en-us/security/security-insider/intelligence-reports/microsoft-digital-defense-report-2024
- 38 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media* (2018) pp.221–233
- 39 US Communications Decency Act of 1996. 47 U.S.C. § 230. Protection for private blocking and screening of offensive material www.law.cornell.edu/uscode/text/47/230
- 40 US. Digital Millennium Copyright Act of 1998, Pub. L. No. 105–304, 112 Stat. 2860 (1998), codified in Title 17 of the US Code. The DMCA implements international copyright treaties, criminalises circumvention of digital rights management technologies, and establishes a 'safe harbour' shielding online service providers from liability for infringing content uploaded by users. www.copyright.gov/legislation/dmca.pdf
- 41 European Union. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act). (2022) p. 1–102. <http://data.europa.eu/eli/reg/2022/2065/oj>
- 42 Ibid. See in particular articles 33–44 on systemic risk management for 'very large online platforms' and 'very large online search engines', articles 14–23 on transparency and accountability, article 36a on the integration of the Code of Practice on Disinformation (February 2025), article 16 on user reporting and redress mechanisms, and article 36 on the crisis response mechanism. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>
- 43 European Commission. Code of Practice on Disinformation (2018) and Strengthened Code of Practice on Disinformation (2022) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>
- 44 Schiffrin, A. 'How Europe fights fake news.' *Columbia Journalist Review*. 26 October 2017. www.cjr.org/watchdog/europe-fights-fake-news-facebook-twitter-google.php
- 45 International Panel on the Information Environment. *Trends in the Global Information Environment: 2023 Expert Survey Results* (2023) www.ipie.info/research/sr2023-3
- 46 Schiffrin, A. 'How Europe fights fake news.' *Columbia Journalist Review*. 26 October 2017. www.cjr.org/watchdog/europe-fights-fake-news-facebook-twitter-google.php; Schiffrin, A. *The Pursuit of Truth: Fixes for the Spread of Online Mis/Disinformation* (2023) https://igp.sipa.columbia.edu/sites/igp/files/2023-12/IGP_Anya_Schiffrin_The_Pursuit_of_Truth-Fixes_for_the_Spread_of_Online_Mis_Disinformation.pdf
- 47 UNESCO, *Journalism, "Fake News" & Disinformation: Handbook for Journalism Education and Training*, (2021); OECD. *Facts Not Fakes: Tackling Disinformation, Strengthening Information Integrity*. (2022) <https://doi.org/10.1787/5b7e3c1c-en>. European Commission, *Report on the Implementation of the 2022 Strengthened Code of Practice on Disinformation* (2023).
- 48 Brezzi, M. Gonzalez, S., Nguyen, D. et al. An updated OECD framework on drivers of public trust in public institutions to meet current and future challenges, OECD Working Papers on Public Governance No. 48 (2021) p.9 www.oecd.org/content/dam/oecd/en/publications/reports/2021/12/an-updated-oecd-framework-on-drivers-of-trust-in-public-institutions-to-meet-current-and-future-challenges_bfa20b1b/b6c5478c-en.pdf
- 49 As an OECD study on COVID-19 measures found, digital prompts reduced people's intent to share false headlines by 21% compared to a control group, especially among frequent online users. OECD. *Misinformation and Disinformation: An International Effort Using Behavioural Science to Tackle the Spread of Misinformation*. Policy Paper No. 21 (2022) www.oecd.org/en/publications/an-international-effort-using-behavioural-science-to-tackle-the-spread-of-misinformation_b7709d4f-en.html See also MIT Initiative on the Digital Economy. *Reducing Misinformation Sharing with Accuracy Prompts*. Research Brief. (2024), reporting on field experiments showing that content-neutral accuracy prompts reduce the sharing of misinformation. https://ide.mit.edu/wp-content/uploads/2024/04/RB_3-31-24.pdf
- 50 Ibid (OECD, 2022)





Chapter 6

Rooted in resilience: Community-first approach to harmful information





Chapter 6



Rooted in resilience: Community-first approach to harmful information

Contents

	Introduction: The importance of community for resilience	213
6.1	Determining community	213
6.2	Community engagement and accountability	214
6.3	What communities say: Primary research insights	225
6.4	Community-led solutions	228
6.5	Unpacking risk: A gender and diversity lens	232
6.6	Building sustainable information resilience	234
6.7	The community within: How volunteers build trust and humanitarian reach	238
6.8	Effective communication in the context of harmful information	242
6.9	Critical reflections: Challenges in practice	247
	Concluding remarks: From communication to collaboration	248
	Endnotes	252

Introduction: The importance of community for resilience

In an era marked by record-breaking climate extremes, a growing number of disasters and emergencies and shrinking humanitarian budgets, the imperative to act to prepare before crises strike has never been more critical. But today's emergencies are not only physical – they are also informational. Harmful information can amplify fear, erode trust and disrupt preparedness and response efforts. It exacerbates existing vulnerabilities and tensions, compounding risks in contexts where trust is already fragile. As with disaster risk reduction, addressing harmful information requires a fundamental shift from reactive response to proactive resilience. This means supporting locally led action, investing ahead of crises and tackling the root causes of risk at the heart of humanitarian strategies.¹

Just as physical disasters need not be deadly or destabilizing if communities are supported to anticipate, withstand and adapt, the same holds for harmful information. The challenge lies in scaling and sustaining locally grounded solutions by embedding information resilience into humanitarian programming. Strengthening resilience through integrated, cross-sectoral approaches shifts the focus from reacting to harmful information crises to managing risks and building communities able to withstand them.

This chapter shifts the focus of previous chapters from identifying the threats posed by harmful information to exploring practical, community-driven solutions that strengthen local capacity and resilience. It also draws on primary research with community members, recognizing that top-down approaches – which often emphasize the roles of governments, organizations or technology companies – may overlook dynamics that critically shape the effectiveness of responses in community contexts.²

Communities are inherently dynamic and complex, so too are the vulnerabilities they face. Multiple, interconnected factors – physical, human, financial, natural, social and informational – shape a community's resilience. In today's increasingly digital and connected world, the ability to access, interpret, trust and act on information is critical to how communities prepare for, respond to and recover from crises. Understanding and strengthening resilience therefore requires a holistic multidisciplinary approach that considers how these factors interact and evolve over time. One aspect of this is how harmful or helpful information flows through and influences community systems. Community-level organizations are uniquely positioned to understand both the needs and specific contexts in which harmful information takes root and spreads. They are often best placed to respond to localized challenges, grounded in a clear understanding of community dynamics, relationships and priorities.

6.1 Determining community

A community is a group of people who may or may not live in the same area, village, neighbourhood or region but who share a common culture, habits, resources or social

connections. Communities are also shaped by shared exposure to the same threats and risks, such as disease, political and economic issues and disasters.³

The term 'community' is diverse and depends on context. It often refers to:

- a group of people living in a defined geographic area
- a group of people sharing common culture, values, norms or social structures
- a collective defined by shared interests or identities, whether local, national or international.

People frequently belong to multiple communities and research shows that the more communities an individual is connected to, the more resilient they are likely to be.⁴

In the digital age, the meaning of community has expanded. Communities are no longer defined solely by geography; technology connects people locally and transnationally, allowing them to gather around shared interests, values or identities. Internet-age recognition – whether of identity, expression or experience – does not just amplify individuals or ideas, it also draws people into contact with others who think and act as they do. Thus, the term 'community' increasingly denotes a group with shared interests and identities that make them distinct from the wider world. Where communities were once primarily place based, many are now interest or identity based, formed online and transcending physical boundaries, including for those who find a sense of belonging in exclusionary or harmful identities.⁵

The evolving notion of community highlights the importance of peer-to-peer relationships and community building in responding to harmful information. Strengthening these connections is central to building information resilience and promoting locally led responses.

6.2 Community engagement and accountability

Local communities are on the frontlines of the risks posed by harmful information and are essential sources of understanding, local knowledge and effective mechanisms to reduce those risks. Local actors are often the most trusted voices. Trust and proximity are critical for developing contextual understanding, leveraging localized knowledge and countering harmful information narratives. As first responders themselves in emergencies, communities are also the first to be affected by rumours, panic and harmful information.

The Red Cross Red Crescent *Guide to Community Engagement and Accountability* (CEA)⁶ outlines five key objectives to engage communities across all stages of programming:

- 1 To understand the community context and needs.

- 2 To deliver better, more effective programmes and operations.
- 3 To build trust, access and acceptance with communities.
- 4 To strengthen community ownership and resilience.
- 5 To uphold commitments and accountability to communities.

These objectives are particularly relevant in the context of harmful information, which often flourishes in environments marked by low trust, unmet information needs or social exclusion. Engaging communities throughout the design, implementation and evaluation of interventions and programmes aimed at countering harmful information helps ensure that responses are relevant, culturally appropriate and sensitive to the crisis context and evolving information landscape. Such engagement also enables practitioners to identify information gaps, social tensions and power dynamics that may shape how messages are shared, interpreted or received. In some cases, the process of co-developing information responses can itself serve as a connector – rebuilding trust, fostering dialogue and creating space for local problem-solving. Whether addressing harmful information or broader challenges to resilience, CEA remains central to effective and principled humanitarian action.

In the face of harmful information, community resilience is key to anticipating, absorbing and adapting to its impacts. The IFRC defines resilience as: “The ability of individuals, communities, organizations or countries exposed to disasters, crises and underlying vulnerabilities to anticipate, prepare for, reduce the impact of, cope with and recover from the effects of shocks and stresses without compromising their long-term prospects.”⁷

Contributor Insight 6.1

Highlighting community engagement and accountability in the Americas

In Central America, the institutionalization of community engagement and accountability (CEA) has been key to guiding actions that place community voices at the centre of humanitarian response. This has been reflected in the development of audiovisual case studies – tools that not only document good practices but also highlight the experiences, concerns and proposals of affected people.

In 2022, the Honduran Red Cross and Guatemalan Red Cross produced participatory videos as part of their responses to Hurricanes Eta and Iota, as well as to COVID-19. In these videos, communities share their challenges and recovery processes in their own words, illustrating how active participation strengthens the relevance and effectiveness of the response.⁸

In 2023, under the [Programmatic Partnership](#), actions were documented through an audiovisual case study. In one of these videos, Red Cross volunteers and staff explain why feedback mechanisms are essential in emergencies and humanitarian programmes.

Feedback enables communities to share their opinions and is a cornerstone of the CEA approach within the IFRC network, ensuring that local voices are heard and integrated into decision-making.⁹

In 2024, through the Building Trust project, inter-institutional coordination was strengthened through active listening and community analysis. This strategy demonstrates how institutionalizing CEA not only improves communication but also transforms the way the IFRC designs and implements actions.¹⁰

Impact of CEA in public health emergencies – Guatemala (2023)

This **case study**,¹¹ part of a global research initiative commissioned by IFRC, analyses the impact of the CEA approach in the context of public health emergencies. The primary objective is to identify, understand and document how the implementation of CEA has influenced Red Cross programmes and community health systems in five countries: Indonesia, Guatemala, Guinea, Georgia and Malawi. In Guatemala, the 2023 study drew on lessons learned during the COVID-19 pandemic and its recovery phase.

The findings highlight that evidence-based actions, both qualitative and quantitative, are essential to guiding interventions that are more contextualized, relevant and sustainable. Combining statistical data with community narratives provides a more comprehensive understanding of the effects of humanitarian actions and the factors that influence their effectiveness. The use of structured feedback mechanisms, perception analysis and community-level information monitoring has significantly strengthened trust, programme relevance and public health outcomes.

Furthermore, effective impact measurement requires the strengthening of local capacities for data collection, analysis and ethical use. This underscores the need for sustained investment in technical training, methodological support and the recognition of community knowledge as a legitimate source for decision-making. Finally, the institutionalization of CEA, combined with its continuous evaluation, has demonstrated its potential to transform both humanitarian response and the relationship between communities and health systems.

Influencing public policy through CEA – Panama (2025)

This **case study**¹² analyses how the Panamanian Red Cross is leveraging its auxiliary role to the state and its active participation in governmental coordination spaces to influence public policy, promote more participatory and accountable disaster responses and strengthen its position as a key humanitarian actor in the country.

The National Society has actively promoted the CEA approach. Working in partnership with institutions such as the Ministry of Health and the Ombudsman's Office, it has jointly established mechanisms to channel community feedback to the relevant authorities, contributing to improved quality and relevance of responses – as demonstrated during the 2024 dengue fever outbreak and flood emergencies.

The Panamanian Red Cross has also implemented targeted training initiatives (such as CEA in emergencies workshops) for its staff, government officials and inter-agency partners, as well as conducting knowledge, attitudes and practices surveys, and disseminating institutional policies.

The study highlights how this strategy has enhanced coordination, prevented duplication of aid efforts and increased community acceptance. It also underscores the importance of sustained investment in the institutionalization of CEA, including increasing geographic coverage, recruiting dedicated technical personnel and integrating CEA across all phases of programmes and operations.

Strengthening social cohesion through community-led evidence – Ecuador (2023)

This study promotes social cohesion between migrants and host communities through strategies grounded in community perspectives. By collecting direct insights from both groups, it avoids dependence on harmful, unverified narratives and supports an evidence-based, community-driven approach to humanitarian programming.¹³

Conducted by the IFRC and the Ecuadorian Red Cross in five neighbourhoods with high levels of interaction between local and migrant populations, the study gathered 837 responses – 61% from host communities and 39% from migrants. Findings show that motivations for migration, length of stay in the country and individual expectations strongly influence perceptions of belonging and coexistence. Longer stays often increase migrants' intention to settle, creating opportunities for more sustainable and inclusive programming.

Key results reveal a shared interest in entrepreneurship and community-based economic activities, especially among women. However, 57% of migrants reported experiencing differential treatment based on nationality, particularly in public spaces. Female respondents reported more discrimination within neighbourhoods, while male respondents more frequently cited discrimination on public transport. These patterns highlight the need for localized, gender-sensitive and time-specific inclusion strategies.

While mutual cultural curiosity exists, gaps remain. Both groups expressed a willingness to engage more deeply through festivals, educational initiatives and shared experiences. Community-generated recommendations emphasize the importance of face-to-face interaction over digital-only engagement, showing a strong preference for direct, human-centred dialogue.

This research supports the development of programmes that are technically sound, ethically informed and socially rooted. It fosters trust by reinforcing the safe and responsible use of data and encourages the use of IFRC-approved artificial intelligence (AI) tools within a clear ethical framework. In doing so, it strengthens humanitarian response through inclusive, context-aware decision-making that reflects the voices of the people most affected.

Andrés Caro
Social Data Science
Specialist
**Central America
Cluster, IFRC**

Carla Guananga
Community Engagement and
Accountability Senior Officer
**Central America
Cluster, IFRC**

Carolina Cortés
Community Engagement
and Accountability Officer
**Central America
Cluster, IFRC**

Gracia Banegas
Graphic Designer
**Central America
Cluster, IFRC**

6.2.1

Listening first: Understanding information ecosystems and enabling local capacity

Understanding how communities navigate information requires mapping both formal and informal communication flows. While formal channels – such as public service

announcements, media outlets and official emergency alerts – play an important role, informal networks often have greater influence, especially in contexts where trust in institutions is low.

Family members, religious leaders, community elders and peer groups frequently serve as primary sources of information, shaping how messages are received, interpreted and acted on. These interpretations are further shaped by social norms, cultural beliefs and collective memory, which all filter whether information is trusted, questioned or rejected. Harmful information often exploits these dynamics, filling gaps where trusted, timely or locally relevant information is lacking or mistrusted. When official communication fails to resonate with or reach intended audiences, harmful information can become more persuasive, not because it is more accurate, but because it feels more familiar, accessible or aligned with lived experiences and perceptions.

Contributor Insight 6.2



Yemen Red Crescent Society's response to harmful information: Strengthening community trust

The proliferation of disinformation and misinformation poses a significant and multi-faceted threat to the Yemen Red Crescent Society, critically hindering its humanitarian operations and endangering the safety of its volunteers. False information reaches the National Society through social media, field reports, direct community interactions and its hotline, and manifests in several harmful ways. Examples include:

1 Challenges in humanitarian access and emblem misconceptions

A persistent challenge has been the misinterpretation of emblems, particularly the Red Cross emblem. In some communities, the presence of this emblem on aid items and banners alongside the Yemen Red Crescent Society logo is misconstrued as a sign of religious affiliation among partners. This misunderstanding has negatively affected local interactions, reducing cooperation from community leaders and generating some opposition to National Society initiatives.

The resulting misconceptions have created significant access barriers, directly impacting the Yemen Red Crescent Society's ability to deliver aid and implement programmes effectively. To address this, it has prioritized humanitarian diplomacy by strengthening engagement with community decision-makers and influential figures. Through dialogue, the National Society has clarified the emblem's universal humanitarian meaning and the neutral, independent nature of its partnerships. It has held numerous meetings to present its mandate, explain the purpose of its work and actively correct misconceptions. In addition, it is conducting focused operational communication sessions with community influencers. These sessions emphasize the humanitarian principles of neutrality, impartiality and independence. By proactively engaging with local leaders, the National Society

aims to build trust, ensure its humanitarian mission is clearly understood and accepted, and strengthen community acceptance and support.

2 Building community trust through proactive communication

Proactive transparency and communication: At the heart of this success lies the principle of radical transparency. During all aid distributions – whether for floods, other disasters or cash assistance programmes – the Yemen Red Crescent Society prominently displays banners with clear and detailed eligibility criteria. These banners specify the exact conditions for receiving aid, along with the items and quantities being distributed. This simple yet powerful measure leaves no room for ambiguity and directly counters harmful rumours that certain individuals or groups are being unfairly excluded. By shifting the conversation from speculation to shared understanding of the distribution process, it ensures communities know how and why decisions are made, reinforcing trust in the distribution process.

Empowering communities through accountability: Beyond transparency, the Yemen Red Crescent Society has strengthened accountability by providing community members with direct channels for feedback and complaints. Dedicated complaint booths and hotlines at distribution sites encourage both beneficiaries and non-beneficiaries to voice their concerns. What could otherwise be a point of friction is transformed into an opportunity for direct engagement: trained staff can immediately address issues, clarify targeting criteria and explain the humanitarian principles guiding the National Society's work. This proactive feedback loop not only resolves concerns but also demonstrates its commitment to fairness and accountability, reinforcing its role as a trusted partner.

Over time, this strategy has both addressed immediate issues and fundamentally changed the perception of the Yemen Red Crescent Society's work in the most vulnerable communities. By turning a source of damaging rumours into a story of trust and community empowerment, the organization has solidified its reputation as an accountable and reliable humanitarian actor.

Rania Alshabibi

Planning, Monitoring, Evaluation and Reporting Manager

Yemen Red Crescent Society

Contributor Insight 6.3



Q&A with the Red Crescent Society of Kyrgyzstan

1 What examples from your work illustrate how harmful information has affected humanitarian response, access or community trust?

Around one-quarter (23%) of the population in Kyrgyzstan was vaccinated with at least one dose of a COVID-19 vaccine as of December 2023 (WHO), one of the lowest coverage rates in the region despite availability of vaccines. Vaccine hesitancy during the COVID-19

pandemic translated into a broader distrust in vaccines afterwards, including routine childhood immunization. As a result, vaccine refusals increased markedly following the COVID-19 pandemic – a trend seen globally, but particularly pronounced in Kyrgyzstan.

In 2023, with IFRC support, the Red Crescent Society of Kyrgyzstan conducted a perception study on routine immunization, which provided deeper insights into people's perceptions towards vaccination.¹⁴ Anti-vaccination sentiment continues to grow in the country. The Ministry of Health of Kyrgyzstan maintains a database of vaccine refusals – where parents or guardians of eligible children officially register their decision through a refusal form. According to the Republican Center for Immunoprophylaxis (the Ministry of Health's main technical agency for immunization), newly registered refusals rose sharply after the pandemic from over 10,000 in 2021, 20,496 in 2022, 20,486 in 2023 and 19,760 in 2024.

In 2024, 93% of vaccine-eligible children received the second dose at the age of two (WHO Immunisation Data portal – Global). While this represents strong coverage, it remains below the recommended 95% threshold needed to achieve herd immunity. As a result, the country experienced a surge in measles cases in 2023, 2024 and 2025. In 2023, there were 7,046 confirmed cases; in 2024, the number rose sharply to 24,380 cases; and in 2025, more than 9,000 cases were reported.

2 How do you define or recognize harmful information in your context and who do you see as most affected by it?

In the case of harmful information about childhood vaccinations, it is ultimately children who bear the consequences of their parents and caregivers being influenced by harmful information on vaccination. For example, measles and rubella vaccinations in Kyrgyzstan are administered at the ages of one and two years, when children are too young to have any voice in the decision. The human cost is stark: in Kyrgyzstan, nine children died from measles-related complications in 2023, five in 2024 and ten in 2025.

3 What strategies or tools have you used or seen used effectively to respond to harmful information during disasters or crises?

Strategies used by the National Society to address vaccine hesitancy in programmes promoting COVID-19 vaccination and childhood vaccinations include:

- a **Engaging religious leaders and community leaders:** In Kyrgyzstan, vaccine hesitancy is often driven by false beliefs, for example, that vaccines made in certain countries are less safe or effective or that they may not meet Halal requirements. Leveraging its established relationships with local leaders and trusted figures, the Red Crescent Society of Kyrgyzstan is well-positioned to manage rumours and counter misinformation – an essential part of building trust and improving health outcomes. This approach has proven effective. Between April and September 2025, trained Red Crescent volunteers worked with 337 religious leaders, equipping them to share accurate information and encourage immunization among their communities. These leaders continue to explain the importance of vaccination from a religious perspective, using messages that combine evidence-based health information with religious interpretations.
- b **Listening to communities and adapting communication strategies:** the Red Crescent Society of Kyrgyzstan has well-established community feedback channels, including hotlines, community meetings and digital platforms,

ensuring that community concerns are regularly collected, analysed and acted on. Staff are trained in active listening and respectful communication, allowing for effective feedback management. Based on community input, it adapts its key messages to better respond to concerns.

- c Reaching out individually with vaccine-hesitant parents:** Despite the Ministry of Health's efforts, there remains a group of parents hesitant about vaccination, leaving children zero-dose or under-immunized. Building on its earlier work (reaching more than 2,000 families with zero-dose children in 2023–2024), the National Society has found that one-to-one engagement is often required. Trained volunteers visit households – working alongside local health workers – to address fears and hesitations through active listening, emotional support and accurate, tailored information.
- d Engaging social media and mass media:** In Kyrgyzstan, harmful information spreads largely through social media. To counter this, since 2023 the National Society partners with bloggers and influencers who share accurate information and positive vaccination experiences with their followers. At the same time, it supports parents and caregivers who actively promote childhood vaccination on social media and other public forums. To reach wider audiences, it also works with mass media agencies creating content and disseminating messages through radio broadcasts and television segments.

4 **How does harmful information influence relationships with local communities, volunteers or authorities in your setting?**

Based on the Red Crescent Society of Kyrgyzstan's experience promoting COVID-19 vaccination during the pandemic, it was observed that even when volunteers successfully encouraged people to seek vaccination at local health clinics, some health practitioners themselves discouraged people from receiving the COVID-19 vaccine. A similar pattern was observed during responses to a surge of cases of measles, where certain health practitioners turned away parents and children, citing potential side effects. This highlights that efforts to counter harmful information must be fully supported by health services and that strengthening public trust in health practitioners and the health system remains essential.

Oyungerel Amгаа

Health and Care Manager for Central Asia

IFRC



Disinformation fuels attacks on humanitarian response during Ecuador's 2019 protests

For many years, Ecuador has faced an environment of constant social mobilizations that have shaped the humanitarian action of the Ecuadorian Red Cross. These complex scenarios have influenced its institutional response, transformed communication dynamics and redefined its presence on social media and other communication channels.

For example, a 2019 presidential decree eliminating the gasoline subsidy triggered strong social unrest in Ecuador and resulted in social mobilizations that lasted for 11 days. The level of violence intensified as major roads were blocked, cities became isolated, and shortages of food, medicine and other essential goods were reported. In response, the Ecuadorian Red Cross activated its response teams, operating in line with pre-established emergency plans.

Amid the crisis, disinformation spread rapidly on social media following the release of a video showing an ambulance in Quito's historic centre distributing equipment intended for police officers, such as bulletproof vests and tear gas canisters. The post quickly went viral and had serious consequences for the Red Cross, though it did not involve a Red Cross vehicle. The misleading and malicious narrative led to emergency vehicles being stopped and inspected to verify whether they were transporting patients or supplies. There were difficulties in transporting blood components, delays in emergency care, tyres punctured by sharp objects, and rocks thrown at four ambulances clearly marked with the Red Cross emblem even while using sirens and flashing lights to reach injured individuals. The spread of false information generated mistrust, uncertainty and the word-of-mouth repetition at protest sites further weakened the credibility of first response organizations, undermining humanitarian action and compromising life-saving operations at a critical time.

Following the events of October 2019 and in the lead-up to the protest movements of June 2022, humanitarian diplomacy with all involved organizations became a key tool to counter disinformation that could directly impact the work of the Ecuadorian Red Cross. The dissemination of key messages through social media was a vital part of this strategy. Implementation took place at all levels under the leadership of the National Society President and Secretary General and in coordination with the management and operations teams throughout the country. The main objectives were to reinforce that Red Cross action is based on the principles of neutrality and impartiality that underpin its humanitarian work, and to prevent harmful information from undermining this work.

Complementary actions were carried out, including:

- 1 Strengthening the institutional image through the proper use of uniforms by all personnel.
- 2 Disseminating informational content about the Ecuadorian Red Cross ambulances to facilitate their identification and distinguish them from the vehicles used by other institutions.

- 3 Promoting the Fundamental Principles of the International Red Cross and Red Crescent Movement as pillars of humanitarian action.
- 4 Implementing the campaign 'We Are Not a Target,' aimed at raising awareness about the neutral role of ambulances in emergency response.
- 5 Disseminating key messages in both traditional and alternative media to highlight the role of the Red Cross.

These actions are part of a strategy aimed at raising public awareness about the humanitarian role of the Ecuadorian Red Cross in the context of social mobilizations. The initiative is reinforced through the use of accessible language – including Kichwa, as the predominant language in Indigenous communities – with the purpose of strengthening understanding and acceptance of Red Cross interventions.

Roque Soria

National President

Ecuadorian Red Cross

Juan Carlos Vizcarra

Secretary General

Ecuadorian Red Cross

A community-centred, trust-building approach played a pivotal role in transforming a highly volatile situation into a collaborative response effort during the 2018–2020 Ebola outbreak in eastern Democratic Republic of the Congo (DRC).¹⁵ Malteser International, active in the region since 2000, piloted the People First Impact Method (P-FIM) – a structured, participatory approach focused on listening first, then co-creating solutions with communities. Through a two-step process of open dialogue and collaborative planning, Malteser International staff worked directly with community members to:

- identify fears and barriers to trust
- address circulating rumours
- co-design communication activities based on local capacities and knowledge.

This locally anchored engagement helped to reduce tensions, rebuild trust and counter harmful information that had previously undermined the Ebola response.

The approach highlights a critical lesson: effective responses to health emergencies and harmful information require early, meaningful community involvement. In fragile and conflict-affected contexts, humanitarian actors must refrain from presenting themselves as experts with all-encompassing knowledge and all the answers. Instead, they should co-create responses with affected populations. Doing so not only counters harmful information, but also builds local ownership, strengthens community capacity and fosters motivation, trust and, ultimately, resilience.

Professor Pierre from the University of California refers to this mindset as 'intellectual humility' – the practice of acknowledging uncertainty, the possibility of being wrong and avoiding unwarranted overconfidence. For humanitarian organizations to embody intellectual humility means acknowledging both *what* they don't know and *that* they don't know¹⁶ while being open to learning from others – especially from the communities most affected.

Contributor Insight 6.5

Canadian Red Cross: Localizing humanitarian action through the Indigenous Peoples Framework

For three decades, the Canadian Red Cross has partnered with more than 500 Indigenous communities across a vast and diverse geography. This work is anchored in the Indigenous Peoples Framework and its four pillars – reconciliation, cultural safety, collaboration and community-led service delivery – which together provide a principled, organization-wide approach to localized, accountable humanitarian action with Indigenous communities.

The framework recognizes that trust is built differently with Indigenous Peoples. The Canadian Red Cross invests in direct, recurring touchpoints with local points of contact, Indigenous community leadership, elders and community-based advisories, and provides critical information in Indigenous languages. These modalities keep engagement grounded in community priorities and help close information gaps where harmful information can emerge along the pre-crisis to post-crisis continuum. Lessons from the COVID-19 pandemic reinforced this approach: as many Indigenous Nations across Canada exercised sovereignty by closing their borders to protect residents, the Canadian Red Cross maintained two-way communication through dedicated virtual operations capacity and secure, trusted channels.

Today's enduring climate risks heighten the need for this model. Although Indigenous Peoples comprise about 5% of Canada's population, they accounted for an estimated 42% of wildfire evacuations in 2023,¹⁷ and 16% of disaster-related displacements.¹⁸ In response, the National Society prioritizes adaptable, localized delivery through mobile teams, embedded liaisons, in-language communications, and evidence-informed preparedness and evacuation supports – ensuring needs are met before, during and after crises while strengthening community resilience.

The Indigenous Peoples Framework aligns with the UN Declaration on the Rights of Indigenous Peoples and is grounded in the principle of free, prior and informed consent. The Canadian Red Cross upholds Indigenous data sovereignty: information shared by communities remains their property and is stewarded according to community protocols. These commitments underpin transparency and position the National Society as a partner of choice across First Nations, Inuit, Métis, urban, remote and northern contexts. Through the framework, it continues to walk alongside Indigenous communities by building trust, reinforcing local capacities and ensuring humanitarian action remains people first, culturally safe and led by the priorities of the communities it serves.

Jean-Philippe Crete

Director of Research and Policy, Office of Indigenous Relations

Canadian Red Cross

6.3 What communities say: Primary research insights

Harnessing the contextual knowledge of volunteers of National Red Cross and Red Crescent Societies is important to understand how harmful information affects communities, how it interacts with trust, and how its impacts are mitigated. To capture this, the **IFRC Solferino Academy** conducted a 1.5-month rapid community intelligence study with 40 volunteers and staff from 10 National Societies, acting as community researchers.¹⁹ Between them, they interviewed 132 volunteers and community members – often people they support or live alongside – who have been affected by landslides, earthquakes, cyclones, floods, civil unrest, epidemics, pandemics, migration-related challenges, economic hardship and more.

Working with communities on a daily basis, the community researchers are uniquely positioned to identify challenges and highlight key considerations for response strategies. Some of the main findings include:

6.3.1 Harmful information affects decision-making and organizations can be hampered by perceptions of their role

In disasters and crises, the absence of accurate information makes it difficult for people to make informed, beneficial decisions. A community member from Madagascar described the impact:

“*Distrust of the products distributed, such as hygiene kits or purified water, often stems from rumours or false information. This leads some people to reject solutions that are beneficial to [their own] health.*”

Ensuring the availability of accurate information is therefore a priority for humanitarian organizations. Yet, even where trust exists, organizational reputations can hamper efforts to counter false information if information-sharing is not perceived as part of their role. A volunteer in Kenya explained:

“*[During an mpox outbreak, community members] perceived us as an organization that should come with goodies to give them ... they didn't take that sensitization seriously. The community ... felt it's better if [we] provided them with items that are [tangible] ... rather than passing down information.*”

The misalignment between organizational reputation, role and capacity can have profound effects. In one of the Qatar Red Crescent Society's international operations involving many organizations, rumours about aid availability damaged trust and created community rifts. Confusion over the roles of different actors, combined with urgent needs, exacerbated the problem. As one community member – a response worker – noted:



- “ *The false information and rumours ... greatly diminished our role as an organization. [Affected community members] started going to these [other] organizations asking where the aid was that had arrived [according to rumours], ‘Why didn’t we receive anything?’ ... people during that period did not see or analyse what this organization was for.* ”

Another response worker described how rumours fuelled inter-community tensions:

- “ *The issue of the mixed news ... started to create a rift ... and they started to give it interpretations based on ethnicity and sectarianism.* ”

6.3.2

“The very ties that bind communities together can make them more susceptible”; ties spread information, for better or worse

Harmful information often spreads most powerfully through trusted networks. A community researcher from Zambia explained:

- “ *Trust within communities is a form of trust that is resilient ... built on shared cultural values, norms and understanding. At the same time, false information is particularly effective when it originates within the circle of people you trust ... the very ties that bind communities together can make them susceptible to misinformation.* ”

At the same time, false or misleading information may fracture community trust and disrupt social balance. As another Community Researcher from Zambia reflected:

- “ *It is evident that harmful information really disturbs the balance within communities.* ”

A community researcher from Cameroon noted the complexity of these dynamics:

- “ *Trust within communities is fluctuating. ... Nevertheless, close ties remain important landmarks, even if the rapid circulation of word-of-mouth [information] can strengthen solidarity as well as sow mistrust.* ”

Social media accelerates these dynamics, rapidly amplifying both accurate and misleading information. A community member (rescue worker) in China pointed to a new challenge:



- “ *Nowadays, with the rise of short videos, some bloggers use AI to create composite images ... you can describe a disaster to an AI, and it will generate images — but these images don’t match the reality. This kind of exaggerated and false information is actually quite common.* ”

Harm affects humanitarians themselves. During the COVID-19 pandemic, misrepresentation of caregivers' roles eroded essential trust between them and the communities they served. A community member in Madagascar recalled:

“*The media landscape ... presented a simplified and often false image: either a ‘hero’ caregiver without weaknesses, or a caregiver ‘on the verge of collapse.’ Information, when exaggerated or minimized, distorts the social perception of the profession ... This lack of nuance has fuelled stereotypes and exacerbated caregivers’ difficulty expressing their emotions. Misrepresentations undermined trust between caregivers and the public.*”

6.3.3

“Trust alone is not enough”: culture and structure play key roles

Even when trust exists, other factors can prevent accurate information from reaching communities effectively. In Sri Lanka, a community researcher explained why systems matter:

“*While trust exists, how this information is communicated is questionable. The Red Cross and government often lack a structured, efficient communication system ... Trust alone is not enough without a proper system for fast, verified communication, even trusted organizations risk failing their communities during disasters.*”

Structural inefficiencies can also cause direct harm. Another community member observed:

“*Slow or delayed information hurts the efficiency of rescue efforts ... When red tape and administrative hurdles prevent timely help from getting through, it wastes goodwill and [may] do more damage than outright false information.*”

Culture is equally critical. A community researcher in Uganda explained:

“*Culture plays a big role in shaping people’s ways of thinking and perception about information being passed to the community. People verify information through their leaders.*”

In Madagascar, a community researcher offered a striking example of how cultural codes shape interpretation:

“*In some parts of Madagascar, the arrival of two hatless men in a village traditionally signalled a death, illustrating the importance of cultural codes in communication. Understanding and respecting these practices can improve the effectiveness of awareness campaigns.*”

Community reasoning is also grounded in experience. A community researcher in Bolivia emphasized:

“ *They had the same information from other doctors ... but they always believed in what people close to them said... their [reasoning was]: if COVID was similar to a cold and in the past herbs cured them, the same could be used with COVID ... we are not irrational, we continue to base ourselves on facts and experiences.*

Identifying trusted groups within a culture is therefore essential. The research highlighted elders, traditional healers, religious leaders and women as particularly influential. In Zambia, a community member described the influence of women:

“ *Women are one of the most influential individuals to advocate for health within communities and even in families... [they are] powerful in the face of public health.*

In Bolivia, a community member highlighted the decisive role of mothers:

“ *... mothers played a very important role ... the information they had was not even questioned, it was believed and complied with ... the mothers and the trust they had towards them, made the difference.*

At the same time, the availability of digital tools is reshaping and sometimes challenging traditional structures. A Zambian community member observed:

“ *The internet has revived and amplified ancient superstitions, spreading them rapidly and giving them new life.*”

This is echoed by a community researcher from Madagascar who emphasized the intergenerational challenge:

“ *Traditionally, local chiefs are focal points for information spread and verification, now youth start spreading information quickly through social media without going through the traditional verification channels.*

Taken together, these insights highlight that building resilience against harmful information requires more than trust. Effective communication depends on the interaction of trust with cultural understanding and structural efficiency, ensuring that communities not only receive information but are also able to use it meaningfully.

6.4 Community-led solutions

Addressing harmful information is most effective when messages originate from within communities rather than being externally imposed. In many settings, participatory content creation has proven more relevant and trusted, as it reflects local language, cultural references and lived experiences. Communities often make sense of events and share knowledge through storytelling, radio, theatre or visual communication – formats that are not only familiar but also emotionally resonant.

Where messages are co-created, they tend to carry greater legitimacy and reach. In contrast, efforts that focus solely on ‘correcting’ falsehoods – without acknowledging and addressing the underlying concerns or mistrust that make them persuasive – risk being ignored or even rejected. Observations across multiple contexts show that meaningful engagement, rather than top-down correction, is more likely to foster trust and shift harmful narratives.

Grand Challenges Canada frames the creation or strengthening of community networks as “initiatives focused on creating or enhancing spaces for dialogue”²⁰ – bringing together community experts and organizing policy dialogues with, for example, local media, influencers and researchers on harmful information. These efforts aim not only to prevent and counter harmful information but also to build shared understanding and community-led solutions rooted in trust and accountability. This includes collaborating on information resilience, responsible journalism, community-based protection, human rights, prevention of hate and genocide, and peacebuilding.²¹ The efforts aim to avoid or mitigate the humanitarian consequences deriving from harmful information and address its implications on trust and integrity in humanitarian action – ensuring that trust, access and principled engagement are preserved even in contested or polarized environments.

By embedding harmful information responses within existing social and community networks, these initiatives help communities recognize, respond to and recover from information threats, going beyond isolated messaging campaigns. They involve long-term partnerships, support community-based protection mechanisms and promote responsible journalism and inclusive policy dialogues focused on human rights, hate speech prevention, peacebuilding and information resilience, among other areas.

Effectively addressing harmful information requires investing in and equipping community members who are already trusted. Depending on the context and issue – and whether online or offline – these trusted messengers may include religious leaders, community elders, women’s groups or youth networks, as well as digital influencers who shape narratives and reach wide audiences via online communities. Local influencers bring credibility by being grounded in lived local realities and close to people in need. Engaging with these actors not only enhances the reach and relevance of messaging but also reinforces trust, particularly where institutional credibility is weak or contested.

Alongside this, strengthening media and digital literacy at the grassroots level is essential as it empowers individuals to critically assess the information they encounter and share. Rather than imposing ready-made, top-down narratives, humanitarian organizations can provide practical tools on the principles and practices of humanitarian action (focusing on what it entails rather than who the organizations are). These approaches include feedback mechanisms, rumour-tracking systems and awareness-raising with local media. They enable communities to engage actively with information, challenge harmful content and contribute to more resilient and informed environments.

Contributor Insight 6.6

Community-first: Insights from engagement with communities

Case 1: Why community engagement in early warnings matters

A recent study²² in Malawi's Nsanje and Phalombe districts showed community engagement is central to how climate change information is shared, trusted and acted on. Most people receive climate and early-warning information through community meetings led by trusted local actors – primarily government officers, NGOs, the Malawi Red Cross Society, local leaders and extension workers. These actors use familiar methods such as megaphones and direct dialogue to foster understanding and prompt action. Formal early warnings – including flood or drought alerts – are disseminated through radio broadcasts, local government officials and NGOs.

However, many community-based structures, such as village civil protection committees, lack the training and resources to share warnings consistently and independently, creating reliance on external actors. In Phalombe, younger and more digitally connected populations also turn to social media and local news, reflecting an evolving landscape of trusted communication.

Importantly, effective communication builds on existing community networks – women's groups, youth clubs and local committees – that blend scientific forecasts with Indigenous knowledge. As one local organization staff member explained,



We use the Department of Climate Change and Meteorology Services for seasonal forecasts, but committees say things like 'the hippopotamus was walking from the river to the community,' meaning it smells a flood. So, the community relies more on Indigenous information than on scientific forecasts."

Local organization staff member, aged 29, Nsanje

This community-rooted approach strengthens the ability to detect and respond to harmful information, building resilience before, during and after crises. Grounding early warning systems in trusted, familiar community structures and communication methods helps humanitarian actors address misinformation and disinformation, ensuring timely and actionable information reaches those who need it most.

Case 2: A community-first approach to navigating harmful information during COVID-19

Harmful information thrives where trust is weak or communities' information needs go unmet. The IFRC's [commitment to CEA](#) – embedded in its policies and Strategy 2030 – recognizes that meaningful, inclusive community participation is essential for building mutual trust and sustainable resilience.

In an era shaped by misinformation and disinformation, trust-building and inclusive, actionable communication are vital. Recent research²³ from Georgia, Guatemala, Guinea, Indonesia and Malawi – where National Red Cross and Red Crescent Societies applied CEA

approaches during COVID-19 to counter harmful information through localized, trusted networks – highlights the following key findings:

- **Georgia:**²⁴ Leveraging local authorities, youth volunteers and trusted community spaces helped combat vaccine rumours. “I am not a passive participant ... I suggested to the Red Cross to organize a first aid course in my village and they did.” (Community leader.) **Key insight:** Trusted local actors are essential to counter misinformation and prevent further marginalization.
- **Guatemala:**²⁵ Cultural complexity and social divisions fuelled distrust and misinformation. Collaboration with Indigenous councils, schools and women’s leaders helped tailor messages to local realities. **Key insight:** Community-led, culturally resonant communication counters disinformation that draws on lived experiences.
- **Guinea:**²⁶ Experience from past epidemics fostered trust in traditional leaders and peer educators who used storytelling and local theatre to challenge rumours. “Before, we didn’t dare, but thanks to persistent awareness and guidance, we dared to get vaccinated. We trusted the Red Cross volunteers because they are from the community.” (Community member.) **Key insight:** Past epidemic memory strengthens resilience against harmful information when combined with two-way communication.
- **Indonesia:**²⁷ Facing physical isolation and digital disinformation, collaboration with religious leaders and women’s groups enabled culturally adapted messaging through offline channels such as radio and community rituals. **Key insight:** While harmful information spreads digitally, trusted face-to-face communication remains a powerful corrective.
- **Malawi:**²⁸ Concurrent COVID-19 and cholera outbreaks fuelled fear, rumours and vaccine scepticism, including misinformation about infertility. Direct engagement with traditional leaders, health workers, youth and women’s groups – through household visits and dialogue – helped rebuild trust eroded by inconsistent messaging. Early delays in integrating community feedback limited initial responses to falsehoods. **Key insight:** Compounded crises amplify harmful information. Timely, inclusive and consistent messaging delivered by trusted peers is essential to restore confidence.

Case 3: Nepal:²⁹ **Rebuilding trust in vaccines through persistent, local engagement**

In Nepal, trust was not built through mass messaging but through repeated, personal interactions. In rural Banke district – hit hard by COVID-19 and vaccine misinformation – trust began to grow when locally known Red Cross volunteers made household visits and listened without judgement. Demonstrating the vaccine’s safety first-hand proved more persuasive than technical facts alone.

Communities need space to ask questions and see proof, not just follow instructions. Rumours about infertility and vaccine trials spread quickly, but Red Cross volunteers stayed embedded in the community addressing misinformation through door-to-door visits, radio messages and by modelling healthy behaviours. Having the same volunteer return repeatedly helped shift scepticism into confidence.

At the onset of the crisis, the lack of clear, trusted information left people vulnerable to fear and falsehoods. As the pandemic evolved, so did the community's need for more nuanced, timely information – especially during the vaccine rollout. Pre-existing mistrust made initial outreach difficult, but continued engagement helped move people from resistance to vaccine advocacy.

The Nepal Red Cross mobilized over 3,000 local volunteers, reaching more than 3 million people with culturally appropriate, localized communication. This **hyperlocal, face-to-face model** built lasting resilience – against COVID-19 misinformation and potentially for future public health crises as well. Empowered community members became messengers themselves, showing how trusted engagement creates ripple effects that strengthen community confidence.

Gefra Fulane

Research Coordinator

IFRC, Geneva

6.5 Unpacking risk: A gender and diversity lens

Applying a gender and diversity lens is essential to understanding the distinct roles, risks and needs of different population groups in the community in relation to harmful information. Such analysis helps identify specific vulnerabilities, how exposure to information risks differ and what tailored interventions are needed to ensure all community members – regardless of gender, age, ability, ethnicity or other identity factors – can access trustworthy information and participate safely and meaningfully in decision-making. By understanding how gender and diversity shape people's exposure to risk, including information-related risk, humanitarian actors can improve the safety, dignity and inclusion of affected populations. This approach also enhances the effectiveness of programmes and reduces the likelihood of exclusion, exploitation or abuse.

According to the International Telecommunication Union (ITU), 70% of men and 65% of women worldwide used the internet in 2024, meaning that approximately 189 million more men are online than women. While the gender digital divide is slowly narrowing, women still make up the majority of the global offline population. This gap is especially stark in least developed countries, where only 29% of women are online compared to 41% of men. Significant disparities also persist in Africa and the Arab States, where access for women continues to lag behind that of men. These digital divides have serious implications: unequal access to information and digital tools reinforces existing social and economic inequalities and limits women's opportunities for education, employment, civic participation and health. It also restricts women's ability to access reliable, life-saving information – especially in humanitarian crises. Addressing this divide is therefore not only a development issue but a matter of equity, inclusion and resilience.³⁰

At the same time, rapid technological change is giving rise to new and evolving forms of violence, particularly against women and girls. The UN Secretary-General's 2024 report³¹ underscores that perpetrators are increasingly using digital platforms and tools to carry out gender-based abuse, harassment and exert control, and misogynistic content

is spreading across mainstream platforms. This reinforces harmful masculinities and discriminatory social norms that perpetuate violence.

The rise of generative AI has further intensified these risks by enabling image-based abuse, deepfakes (see [Annex I: Glossary, on page 353](#)) and the rapid spread and amplification of misogynistic narratives. There is growing evidence that online violence is closely linked to real-world harms, including gender-related killings and femicides. Online abuse also has profound psychological, social and health impacts, often deterring people – especially women and girls – from participating in public life, education and employment. Many people self-censor, reduce their online presence or withdraw from digital spaces altogether. In doing so, they lose access to the very tools and platforms that support resilience.

Many responses to harmful information remain gender-blind, failing to consider the distinct ways in which harmful content targets individuals and how different groups experience and respond to it. Gendered patterns of harm include targeted abuse, threats of sexual violence and violations of digital privacy, particularly against women and girls. In many contexts, structural barriers further restrict access to information, limiting the ability to verify or challenge misleading content. In addition to gender-based risks, age-related vulnerabilities are often overlooked.³² Children and older adults may lack the digital literacy, access or support systems needed to critically assess information or respond safely to online threats. Addressing harmful information responses effectively requires inclusive strategies that recognize and respond to the diverse risks and needs of different population groups.

Contributor Insight 6.7

Safeguarding and harmful information

Safeguarding in the humanitarian sector aims to prevent and protect people from harm caused by humanitarian actors delivering responses, including harm from sexual exploitation, abuse and harassment. Harmful or inadequate information, whether deliberate or unintentional, can undermine safeguarding by silencing survivors, deterring reporting or spreading false accusations that damage trust in complaint mechanisms. Misinformation can distort perceptions of what behaviour is and is not acceptable, and who is safe to approach, while disinformation may be used to discredit whistleblowers or cover up abuse.

Increasingly, sexually explicit or illegal materials are used to intimidate, harass, extort or discredit individuals. Intimate images may be edited, misused or circulated without consent; false rumours can target staff or community members; fake profiles or forged documents can manipulate perceptions and distort aid delivery and distribution models, reinforcing inequality and exclusion; and AI tools can create harmful fake content. Such narratives can erode community confidence in humanitarian actors and prevent access to protection services, reducing accountability and the humanitarian 'licence to operate'.

Addressing this risk requires proactive communication: clear, consistent messaging on safeguarding commitments; transparent reporting processes; and staff, volunteer and community feedback mechanisms that are safe, accessible and trusted. By integrating

safeguarding into information and communication strategies, humanitarian organizations can both protect individuals and reinforce the trust that underpins effective, accountable humanitarian action.

Joanne Dunn

Head of Safeguarding

IFRC, Geneva

6.6 Building sustainable information resilience

Building resilience to harmful information requires more than short-term awareness campaigns: it demands sustained investment in local information infrastructure. In many contexts, trusted communication channels are fragile or underdeveloped, creating gaps that harmful content can easily exploit. Strengthening community radio, local media, feedback systems and digital access points ensures that reliable and relevant information is consistently available – not only during crises but before and after as well. As highlighted throughout this report, information is a critical need across the entire crisis continuum.

“

Well, as a primary source, yes, we relied a lot on the radio station since it is the main means of communication in my town. Even though nowadays there are new technologies and new ways to access information through social media, ... the radio station is still very relevant, and in fact, it's one of the media that has given us the most accurate information. Because while on social media they might say that such-and-such community is burning or that such-and-such town needs help, the only way to verify if it was true was through the radio station, since it did provide accurate information about what was happening at that time. ... the radio station remains active and is something truly essential, I believe, for the town and where people listen the most.”

Community member, Bolivia

The IFRC emphasizes that timely, trusted and accessible information enables communities to anticipate risks, make informed decisions and recover more effectively. Crucially, two-way communication allows communities not only to receive information, but also to voice concerns, challenge harmful information and help shape responses. Over time, this kind of meaningful engagement helps to reinforce broader forms of resilience. Communities that can navigate information critically and collectively are often better equipped to manage wider challenges because trust, agency and access to knowledge are foundational to effective response and recovery in any domain.

In Chad, the fight against infectious diseases has been strengthened through a community-driven radio initiative that prioritizes trust, participation and local relevance. Through the Programmatic Partnership, the Red Cross of Chad and French Red Cross

used Radiobox – a portable, easy-to-use broadcasting kit – to deliver accurate, life-saving health information to remote and underserved communities. Despite the rise of digital technologies, radio remains a vital tool in areas with limited access to health services. Radiobox allowed Red Cross volunteers to broadcast 140 live shows in local languages, focusing on epidemic-prone diseases, the importance of vaccination, and water, sanitation and hygiene practices. Crucially, the format was participatory: community members shaped the programme content, asked questions and voiced concerns around the health topics that mattered the most to them. This enabled volunteers to better understand and respond directly to the community's needs, as well as to dispel any harmful health information. Over 10,000 people participated in the broadcasts and an April 2025 evaluation found strong evidence of impact: 80% of participants surveyed felt the health messages were clear and 91% reported changing daily health behaviours. Radiobox's success demonstrates how trusted, inclusive communication channels can reinforce community preparedness, challenge harmful information and strengthen public health resilience.³³

Contributor Insight 6.8

Japanese Red Cross Society: Enhancing transparency of activities and donations during disasters

Japan's unique linguistic and media environment provides a partial buffer against the global spread of harmful information. Because Japanese is spoken almost exclusively inside the country, the language itself acts as a kind of firewall, limiting the impact of harmful information campaigns circulating in other languages. Japan's strong domestic media landscape – including public broadcasters – also helps shape a more contained information environment. This reduces dependence on foreign media sources and helps limit exposure to externally driven harmful narratives.

Social media usage patterns also differ in Japan. Platforms like LINE and X (formerly Twitter) are more commonly used than platforms like WhatsApp or Facebook. This creates a different digital ecosystem, where harmful information tends to be more localized and less influenced by international trends.

The Japanese Red Cross Society has consistently maintained a strong public communication strategy that contributes to trust and transparency. It offers information to donors, volunteers and members of the public primarily through its official website, social media accounts (X, Instagram, Facebook, YouTube), a monthly newspaper and other regular publications. Whenever a disaster occurs in Japan or abroad, it promptly communicates Japanese Red Cross Society disaster relief operations to donors and the public. In all donor communications, it directs people to clear explanatory resources such as its donation guide. As a result, instances of harmful information targeting the National Society have

been extremely rare in Japan. Trust is cultivated not only through emergency communication, but through consistent, transparent dialogue over time.

Akihisa Okayama

Director General of PR Office

Japanese Red Cross Society

Contributor Insight 6.9



Strengthening information integrity, towards a preventive and inclusive approach to disinformation

In the Sahel and Central Africa region, rumour and disinformation have a devastating impact. One example comes from the Central African Republic (CAR), where local populations shared how false news of an imminent conflict prompted several families to flee their homes, abandoning their possessions and deepening distrust among neighbours. Fondation Hirondelle has been working in the region for over 30 years. In our view, the main global risk linked to misinformation, malinformation and disinformation – which are increasing with digitalization and the rise of generative AI – is that of epistemic rupture. This concept refers to the breakdown of thought structures within a social group and it is unfolding along three fault lines:

- A growing segment of the population is being drawn into counterfactual narratives.
- The capacity for dialogue between those with opposing views is eroding, fuelling polarization.
- Less-connected, under-resourced populations – particularly those speaking marginalized languages – are being excluded from the production and sharing of knowledge.

To address this risk, it is not enough to merely counter disinformation as this approach is inherently reactive. Instead, there must be proactive and preventive investment in promoting information integrity.³⁴

1 Fill the information vacuum:

The first step in this preventive approach is to act before rumour and disinformation become the most accessible response to the insecurity caused by a lack of credible information. This is best achieved by strengthening the capacity of local journalists and media to provide trustworthy, timely and relevant information that is accessible and in local languages and available across both online and offline platforms. This approach precedes fact-checking, which aims to identify and react to disinformation after it has

already spread. It also differs from prebunking (see [Annex I: Glossary, on page 353](#)), which focuses narrowly on topics likely to attract falsehoods.

Promoting information integrity, by contrast, seeks to meet the broader information needs of communities. Scientific literature confirms that supporting local journalism is one of the most effective ways to prevent disinformation.³⁵ Research conducted in CAR, in partnership with US-based academic partners, also confirmed the effectiveness of a journalistic approach in building audience trust and helping people differentiate fact from fiction in contexts where disinformation is pervasive.³⁶

The key conditions for this preventive approach to have a positive impact are:

Grounding in journalistic ethics: Local journalists not only require proper training but also sufficient technical and financial resources to uphold professional standards. Their editorial independence must also be upheld. Blurring the line between the role of journalist and that of a spokesperson, activist or lobbyist promoting a particular point of view undermines public trust in journalism itself. Moreover, they need safety. Journalists are protected under international humanitarian law: as civilians, they must not be deliberately targeted and their equipment and installations must be respected.

Ensuring proximity and inclusivity: A journalist must be first and foremost a good listener – someone who actively seeks to hear and understand the diverse perspectives and information needs of their audience. This requires a local presence and proximity that enables journalists to connect with different communities, including the most marginalized, and to build trust. Proximity must be not only geographic, but linguistic: journalists must be able to communicate in local languages and in a manner that is accessible to their audience.

Accessibility also depends on the technologies used to share content. We recommend a hybrid (offline and online) and generalist (covering all types of news) approach to reach the widest possible audience. This involves producing content in various formats – audio, text and video – for multiple platforms, including TV, radio, newspapers, websites, social media and messaging apps, and in multiple languages. Content should be tailored to the needs of diverse population segments, especially the most vulnerable. This includes young people and women who are often among the first targets of discriminatory speech yet also play a critical role in spreading information within families and communities.

Digital technologies and AI can enhance the ability to listen at scale and improve the efficiency of content production. However, these tools come at a cost and require training and adaptation of professional practices. They also raise ethical concerns and must be used transparently. All AI-generated content must be reviewed by humans to avoid the spread of incomplete, unsourced or reductive information – or even disinformation. This is particularly critical in situations of armed conflict, humanitarian crisis and digital divide. AI systems operate primarily on digitalized data in majority languages; they are not neutral and often reflect cultural, linguistic and ideological biases. These technologies can support, but not replace, field presence, contextual expertise and human connection.

2 **Improve information and media literacy, through the media and with communities:**

Alongside efforts to improve the supply of information by enhancing the quality and accessibility of informative content, a second vital step in preventing the impacts of disinformation is to improve the demand side. This involves strengthening the ability of audiences to evaluate the quality of the content they encounter, choose to consume and share.

Information and media literacy through the media can be fostered in two key ways. First, through **greater transparency**, by explaining how journalists gather, process and prioritize information. Second, by developing **dedicated features** such as regular fact-checking segments and educational programming focused on subjects vulnerable to disinformation.

Such initiatives help alert audiences and raise awareness by using concrete examples of disinformation, while also offering ways in which they can guard against it. In parallel, and drawing on the example of Fondation Hironnelle's media work across the Sahel and Central Africa, physical spaces for discussion and dialogue can be established in neighbourhoods or villages to talk about the information circulating in a community. These forums – led by local journalists in partnership with community radio stations, local leaders or mediators – create opportunities to clarify rumours and share verified information. Additionally, interactive public programmes recorded live in communities, rather than in studios, allow opportunities for community members to learn how the media works.

Sacha Meuter

Head of Research and Policy

Fondation Hironnelle

6.7 The community within: How volunteers build trust and humanitarian reach

Volunteers are not outsiders who arrive to assist in times of crisis – they are the community itself. Drawn from the very neighbourhoods, villages and networks they serve, volunteers embody the lived realities, concerns and strengths of people in need. Their proximity builds trust, their presence ensures cultural and linguistic relevance, and their credibility anchors humanitarian action in local resilience. In many contexts, volunteers are the first to respond and the last to leave, making them both the frontline of assistance and a vital bridge between humanitarian organizations and the communities they aim to support.



Trust is built through continuous engagement with communities, from health promotion during cholera, measles and mpox outbreaks to supporting livelihoods when disasters such as fires occur. We listen to their needs, fostering acceptance that allows us to respond effectively in times of crisis, such as during this year and last year's [2024 and 2025] challenging anti-government demonstrations. Youth initiatives that nurture talent through sports and arts, alongside psychosocial support programmes, have also strengthened relationships and changed perspectives, further deepening trust and cooperation."

Philip Thuo Wachira, Volunteer, **Kenya Red Cross Society**

In an era marked by harmful information and growing public mistrust, humanitarian volunteers are both a vital asset and an increasingly vulnerable group. Their close proximity to communities and their lived experiences make them among the most trusted voices in the humanitarian ecosystem, but this trust is being tested. Volunteers are frequently exposed to the suspicions, misinformation and accusations directed at humanitarian organizations, especially when harmful narratives spread unchecked.

Contributor Insight 6.10



Trust-builders: The role of volunteers in countering disinformation

Volunteers play a crucial role in building resilience within their communities, branches and National Societies. They enhance each system's ability to withstand and recover from disruptions or stressors – precisely because they are an integral part of those systems. It is important to think of 'community' in a broad and inclusive sense: communities can include schools, sports centres, churches, digital platforms, social and support services, LGBTQ+ groups, neighbourhoods and more.

One of the most significant challenges facing humanitarian volunteers today is the emotional toll of dealing with hate speech and the constant flow of misinformation and disinformation. Volunteers are proud to serve and to be part of the world's largest humanitarian movement. That sense of identity can feel under threat when harmful or misleading information, especially hate speech, targets the systems they represent, whether at the community, branch or national level.

At the Italian Red Cross, we've learned through responding to emergencies, crises and disasters, including the COVID-19 pandemic, that restoring balance is closely linked to prevention and preparedness. This lesson also applies to information-related challenges. Strengthening media literacy and building a strategic approach to harmful information are essential components of building resilience.

We can and should develop clear strategies to respond to harmful information. Preparation is key. Not everyone has the time, training or expertise to fact-check content, analyse sources, interpret emotionally charged language or assess the credibility of online content. These are skills that require time and dedication.

Well-trained volunteers and informed community members are invaluable in meeting this challenge. Structured activities focused on filtering information – identifying reliable sources, validating content and distinguishing between trustworthy and misleading narratives – should be clearly defined and included in preparedness and response plans, with volunteers and community members actively involved. This work must take place across all levels: within the community, branches, headquarters and in collaboration with stakeholders and partners who bring relevant expertise.

Finally, acknowledgement and empathy are essential. Harmful information is part of our daily reality and its spread is accelerating. We must recognize – and remind others – that

being affected by harmful information is not a sign of naivety or ignorance. It touches everyone, everywhere – including managers, leaders and people with strong cultural or educational backgrounds. We must confront it together, as one Movement.

Carla M Orizondo Martinez

Head of Membership, Volunteering and Civil Services

Italian Red Cross

Volunteers occupy a dual role: they are both defenders of humanitarian principles and, increasingly, first responders to misinformation. They stand on the front line of community trust-building and rumour management – defending humanitarian principles, often while navigating complex and polarized information environments. Their vulnerability is further intensified by the simultaneous flow of information, images, video and audio across both digital and physical spaces related to the contexts in which they operate. To carry out this critical role effectively, volunteers must be part of developing clear guidance, practical tools and sustained support to engage confidently and credibly with affected communities on harmful information.

Voluntary service is a Fundamental Principle of the International Red Cross and Red Crescent Movement. As the nature of volunteering evolves, particularly in response to digital transformation and increased global connectedness, addressing the impacts of harmful information must become a central component of volunteer engagement strategies. Recognizing and responding to these challenges is essential to ensuring that volunteers can continue to serve safely, effectively and in accordance with the fundamental principles.³⁷

6.7.1

Red Cross and Red Crescent Safer Access Framework

The Safer Access Framework offers practical guidance for strengthening trust, acceptance and security in complex and high-risk environments, and is particularly relevant in the context of harmful information. It highlights the importance of understanding context, building community acceptance, managing risks and communicating effectively – both internally and externally.

The Safer Access Framework supports National Societies to proactively address such risks by investing in localized communication strategies, ensuring visible and consistent identification, and training staff and volunteers to navigate information threats. By embedding these practices into preparedness and response, humanitarian actors can better safeguard their operations and uphold the principles of neutrality, impartiality and independence in increasingly contested information environments.

The framework was developed by the International Committee of the Red Cross (ICRC) with input from 50 National Societies and is being updated in its operationalization to reflect the challenges of harmful information.

The framework has eight different elements:

- 1 Context and risk assessment for a clear understanding of the operational environment

- 2 Sound legal and policy base from which to carry out the humanitarian mandate in conformity with Movement partners' policies, international humanitarian law and domestic legislation
- 3 Building acceptance of the organization among key stakeholders
- 4 Building acceptance of individual staff and volunteers through conduct in accordance with the fundamental principles
- 5 Proper protection and promotion of the organization's visual identity
- 6 Implementing internal communication and coordination mechanisms
- 7 Implementing external communication and coordination strategies
- 8 Developing and implementing an operational security risk management system and structure

Contributor Insight 6.11

Trust and resilience: Key principles from BBC Media Action's practice

Drawing on BBC Media Action's experience across dozens of countries and crises, several principles can be identified for strengthening community trust and resilience against harmful information:

- 1 **Start with listening:** trust is built on two-way communication. Listening to concerns, questions and perceptions – even when they involve misunderstandings or harmful narratives – creates a basis for respectful engagement and more effective response.
- 2 **Invest before the crisis:** building trusted communication channels, supporting independent local media and forming partnerships with community leaders are pre-crisis tasks. These relationships determine how quickly and credibly information will be accepted in emergencies.
- 3 **Build community capacity:** to communicate within their own communities and represent community interests within the systems designed to help them.
- 4 **Work through trusted messengers:** community trust often rests with individuals – local radio hosts, health workers, religious leaders – who may not be part of formal humanitarian networks. Supporting and equipping these voices ensures vital information reaches people in ways they believe and will act on.

- 5 **Make communication locally relevant:** language, tone, timing and framing all matter. Messages must address real community concerns, using local languages, familiar formats and relatable stories that link advice to people's lived realities.
- 6 **Be transparent about uncertainty:** in fast-changing crises, it is better to explain what is not yet known than to risk later corrections that damage credibility. Acknowledging uncertainty fosters trust and reduces the space for harmful speculation.
- 7 **Embed verification and trust signals:** in an age of AI-generated and manipulated content, adopt verifiable methods, such as content provenance technologies, e.g., the Coalition for Content Provenance and Authenticity (C2PA), to make humanitarian information identifiable, authentic and harder to impersonate.
- 8 **Sustain engagement after the crisis:** post-crisis communication is essential for recovery, learning and preparation. It also helps maintain trust networks so they can be activated quickly in the next emergency.

By grounding these principles in long-term partnerships and shared ownership of the information space, humanitarian actors can build communities' resilience from the inside out, reducing the space for harmful information and strengthening the social fabric that underpins effective crisis response.

Melissa Everleigh

Senior Advisor

BBC Media Action

Alasdair Stuart

Head of Policy

BBC Media Action

6.8 Effective communication in the context of harmful information

In an era where harmful information spreads faster than ever, communication is no longer just a tool for outreach – it is a form of protection. For humanitarian actors, the ability to communicate effectively can shape whether communities trust guidance, follow life-saving instructions or turn instead to harmful information. The stakes are particularly high in crises, where clarity, timeliness and trust determine how people act under pressure.

Communication in humanitarian contexts serves multiple, overlapping purposes. It is, first, a channel for information-sharing – providing communities with timely, accurate guidance that can shape life-saving decisions. It is also a means of reinforcing reputation and credibility, helping to safeguard the principles and purpose of humanitarian action in environments where harmful narratives risk eroding trust. Communication also serves as a tool of community engagement, helping to reduce fear, clarify uncertainty and foster trust, cohesion and participation.

Above all, it is an act of bridge-building: creating and sustaining trust between humanitarian organizations and the people they serve, as well as among communities themselves. When these roles are aligned, communication becomes more than messaging – it becomes an essential part of humanitarian action, engagement and resilience.

Contributor Insight 6.12

Harmful information and the DANA floods in Valencia, Spain, 2024 (part 2 of 2)

(See Chapter 1 – Harmful Information and the DANA floods in Valencia (Spain) 2024 (part 1 of 2), on page 32)

In late October 2024, Valencia was struck by one of Spain's deadliest floods in recent history, caused by a DANA (*Depresión Aislada en Niveles Altos*, or 'cold drop'). Intense rain fell over several hours, overwhelming infrastructure, leaving 236 people dead and devastating communities.

Specific incidents of harmful information against the Spanish Red Cross spread on social media including negative comments, insults and threats. In response, all Spanish Red Cross communications on other topics, both nationally and locally, were paused on social media. We established centralized control over all press and social media communications across our 1,200 offices in Spain. In the first week, we set up and trained a 24-hour monitoring and response team of 30 members.

The most harmful messages to our reputation were identified and reviewed for possible responses. We replied directly to the false messages that had the most interactions. When replying only through social media was not sufficient, we contacted press experts and verification websites who analyse fake news. We then used their verified responses as reference links on our social media channels.

Dedicated monitoring of the DANA case began immediately on the night of 29 October. For greater reliability, we used two listening tools simultaneously – Clarabridge and Talkwalker – which allowed us to find and resolve possible monitoring gaps that emerged. Through these actions, we were able to detect new threats in real time and prepare responses. However, some cases – such as accusations about financial issues – were too complex to address quickly, which allowed certain false claims to spread for longer.

Some people in areas where we had not yet been able to help also began to criticize us. A small number of volunteers who wanted to help, but had not been mobilized, joined the criticism, expressing disagreement with our actions on the ground. Some political parties that do not support the Red Cross also used this moment to criticize the Spanish Red Cross. Certain people affected by the crisis who aligned with these parties even shouted at us in the street. However, the majority of people we helped expressed their support.

Since the first day of the DANA, there has been a sharp increase in messages against the Red Cross. In November and December, our social media mentions rose by 5,000%, with

around 75% of these being negative messages, an increase of 8,824% from the previous two months.

Fig 6.1

Tone of social media messages about the Spanish Red Cross, October–December 2024

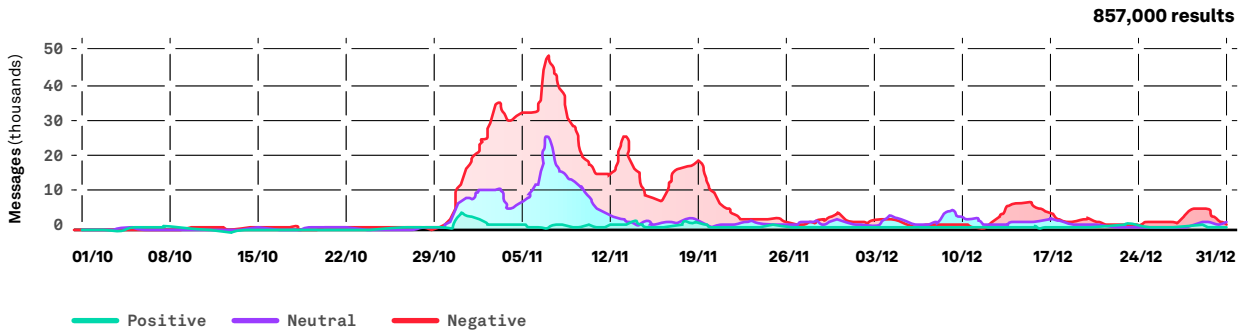


Fig 6.2

Proportion of social media messages about the Spanish Red Cross that were positive, neutral and negative, October–December 2024



368,900 results

75.6% negative
 19.9% neutral
 04.5% positive

In November and December alone, the newspaper *La Gaceta* published 32 articles criticizing the Spanish Red Cross. During the same period, some politicians posted 104 messages against the Spanish Red Cross. Pro-Israel groups also seized the opportunity to amplify the posts against the Red Cross.

Most of the accounts attacking us were anonymous. While much of the criticism focused on the DANA, it also targeted issues such as migration, the financial management of the Spanish Red Cross and claims that the Red Cross was not present and working on the ground.

The harmful information spread about our National Society and other humanitarian actors in this case did not have a significant impact on public institutions, as they were being attacked equally – along with civil society in general. However, it did impact some of the companies that collaborate with us, which were also targeted. The broader public impact was evident in the many questions raised on social media, as well as by individual members contributing through their membership fees.

Lessons learned and opportunities for improvement

Team resizing: The internal expansion of the social media and press team would have been more effective if it had been carried out earlier in the emergency. At the time, we lacked local or international references on the right timing; now we have learned when such adjustments should be made.

Partnerships and external support: Engaging a reputation management agency earlier would have been more beneficial. A temporary upgrade of monitoring and response tools could also have strengthened our capacity. Our partnerships with media groups proved to be key in the crisis management.

In the Spanish Red Cross social media protocol (an internal document), introduced in January 2025, we now outline steps to follow in any emergency or communication crisis. It sets out, step by step, the roles and responsibilities of each person, warning levels, risk assessment measures and corresponding actions. Over the year following the DANA, we strengthened the completeness of information across all communications, including internal communication and brand protection. Alongside the crisis and social media protocol, all our social media accounts are protected with triple security measures, with a limited number of people holding full access rights.

We have also redefined our global communication strategy, recognizing that such incidents are likely to recur. In addition, we actively participate in various groups and initiatives on harmful information from a communications perspective, including within the Movement, on the *Plataforma del Tercer Sector* (Third Sector Platform), and in a new initiative led by the Spanish Government's communications department in collaboration with other humanitarian actors' communications teams.

Recommendations

- 1 Greater involvement of Red Cross volunteers in social media is needed. Establishing a digital volunteer group to serve as brand ambassadors and defenders, in a systematic way, could be effective.
- 2 Better regulation of social media accounts, as account anonymity is increasingly harmful, allowing false messages and hate speech to spread without accountability. The same laws that apply offline should also apply online.

Observations

- 1 Disinformation accusing the Spanish Red Cross of prioritizing helping migrants and/or people overseas over Spanish citizens undermines the perception of the principle of impartiality.
- 2 Disinformation accusing the Spanish Red Cross of supporting government immigration policies undermines the perception of our principles of independence and neutrality.

- 3 Disinformation alleging that the Spanish Red Cross is not present on the ground undermines the perception of the principle of humanity.

María Trénor Alvargonzález

Director of Communication

Spanish Red Cross

Kenan Terzic

Social Media Manager

Spanish Red Cross

The critical elements of strategic communication in a digital age – the **right audience, right message, right time and right messenger**³⁸ – are especially important in efforts to counter harmful information. These four components are deeply interdependent. Even a well-crafted message may fail to resonate if it is delivered to the wrong audience, is delivered by an untrusted source or arrives at an inopportune moment. Missteps in any one of these areas can unintentionally reinforce harmful information, deepen distrust or drive disengagement.

Effective communication requires strategic precision: understanding who needs the information, what matters to them, when they are most receptive and who they are most likely to trust. In the context of harmful information, this approach is not just good practice – it is essential to building and sustaining trust, engagement and resilience.

Table 6.1 **The critical elements of strategic communication in a digital age**

Element	Guiding questions	Considerations in practice
Right audience	Who is affected or at risk? Whose behaviour, perception or trust needs to shift?	Disaggregate by gender, age, language, access, vulnerability. Map trusted networks and local influencers.
Right message	What do they need to know? What matters to them? What concerns are they expressing?	Ground messages in community concerns. Ensure cultural and contextual relevance. Focus on what people want to understand, not just what responders want to say.
Right time	When is the audience most receptive? When do they need the information to make decisions or act?	Provide information early enough to act. Share continuously as situations evolve. Align delivery with key decision points or moments of highest relevance. Avoid overload; repeat key messages to reinforce understanding and trust.

Element	Guiding questions	Considerations in practice
Right messenger	Who do they trust? Who can deliver the message with legitimacy and empathy?	Use trusted local figures such as community leaders, health workers, religious leaders, youth, peers. Ensure diversity, credibility and cultural sensitivity.

6.9 Critical reflections: Challenges in practice

Silence is not a strategy when confronting harmful information. Choosing not to respond – whether out of caution, fear of amplifying the issue or uncertainty – can allow false narratives to take root and spread unchecked. In many contexts, the absence of credible voices creates a vacuum that harmful actors are quick to fill and exploit, often reinforcing fear, confusion or division. While responses must be carefully considered and grounded in local realities, remaining silent can be interpreted as indifference, weakness or even complicity. Experience across crises shows that timely, trusted and context-sensitive communication is essential – not just to counter falsehoods but to build confidence, foster dialogue and strengthen community resilience to the real-world consequences of harmful information. A decision-making framework or response tree specific to each organization may help guide whether, where, when and how to respond.

At the same time, community engagement, while essential in responding to harmful information, is not without risks. Community voices, if not carefully supported, can be co-opted, censored or politicized, especially in polarized environments or situations of armed conflict and violence. Individuals may face retaliation or backlash from authorities, political factions or other groups. Ethical concerns also arise when community members are asked to act as monitors, especially if roles and expectations are unclear or if the engagement puts them at risk. Without clear safeguards around consent, confidentiality and safety, community-based efforts risk undermining trust or unintentionally causing harm.

Digital exclusion adds another layer of complexity. Unequal access to devices, connectivity and digital literacy limits who can participate in information ecosystems. These disparities frequently mirror and reinforce existing inequalities, particularly along lines of gender, age, geography and socioeconomic status.

As discussed earlier in this report, migration remains one of the most politically charged and polarizing issues of our time. The high political stakes surrounding migration control and management strategies as well as public perceptions of migration present serious challenges and dilemmas for humanitarian responders. In particular, harmful information is used to delegitimize humanitarian activities, for example, in the Mediterranean,

where NGOs involved in search and rescue missions have faced media attacks, political accusations of collusion with smugglers and public backlash. These narratives not only erode trust in humanitarian actors but can also restrict operational access, reduce funding and endanger them and the people they seek to help.

Migrants themselves often rely primarily on informal sources, such as family, friends and social networks, both before and during their journeys. Barriers to information access are also acute. Migrants may avoid seeking information out of fear of detection, deportation or discrimination. Language barriers, limited connectivity and the perception that official channels are untrustworthy further compound the challenge. In such contexts, harmful information can thrive – exploiting gaps in access, fear and exclusion – making it even more critical to ensure migrants receive timely, accurate and trusted information that enables informed decision-making and protects their dignity and rights.

◇ Common feedback from contributors to this report: Top five take-aways

- 1 **A coordinated approach to harmful information** – including common protocols for identifying, verifying and responding to rumours and harmful information.
- 2 **Accessible training and tools** – to help staff and volunteers understand and navigate digital and information risks, including harmful information tactics.
- 3 **Greater investment in digital and information literacy** – for both humanitarian actors and the communities they serve.
- 4 **Stronger engagement with media and technology platforms** – to flag harmful content, promote trusted sources and amplify accurate narratives.
- 5 **Guidance on principled and context-sensitive communication** – particularly neutrality and impartiality – when addressing politically charged or emotionally sensitive harmful information. Guidance should support proactive strategies such as prebunking and accurate framing, ensuring responses are principled, context-aware and protect both staff and affected communities.

Concluding remarks: From communication to collaboration

Addressing harmful information effectively requires more than delivering messages, it demands a fundamental shift in mindset: from disseminating information to enabling participation and agency. Too often, communication efforts are designed as one-way flows, with external actors determining what information is relevant or important, how it is framed and who gets to speak. This top-down approach not only risks reinforcing existing power imbalances but also overlooks the insights, lived experiences and priorities of people in need.

In contrast, when communities are actively involved in shaping the narratives, identifying risks and driving the response, they are no longer passive recipients of information but co-creators of solutions. Building resilience to harmful information is not simply about correcting falsehoods – it is about restoring trust, deepening dialogue and empowering people to navigate information environments.

This shift toward shared agency requires:

Recognizing and supporting trusted local actors, from youth leaders and women’s groups to religious figures and community media, who are best positioned to engage meaningfully with their peers and people in need.

Investing in local information ecosystems – community radio, local journalists, digital access points, feedback mechanisms – that provide continuous, context-relevant communication before, during and after crises.

Embedding media and digital literacy into humanitarian programming, enabling individuals to assess and challenge harmful content and engage in informed decision-making.

Creating space for participatory content creation, allowing communities to express their perspectives in formats that resonate culturally and emotionally, whether through storytelling, theatre, radio or social media.

Safeguarding community engagement, ensuring that participation does not expose individuals to risks, especially in politically polarized or fragile environments.

As highlighted throughout this report, harmful information is not just a communications challenge, it is a threat to humanitarian access, acceptance and security, and as such requires a whole-of-organization and whole-of-society approach. Addressing it requires systems-thinking, sustained collaboration and a willingness to share control. Just as humanitarian actors need to fully embrace the localization agenda in service delivery, the same commitment is needed in response to harmful information: to listen, co-create and build responses grounded in trust, transparency and mutual accountability.

The critical elements of strategic communication in a digital age – the right audience, right message, right time and right messenger – are especially important in efforts to counter harmful information. These four components are deeply interdependent.

Ultimately, resilience to harmful information is not achieved through better messaging – it is built through better relationships. By investing in local voices, community ownership and inclusive approaches, the humanitarian sector can move beyond reactive counter-narratives and toward a more sustainable, principled and people-centred response. Shifting toward shared agency means recognizing that trust and legitimacy come not from broadcasting facts, but from meaningful participation, dialogue and mutual accountability.

Note: The author gratefully acknowledges the contributions of the following individuals to [section 6.3](#) and the insights reflected in the community member quotes included in each chapter: Ewané Fabrice Adrien, Nnang Minkoulou Pauline Santana, Essounga Jeanne Claris, Hadidjatou Sali, Fadma Mahamat, Imashela Abram Nambayo, Mambepa Mulilo, Mubila Majaluwa, Ziyi Min, Chen Xi Iris, Pengye Zhang, Mark Cainen, David Mackay, Andriamirado Malalatiana Michel, Rasoanomenjanahary Elisah Nahavitatsara, Rakotolava Safidinirina Précila, Be Ravololona Stanicia, Raveloson Elodye Larissa, Raharimanana Hery-Zo, Aaron Onyede, Ocaya Sunday, Hadaya Alli, Abdalla Fouda, Zinab Al Mohammad, Safa Al-Kuhlani, Dr Abrar Hayder Alameen Ahmed, Ron Adamba, Otero Otieno Wycliffe, Angellah Praygod Mochawa, Naomi Odwa, Joseph Karanja, Martha David Ibon, Guisel Guadalupe Serrano, Ximena Carmen Vicente Machaca, Samantha Mendez, Nicol Victoria Grimaldis Oño, Navindu Ratnayaka, M.E Praveen Waas, Suvini Rathnayake and Sachintha Ayesha Abeyasinghe.

Asks, aims and recommendations

Asks

Put communities at the heart of information resilience by investing in inclusivity, accountability and collaboration. This includes supporting feedback mechanisms, local verification and community-owned narratives that strengthen locally trusted information systems, enabling principled humanitarian action and more effective crisis response.

Aims

Strengthen principled humanitarian action by embedding community voices and verification systems within information ecosystems.

Ensure humanitarian responses are principled and accountable by using information and narratives grounded in local perspectives, reflecting diverse community voices, and promoting trust and inclusivity in decision-making.

Strengthen operational resilience by ensuring diverse voices guide humanitarian decisions and responses.

Recommendations

States and policy-makers

- Resource community-led verification and communication by funding rumour tracking, independent journalism and inclusive channels (radio, digital access points, offline formats) in local languages.
- Strengthen information ecosystems critical to humanitarian action by investing in local capacity and information literacy, and promoting stronger responses from technology

platforms and proactive measures to counter harmful information.

- Empower trusted local actors – including youth groups, women’s networks and community media – through partnerships, resources and safeguards, and integrate them into local information ecosystems to support verification, feedback and principled humanitarian action during crises.

Humanitarian actors

- Build accountability and trust through co-created messages, community feedback loops, transparent reporting and joint review with communities – ensuring that

engagement respects humanitarian principles and avoids politicization.

- Establish early-warning and rapid response systems to monitor

information patterns and narratives that affect humanitarian action, amplify trusted messengers and enable timely responses at the local level – while safeguarding privacy.

- Support community-led content by enabling safe, independent storytelling and locally relevant information in local languages, with robust safeguarding measures to protect contributors and ensure inclusion.
- Strengthen local journalists and media through capacity building for timely and accessible multilingual content across both online and offline platforms – while maintaining neutrality and avoiding alignment with political agendas – to

support principled humanitarian crisis response and a resilient local information ecosystem.

- Measure and learn by tracking trust, participation quality, the inclusion of marginalized voices and timeliness of corrections – using anonymized or aggregated data – and share lessons learned to improve principled humanitarian responses across contexts.
- Ensure strategic communication in humanitarian contexts aligns the right audience, message, timing and messenger, recognizing their interdependence, to effectively counter harmful information and support principled, context-sensitive responses.

Communities and local leaders

- Generate and amplify community-owned narratives that build trust and support principled humanitarian actions during crises.
- Act as trusted intermediaries facilitating – where required – dialogue between humanitarian actors, authorities and populations in need to ensure two-way

accountability and respecting and supporting the role of principled humanitarian organizations.

- Safeguard inclusivity in local information systems so that diverse voices – including groups that are marginalized – are represented and heard.

Endnotes

- 1 IFRC. Global Platform for Disaster Risk Reduction (GPDRR) 2025: Call to Action (2025) www.ifrc.org/sites/default/files/2025-06/IFRC_Global%20Platform%20for%20Risk%20Reduction%202025_Call%20to%20Action.pdf
- 2 See, for example, Grand Challenges Canada. Creating Hope in Conflict. *Navigating the Information Landscape: Misinformation, Disinformation, and Hate Speech in Humanitarian Contexts*. Scoping study. (2024) p.30 <https://humanitariangrandchallenge.org/mdh-report-launch>
- 3 IFRC. *Framework for Community Resilience* (2014) p.6 www.ifrc.org/sites/default/files/IFRC-Framework-for-Community-Resilience-EN-LR.pdf
- 4 Ibid, p.10
- 5 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) p.169
- 6 IFRC and ICRC. *A Red Cross Red Crescent Guide to Community Engagement and Accountability*. (2021) www.ifrc.org/sites/default/files/2022-05/RCRC_CEA_Guide_2022.pdf
- 7 IFRC. *Framework for Community Resilience* (2014) p.6 www.ifrc.org/sites/default/files/IFRC-Framework-for-Community-Resilience-EN-LR.pdf
- 8 IFRC Americas. *Impacto de los Huracanes Eta e Iota en Honduras: Participación Comunitaria en la Respuesta*. YouTube. (Spanish) (2025) www.youtube.com/watch?v=iSV0U1PqeM0
- 9 IFRC Americas. *La Importancia de la Retroalimentación en la Respuesta Humanitaria*. YouTube. (Spanish) (2025) www.youtube.com/watch?v=hizXEfbwdt8
- 10 IFRC Americas. *Strengthening Coordination for Effective Humanitarian Action*. YouTube. (2025) www.youtube.com/watch?v=We5qhbJtsuo
- 11 IFRC. *Community Engagement and Accountability Impact Research: Guatemala Case Study* (Spanish) Community Engagement Hub. (2023) <https://communityengagementhub.org/resource/community-engagement-and-accountability-impact-research-guatemala-case-study>
- 12 IFRC. *Panamanian Red Cross Coordination with Government Leads to Better Responses*. Community Engagement Hub. (2025) <https://communityengagementhub.org/resource/panamanian-red-cross-coordination-with-government-leads-to-better-responses>
- 13 Ecuadorian Red Cross and IFRC. *Cohesión Social Ecuador 2023*. Community Engagement Hub. (2024) Analytical report: <https://communityengagementhub.org/es/resource/cohesion-social-ecuador-2023-reporte-completo> and executive report: <https://communityengagementhub.org/wp-content/uploads/sites/2/2024/02/REPEJECUT-COHESOCIAL-IFRC-CRE-PPP.pdf>
- 14 See Red Crescent Society of Kyrgyzstan. www.redcrescent.kg/ru/about/docs
- 15 Malteser International. *Democratic Republic of the Congo: Applying the People First Impact Method (P-FIM) in the Context of the Ebola and Covid-19 Response*, Cologne (2020) www.malteser-international.org/fileadmin/Files_sites/malteser-international/B-Our_Work/Africa/DR_Congo/P-FIM_in_Ebola_and_Covid-19_response_Malteser_International.pdf
- 16 Pierre, J. *False: How Mistrust, Disinformation, and Motivated Reasoning Make Us Believe Things That Aren't True*. (2025) p.178
- 17 Public Health Agency of Canada. *Rapid Review: An intersectional analysis of the disproportionate health impacts of wildfires on diverse populations and communities*. (2024) www.canada.ca/en/public-health/services/publications/healthy-living/rapid-review-intersectional-analysis-disproportionate-impacts-wildfires-diverse-populations-communities.html
- 18 Internal Displacement Monitoring Centre. *Canada's Indigenous peoples increasingly at risk of disaster displacement*. PreventionWeb and UNDRR note summarizing IDMC data. (2024) www.preventionweb.net/news/canadas-indigenous-peoples-increasingly-risk-disaster-displacement
- 19 Each community researcher was trained and supported to conduct semi-structured interviews in their preferred language (Arabic, French, Chinese, English, Spanish, Kiswahili, Malagasy or Acholi), analyse the data and summarise key issues. Volunteer academics from the University of Michigan, Northumbria University and Open Lab (Newcastle University) then synthesized these findings at the global level, while maintaining local nuance.
- 20 Grand Challenges Canada. Creating Hope in Conflict. *Navigating the Information Landscape: Misinformation, Disinformation and Hate Speech in Humanitarian Contexts*. Scoping study. (2024) p.81 <https://humanitariangrandchallenge.org/mdh-report-launch>
- 21 Ibid
- 22 Malawi Red Cross Society, IFRC and USAID. "When the Hippo Walks, We Know a Flood Is Coming": Community Perceptions on Locally Led Adaptation to Climate Change in Malawi. (2024) https://communityengagementhub.org/wp-content/uploads/sites/2/2025/07/REPORT-Malawi-Perception-Study-on-LLA-_2024-FINAL.pdf
- 23 IFRC and Plan Eval. *Engaged, Informed and Empowered: The Impact of Community Engagement on Public Health Outcomes*. Community Engagement Hub. (2025) <https://communityengagementhub.org/resource/engaged-informed-and-empowered>
- 24 IFRC. *Community Engagement and Accountability Impact Research: Georgia Case Study*. Community Engagement Hub. (2023) <https://communityengagementhub.org/resource/community-engagement-and-accountability-impact-research-georgia-case-study>
- 25 IFRC. *Community Engagement and Accountability Impact Research: Guatemala Case Study (Spanish)*. Community Engagement Hub. (2023) <https://communityengagementhub.org/resource/community-engagement-and-accountability-impact-research-guatemala-case-study>
- 26 IFRC. *Community Engagement and Accountability Impact Research: Guinea Case Study (French)*. Community Engagement Hub. (2023) <https://communityengagementhub.org/resource/community-engagement-and-accountability-impact-research-guinea-case-study>
- 27 IFRC. *Community Engagement and Accountability Impact Research: Indonesia Case Study*. Community Engagement Hub. (2023) <https://communityengagementhub.org/resource/community-engagement-and-accountability-impact-research-indonesia-case-study>
- 28 IFRC. *Community Engagement and Accountability Impact Research: Malawi Case Study*. Community Engagement Hub. (2023) <https://communityengagementhub.org/resource/community-engagement-and-accountability-impact-research-malawi-case-study>
- 29 Nepal Red Cross Society and IFRC. *From managing misconceptions to building trust in Covid-19 vaccination*. (2023) https://communityengagementhub.org/wp-content/uploads/sites/2/2023/02/From-managing-misconceptions-to-building-trust-in-Covid-19-vaccination_Sarif-31012023.pdf
- 30 International Telecommunication Union (ITU). *Facts and Figures 2024. The Gender Digital Divide*. (2024) www.itu.int/itu-d/reports/statistics/2024/11/10/ff24-the-gender-digital-divide.
- 31 UN General Assembly. *Intensification of efforts to eliminate all forms of violence against women and girls: technology-facilitated violence against women and girls: Report of the Secretary-General*. (A/79/500) 79th session (2024) <https://undocs.org/en/A/79/500>
- 32 ITU-Broadband Commission. *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression*. (2020) p.256–257 www.broadbandcommission.org/publication/balancing-act-countering-digital-disinformation

- 33 IFRC. *Stronger faster safer: Epidemic preparedness and response success stories from the Programmatic Partnership 2022-2025*. (2025) p.22–23 www.ifrc.org/sites/default/files/2025-06/2025_IFRC_ECHO_Epidemic_Preparedness_Stories_EN.pdf
- 34 This objective acknowledges that citizens must have access to trustworthy, balanced and comprehensive information about current affairs, the actions of decision-makers and other factors relevant to their personal decisions. It underscores the need for proactive engagement and holistic strategies that address the multiple factors required to maintain a healthy information ecosystem. Compare: UN Development Programme. *Information Integrity: Forging a Pathway to Truth, Resilience and Trust* (2020) www.undp.org/sites/g/files/zskgke326/files/2022-02/UNDP-Information-Integrity-Forging-a-Pathway-to-Truth-Resilience-and-Trust.pdf
- 35 Bateman, J. and Jackson, D. *Countering Disinformation Effectively: An Evidence-Based Policy Guide*. Alverne, C., Ross Arguedas, A., Banerjee, S., et al. The Electoral Misinformation Nexus. *Public Opinion Quarterly* 2024, 681–707
- 36 Fondation Hironnelle. *Building Trusted Media in a Fragile Setting – An Evaluation of Radio Ndeke Luka in Central African Republic* (2024) www.hironnelle.org/wp-content/uploads/2025/04/RNL-External-Evaluation-Report_final_clean.pdf; compare: Fondation Hironnelle. *The impact of Radio Ndeke Luka before and after its launch in three remote regions of CAR* (2025) www.hironnelle.org/en/bekou-study-the-impact-of-radio-ndeke-luka-before-and-after-its-launch-in-three-remote-regions-of-car
- 37 IFRC. Volunteering Policy (2022) www.ifrc.org/sites/default/files/2022-10/20220822_IFRC-Volunteering-Policy-EN.pdf
- 38 DiResta, R. *Invisible Rulers: The People who Turn Lies into Reality*. (2024) p. 357



Chapter 7

Upholding humanitarian principles in the age of echo chambers



Chapter 7



Upholding humanitarian principles in the age of echo chambers

Contents

	Introduction: Outpaced – humanitarian action in the era of instant narratives	259
7.1	Reaffirming and applying humanity in an age of distrust	261
7.2	Dehumanization in the digital age	262
7.3	Promoting tolerance in diverse and divided societies	264
7.4	Neutrality at a cost: The price of not taking sides	267
7.5	Patriotism and humanitarian principles	273
7.6	Independence as a condition for trust	274
7.7	Impartiality as a compass	275
7.8	Organizational integrity: Internal alignment with principles	276
7.9	A framework for applying humanitarian principles in the information age	282
7.10	Influencers: Connectors and dividers	290
7.11	Reaffirming humanitarian principles in the age of digital tools	292
	Concluding remarks: Principled action requires more than declarations	293
	Endnotes	297

Introduction: Outpaced – humanitarian action in the era of instant narratives

In an information environment distorted by rumour, manipulation and politically charged content, trust in humanitarian actors is increasingly fragile. Harmful information can undermine the perceived impartiality, neutrality and independence of humanitarian actors, fuel suspicion among communities and endanger staff, volunteers and the people they serve.

Humanitarian action now takes place in a contested information environment, one where trust is fragile and emotion often displaces fact. As TS Eliot cautioned, “When we do not know, or when we do not know enough, we tend always to substitute emotions for thoughts.” In the sentiment-driven economy of social media and now artificial intelligence (AI), principled humanitarian action is easily misunderstood or misrepresented – especially when it fails to align with dominant political or emotional narratives. In such settings, the consistent application of humanitarian principles becomes not only more difficult but more essential.

Information now competes not on the strength of accuracy, evidence or logic, but on its ability to provoke emotion – amusement, outrage and shock – and to change behaviours. Harmful content spreads rapidly and often unchecked, not because it is true, but because it feels true. Harmful information and distortion flourish where trust is already weakened. A single manipulated image, misleading headline or viral post can cast doubt on an organization’s motives or affiliations, reinforcing narratives that challenge principled humanitarian action and endanger humanitarian actors and the communities they serve.

This chapter begins with the principle of humanity – the foundation of all humanitarian work – and builds on [Chapter 2, on page 69](#) where the focus was on trust and the principle of neutrality. It explores a growing strategic vulnerability: humanitarian action is designed to be deliberate, impartial and context specific. Digital information, by contrast, is immediate, emotional and virally amplified. This creates a dangerous speed mismatch. Humanitarian actors generally take time to verify facts, assess needs and coordinate responses. Harmful narratives and outrage, by contrast, travel in seconds – unconcerned with humanity, neutrality, accuracy, proportionality or consequence.

Humanitarian responses – at least the visibility around them – often arrive *after* public narratives have hardened, reputations damaged and trust eroded. The cost of delay is that narratives are settled early, often by politicized, dehumanizing or false information – creating operational risks. The humanitarian sector’s commitment to accurate, non-political communication is too often misread as indifference, lack of empathy, evasiveness or even sometimes complicity.

Rapid, emotionally charged narratives can shape donor, community, media and influencer perceptions before humanitarian organizations can establish the facts. Deliberate delays in communication – once a safeguard to establish facts and engage – can now erode humanitarian space for communication. Principled, timely engagement in the information environment is essential. What was once considered caution may now be a liability: inaction in the information space can cause more harm than timely, principled engagement. This raises difficult questions: is the slow pace of visibility

around humanitarian action truly necessary or have organizations become institutionally over-cautious? Who is the audience for such communications today? Do some organizations still treat information as secondary to operations, even though it now shapes how operational success or failure is perceived?

As harmful narratives accelerate while communication about humanitarian action lags, the sector faces declining trust and shrinking operational space. Addressing this challenge requires confronting these dilemmas directly and adapting communication practices to keep pace with the evolving information environment.

◇ Fundamental and humanitarian principles

In 1965,¹ the International Red Cross and Red Crescent Movement proclaimed a set of ethical and operational principles to guide humanitarian work, especially in polarized and contested environments. The seven fundamental principles were adopted in the Statutes of the Movement by the International Conference in 1986: humanity, impartiality, neutrality, independence, voluntary service, unity and universality. Four of them – **humanity, impartiality, neutrality and independence** – have since become the foundation for international humanitarian action and are widely recognized as the **humanitarian principles**.

The principles are both hierarchical and complementary, providing an ethical and a practical framework. Humanity and impartiality are the two substantive principles. **Humanity** expresses the purpose of humanitarian action: to prevent and alleviate suffering wherever it is found. **Impartiality** ensures that this action is driven solely by needs, without discrimination based on nationality, race, religious beliefs, class or political opinions.

Neutrality and independence are operational tools – pragmatic means to secure access and maintain a protected space for humanitarian action in situations of armed conflict and violence. Both are designed to build trust or at least acceptance to act from all sides. **Neutrality** requires humanitarian actors to refrain from taking sides in hostilities or engaging in political, religious or ideological debates. **Independence** ensures that decisions, particularly around needs assessment and response, are made autonomously, free from external influence or agendas.

The humanitarian principles are not unique to the Movement, though they form its ethical and operational core. Over time, they have shaped global humanitarian norms, including UN General Assembly resolutions, such as Resolution 46/182² (1991) which affirms these principles as the foundation of UN-led humanitarian coordination, and several hundred NGOs have now adopted the Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief (1994).

7.1 Reaffirming and applying humanity in an age of distrust

Fundamental principle

HUMANITY

The International Red Cross and Red Crescent Movement, born of a desire to bring assistance without discrimination to the wounded on the battlefield, endeavours, in its international and national capacity, to prevent and alleviate human suffering wherever it may be found. Its purpose is to protect life and health and to ensure respect for the human being. It promotes mutual understanding, friendship, cooperation and lasting peace amongst all peoples.

If trust is rooted in vulnerability and expectation, humanity is grounded in respect and compassion. The principle of humanity is widely recognized as the foundational driver of humanitarian action. It is what compels humanitarian actors to act: to save lives, reduce suffering and uphold the dignity of people affected by crises. Among the core humanitarian principles, humanity is superior to the others because it captures the moral imperative, the core motivation and founding values of humanitarianism and underpins the other principles. In his 1979 *Commentary on the Fundamental Principles*,³ Pictet described humanity as requiring not only the preservation of life and physical integrity but also a respect for individual personality and dignity. Pictet emphasized that the principle must evolve with ‘circumstances’. The information age is one such circumstance.⁴

Philosopher Jonathan Glover in his exploration of moral identity and psychological distance argues that **humanity is rooted in our human responses**: the capacity for respect and for sympathy, our ability to care about the suffering of others and to see them as fully human.⁵ Yet these responses are inherently fragile and it is precisely this fragility that harmful information exploits across social media platforms and beyond. Glover observes that sympathy can be overwhelmed, weakened, narrowed or eliminated by psychological or physical distance, by tribalism and belief or through the normalization of dehumanizing language. He warns that distance narrows human responses, cutting people off from whole groups of other people, while remoteness makes it natural to think all this is not really happening: “This enables the propaganda of atrocity, often directed against the dignity of victims”.⁶

Harmful information thrives in such environments and it erodes empathy, undermines solidarity and legitimizes exclusion or violence. The result is not only a more hostile environment for humanitarian operations, but one in which suffering is selectively acknowledged and humanitarian response becomes politically contested. As humanitarian space narrows, as physical proximity to affected populations for principled humanitarian actors becomes more difficult, restricted and/or delegitimized, the principle of humanity itself is under growing threat.

7.2 Dehumanization in the digital age

Dehumanization is no longer confined to explicit hate speech or fringe ideologies. It is increasingly embedded in everyday language, imagery and digital behaviours. It surfaces in manipulated images, inflammatory headlines and algorithmically amplified narratives that distort, erase or deny an individual's humanity. In already fragile environments, the erosion of empathy is not just a side effect; it is often the intent. Dehumanization is not always overt. It can manifest subtly and systemically, including through digital profiling, unequal access to services or the commodification of personal data. At its core, dehumanization involves perceiving or treating someone as *less than human* – a denial of one or more elements of their humanity.⁷ While people never cease to be human, failing to recognize their humanity has real and often violent consequences. This includes reducing individuals to group identities (such as 'migrant', 'refugee'), equating them with animals or vermin, or disregarding their legitimate human interests and agency. When such rhetoric goes unchecked, it not only diminishes empathy and compassion but also increases public support for exclusion, retribution or even violence. It undermines helping behaviours, obstructs reconciliation and fuels policies and practices that may violate international law.⁸

Dehumanizing language and imagery directed at humanitarian action pose serious risks. They can be used to justify or encourage the obstruction of aid, lead to the targeting of humanitarian organizations, criminalize humanitarian assistance and render entire populations invisible or undeserving in the eyes of the public or decision-makers. These risks are particularly acute when such rhetoric originates from official sources or when humanitarian access is framed as politically motivated or a threat to national security.



Extracts from 2024 Council of Delegates Resolution 5: Call for respect and support for principled humanitarian action

“expressing deep concern for the safety and well-being of all people affected and for their ability to access life-saving assistance, and expressing sorrow for the number of lives of humanitarian workers and volunteers lost in the service of humanity.”

“expressing deep concern about the greater impediments to the ability of principled humanitarian actors to deliver assistance and protection to people who most need it, owing to increasing political and societal polarization, the politicization of aid, false characterization of the humanitarian principles of impartiality and neutrality as furthering the interests of an adversary ...”⁹

Upholding the principle of humanity in the digital era requires vigilance against the ways in which the information space and technology can unintentionally harm (online and offline), exclude or depersonalize. It demands sustained commitment to human presence and relational accountability – one reason why traditional humanitarian actors have long emphasized the importance of physical proximity and human connection to people in need.

Dehumanization not only challenges the principle of humanity, it also undermines impartiality, by denying the equal value of all people and the right to humanitarian response based solely on need. It casts suspicion on neutrality, framing humanitarian actors as aligned with or sympathetic to those who have been publicly stigmatized or discredited; it erodes perceptions of independence, as humanitarian intentions and affiliations are deliberately misrepresented.

The principle of humanity compels us to affirm the intrinsic value of every individual, especially when their dignity, rights and personhood are denied. Impartiality reinforces this imperative by requiring humanitarian action to be based solely on need. Neutrality and independence enable humanitarian actors to maintain focus on those people most affected, while navigating highly politicized environments and pressure to devalue or exclude individuals or groups. Together, they are not constraints but the basis for principled action, especially in complex contested spaces.

Yet even well-intentioned humanitarian actors can inadvertently replicate dehumanizing patterns if they are not actively mindful of the language they use or the power dynamics they reinforce. This is why principled humanitarianism must go beyond ethical commitments to uphold humanity: it must include deliberate efforts to detect and disrupt dehumanizing narratives, especially those normalized “in the undramatic episodes of the day-to-day”.¹⁰ The antidote to dehumanization is not simply more information – it is recognition. It is the act of restoring visibility and dignity to those denied it and reaffirming – through words, images and action – that no one is ever less than human. **The humanitarian principles are not only capable of surviving today’s crisis of harmful information, they were designed to meet precisely this type of challenge.**

Contributor Insight 7.1



Navigating harmful narratives in a migration hotspot

Since June 2023, the Italian Red Cross has been running the migrant reception centre (‘hotspot’) on Lampedusa. The hotspot plays a critical role in managing sudden migrant arrivals. Previously, the Lampedusa hotspot has been the target of numerous harmful information campaigns, particularly on social media. Misleading videos have also spread widely, including footage falsely portraying migrants lying on the ground and episodes of violence in the centre. These distortions, often amplified by sensationalist rhetoric, framed the situation as an ‘invasion’ and are frequently used to stoke fear. From the outset, our goal was to transform the hotspot into a hub of humanity. This initially presented serious challenges: to improve operational management, expand reception capacity and build a constructive relationship with the Lampedusa community, institutions and authorities.

From the very first days, we worked to adapt the centre and align its operations with the standards of the Italian Red Cross and the broader International Red Cross and Red Crescent Movement. One of our first steps was to inform the media about the changes made – both to the infrastructure and the reception system. This allowed us to

communicate the improvements achieved and the increased care provided to migrants arriving at the hotspot.

Authorities, institutions and the local population quickly realized that there had been a significant shift in the way reception of migrants was managed on the island. In a short time, Italian media outlets began to report positively on our work at the hotspot. However, in some European countries where coverage of migrant hospitality is more negative, media engagement required a different approach. We developed tailored strategies for English and French language media to better convey our values. Media in Dutch, German and Arabic languages quickly highlighted the positive impact of the hotspot change of management approach.

Beyond operational and logistical improvements, we also launched a storytelling initiative centred on the emotions and lived experiences of the migrants arriving in Lampedusa. This helped shift the narrative – from one focused solely on numbers to one showing human stories and dignity. Over time, we have observed the media spreading false information or inaccurate numbers. Thanks to our strong relationships with newspapers, television and radio outlets, we have been able to position the Italian Red Cross as the most trusted source for migrant-related news on Lampedusa – and beyond. Today, more than two years since we began managing the hotspot, journalists continue to turn to us for data, updates and fact-checking. They regularly seek to verify news – often false – that they receive via social media or private messages. There have been particularly intense moments, such as during periods of overcrowding at the hotspot or the arrival of survivors from shipwrecks. Yet, our consistent presence and reliability have enabled us to navigate these difficult situations – delivering timely messages across both traditional and social media.

Marco Ottaviani

Head of Press Office and National President Spokesperson

Italian Red Cross

7.3 Promoting tolerance in diverse and divided societies

Over 20 years ago, the Movement acknowledged the pervasive problems of discrimination, intolerance and lack of respect for human diversity in many parts of the world through the 2003 Council of Delegates Resolution 9.¹¹ This called on all components of the Movement, within their respective mandates, to promote tolerance, non-discrimination and respect for diversity at the local, national and international levels. It outlined five critical areas for action:

- 1 Ensuring **openness and diversity** within the Movement itself, ensuring inclusive representation and internal practices that reflect humanitarian values.
- 2 Engaging externally to **build understanding and insights**, including building partnerships to foster dialogue and inclusion.
- 3 Promoting **public dialogue and advocacy** to foster social cohesion.

- 4 Strengthening **preparedness**, both proactive and reactive, for promoting tolerance and respect, with a focus on addressing the needs of marginalized and at-risk groups – supporting dialogue, **trust-building among communities** and coexistence.
- 5 **Learning from experience** and developing new initiatives both internally and with other organizations to identify best practices and foster collaboration in combating intolerance, discrimination and lack of respect for diversity.

Although adopted before the rise of today's digital information ecosystems, the resolution's relevance has only grown. What has changed – and now demands urgent attention – is the speed and scale at which harmful information spreads, rooted in the interplay between online and offline spaces. Digital platforms have become accelerants for harmful information – including intolerance, discrimination and lack of respect for diversity – that fuel division, mistrust and violence. These dynamics not only reinforce the very forms of intolerance the resolution sought to address, but they also directly undermine principled humanitarian action – distorting perceptions of neutrality, impeding access and putting humanitarian actors and communities at risk.

A renewed focus is therefore needed on how the Movement responds to these evolving challenges. Upholding the commitments of Resolution 9 today requires confronting the digital dimensions of intolerance and embedding the responses to harmful information into broader strategies for inclusion, protection and principled engagement – both online and offline. This is not only an internal imperative but also a necessary response to the broader information environment that shapes perceptions, access and the safety of humanitarian action.



Can you think of anything like yourself, you know, where there was a big news event and it kind of impacted how you felt about the country or the community? Well, the one that's ongoing just now is there's a lot of news stories about immigration and that kind of thing. I just think, personally, it makes me feel that I think we're a little bit backward and not accepting them. And a lot of the stories that I'm seeing, especially on X (Twitter), for example, is anti-immigration, which I can't quite figure out because it's not a stance that I take. So, you know, it's trying to send me down a particular line or trying to influence my thoughts on it, or brainwashing me into thinking another type of way."

Community member, UK

Contributor Insight 7.2

60 years of the fundamental principles: Urgent call to action

The 32nd International Conference of the Red Cross and Red Crescent Movement, marking the 50th anniversary of the Fundamental principles,¹² was held in 2015 under the overarching theme: Fundamental Principles in Action. The Conference reaffirmed the central role of the Movement's leadership in fostering and strengthening these principles. It emphasized that this responsibility should be exercised through four key areas of focus: a) strengthening the legal and statutory basis of National Societies and reinforcing their auxiliary role to public authorities; b) establishing and maintaining a sustained dialogue with public authorities and also with external partners and the broader public; c) investing in (humanitarian) education and providing contextualized practical training on the fundamental principles; and d) promoting the exchange of good practices and peer-to-peer learning among all the components of the Movement.

In 2024, the IFRC's Global Think Tank on Fundamental principles affirmed that while the Movement's leadership continues to bear responsibility for their promotion, the environment in which this responsibility is exercised has changed significantly since 2015. In recent years, principled humanitarian action has come under increasing pressure from growing social and political polarization, the politicization of aid, the spread of harmful information and the rise of dehumanizing rhetoric. These trends have contributed to an erosion of trust in humanitarian actors and their work – one that has been further intensified by the COVID-19 pandemic and other ongoing crises and conflicts.

In many contexts, neutral and impartial humanitarian actors face abuse, threats and even acts of violence, seriously undermining the Movement's ability to help people in need. Upholding respect for principled humanitarian action is key to enable National Societies, as mandated local responders, to carry out their humanitarian mission effectively – both within their own country and in international contexts. It is also critical for the IFRC and the ICRC to fulfil their respective mandates.

To address these challenges and concerns, the 2024 Council of Delegates adopted **Resolution 5, 'Call for respect and support for principled humanitarian action'**, and an accompanying Appeal to States at the 34th International Red Cross and Red Crescent Conference in 2024. The resolution urges all Movement components to intensify their focus on the fundamental principles and to take concrete action in the following areas:

Enhancing knowledge and learning to foster behavioural competencies aligned with the fundamental principles.

Strengthening National Societies' legal base and auxiliary role.

Reviewing and revising National Societies' Statutes (in line with the Statutes guidance).

Strengthening integrity, accountability and trust, including through initiatives such as the recently launched Community of Practice.

Advocating to external stakeholders on the importance of principled humanitarian action.

The fundamental principles are the moral foundation and compass of the Movement, forming the ethical backbone of principled humanitarian action. While the primary responsibility of the Movement's leadership to uphold and promote these principles has remained constant, the global context has changed dramatically. As the Movement marked in 2025 the 60th anniversary of the fundamental principles, all components are called on to take urgent action based on Resolution 5. This is not only a reaffirmation of shared principles, but a necessary step to ensure that principled humanitarian action can be preserved and protected in the years ahead.

Markus Mader

Strategic Advisor, Chair of Global Think
Tank on Fundamental Principles

IFRC, Geneva

7.4 Neutrality at a cost: The price of not taking sides

Fundamental principle

NEUTRALITY

In order to enjoy the confidence of all, the Movement may not take sides in hostilities or engage at any time in controversies of a political, racial, religious or ideological nature.

In today's volatile information environment, humanitarian organizations face growing challenges in communicating principled messaging. Even when grounded in facts and neutrality, such messaging is often drowned out, distorted or reframed as political or partial, especially on digital platforms. The polarizing comments and narratives that become attached to such statements, for example, in comment sections, leave little or no room for genuine engagement nor are they necessarily intended to. This dynamic shuts down the space for communication and engagement dialogue at precisely the time when it is most needed. This presents a serious perception gap: neutrality and impartiality are misunderstood, mistrusted or dismissed as detached or complicit. Meanwhile, silence or restraint in speaking out can be interpreted as moral indifference. Online dynamics, amplified by algorithms and public emotion, leave little room for nuance or principle-based explanations. As a result, neutrality itself can become controversial, ironically undermining its core purpose: enabling neutral humanitarian actors to "not take sides in hostilities or engage at any time in controversies of a political, racial, religious or ideological nature".

While neutrality is vital to preserving humanitarian space, it should not be invoked unthinkingly or at the expense of the most essential principle: humanity. Neutrality is often the bridge that allows the Red Cross and Red Crescent to operate where others cannot, yet at times it is applied without recognizing opportunities to raise a voice. Upholding both principles means navigating each decision carefully – balancing the

need to remain neutral with the responsibility to speak out when issues fundamentally affect humanitarian action, always weighing the consequences of each action.

A deliberate hierarchy of principles

Humanity is the highest imperative.

Impartiality defines how to serve that goal. Neutrality and independence are the means to achieve it. Voluntary service, unity and universality are the enablers that make it possible. Each principle flows from the one before, in a deliberate hierarchy that begins – and ends – with humanity.

Critics often define neutrality by what it is not – failing to recognize that it is not neutral toward suffering. Neutrality is neither passive nor indifferent: it is a deliberate, principled stance aimed at building trust, protecting access and preserving humanitarian space. Refraining from engaging in ‘wars of words’ that could compromise acceptance, access and trust is very hard. Upholding neutrality today requires more than internal discipline; it demands careful management of perception, timing and consistency, as well as strategic decisions about where and how to communicate. At times, that may include disabling or disengaging with comment sections or stepping back from online engagement entirely, while deepening efforts offline in trust-building with affected communities.

Neutrality is a strategic and tactical tool that aims to enable humanitarian organizations “to enjoy the confidence of all.” The challenge today is that the “all” has become global, more immediate, more fragmented and increasingly polarized. **Public discourse often imposes binary frames: for or against, us or them, victim or aggressor, innocent or guilty.** These dichotomies flatten complex realities and place humanitarian actors under pressure to take such sides. For example,

- **“For or against”**: neutrality is often equated with complicity and any failure to publicly condemn is portrayed as a moral failing.
- **“Us or them”**: the expectation to align with a particular group’s narrative, perspective or cause undermines impartiality and oversimplifies complex situations.
- **“Victim or aggressor”**: humanitarians may face pressure to assign blame, but their role is to provide principled assistance and protection based on need, not judgement.
- **Innocent or guilty**: providing aid to those perceived as ‘guilty’ can be misinterpreted as support for their actions, posing a challenge to the principle of neutrality.
- **“Suffering or staging”**: humanitarian needs are often framed as exaggerated, fabricated, politically motivated or evidence that those suffering are a security risk or political tool, casting doubt on humanitarian responses and framing humanitarian actors as naïve, complicit or politically driven.

- **“Deserving or undeserving”**: this framing erodes empathy and casts suspicion on both affected populations and humanitarian actors, politicizing even the act of alleviating suffering.

In such binary framings, humanitarian needs are not simply acknowledged – they are scrutinized for motive. These narratives leave little space for neutrality and impartiality, which require humanitarians to act based on need, not affiliation, identity or perceived moral standing. Yet, articulating this is increasingly difficult in an information environment that discourages nuance, favours emotive soundbites and overlooks the operational consequences of alienating key actors or compromising access. Humanitarian actors, by contrast, must protect access to all sides, safeguard staff and volunteers operating in harm’s way and communicate in ways that do not endanger the very people they seek to assist. In this context, how suffering is framed – and by whom – has never been more difficult – or more consequential. These binary framings erode empathy and cast suspicion on both affected populations and the principled humanitarian actors who serve them. They undermine impartiality by implying that suffering must be politically justified or morally validated. Within such polarized discourse, even the act of responding to human suffering risks becoming politicized.

Nowhere is this more evident than in the suffering of Palestinian civilians in Gaza and of Israeli hostages and their families since the 7 October 2023 attack, including during the period of captivity. Public discourse was dominated by binary framings: ‘either you stand with the hostages or with Gazan civilians’; ‘either you condemn Hamas or you ignore Israeli suffering’; ‘either you speak out or you are complicit’. This narrative architecture leaves little space for nuance, and even less for neutral, impartial humanitarian action. Humanitarian actors who draw attention to the devastating toll of bombardments and the lack of food and other essentials for civilians in Gaza were accused of ignoring the suffering of the hostages and Israeli trauma. Conversely, those who spoke out about the plight of hostages may be seen as aligning with one side of the conflict. In both cases, humanitarian discourse and action are quickly politicized, judged not by adherence to humanitarian principles or operational impact, but by perceived allegiance.

This binary framing strikes at the heart of neutrality: it denies space for holding multiple truths at once with regard to human suffering – for example, that both the loss, deprivation and suffering caused by the taking, captivity and death of hostages from Israel and the loss, deprivation, suffering and death of civilians in Gaza are serious humanitarian concerns (each raising obligations of parties to armed conflict under international law). For humanitarian organizations, the imperative is not to choose sides, but to protect the dignity and rights of all affected people, irrespective of political narratives. Yet doing so, particularly in the public information space, requires careful navigation often with an emphasis on maintaining humanitarian access. Statements must be weighed not only for their content, but for how they may be received, misinterpreted or instrumentalized. In such contexts, neutrality does not mean silence; rather, it calls for speaking with clarity, precision and consistency, grounded firmly in principle, even when those principles risk being misrepresented.

Contributor Insight 7.3



Clarifying neutrality: Countering online distorted narratives

In recent years, particularly since the escalation of the crises in the Middle East, the Italian Red Cross has witnessed the spread of harmful narratives online concerning the role of the International Red Cross and Red Crescent Movement and the principle of neutrality.

Much of the disinformation and misinformation has focused on the role of the ICRC as a neutral intermediary, particularly in its facilitation of hostage release operations in the Gaza Strip. These narratives¹³ have distorted public understanding of the principle of neutrality, and more broadly, of the roles of different Movement components during humanitarian crises. We have observed these distortions directly in user interactions on the Italian Red Cross social media platforms, as well as in articles published by prominent Italian media outlets.

In response, the Italian Red Cross organized a webinar in March 2025 on 'The Importance of Neutrality in the Current Humanitarian Landscape'. The event featured contributions from members of the Movement, academics and journalists. It was designed for all Italian Red Cross volunteers, with the aim of addressing and countering distorted narratives while deepening awareness of the operational dimensions of the principle of neutrality. To further clarify and support understanding, the Italian Red Cross also created a dedicated frequently asked questions (FAQs) page on its website addressing the principle of neutrality and the conflict in Gaza.

Tommaso Natoli

Head of Humanitarian Diplomacy
and International Law

Italian Red Cross

Giulia Marcucci

Humanitarian Diplomacy Officer

Italian Red Cross

Contributor Insight 7.4



Truth is the first casualty of war

Since the beginning of the armed conflict in Ukraine, the Ukrainian Red Cross Society has faced the life-threatening consequences of false information. In March 2022, a single picture ignited the war on information: a photo of a handshake between the former ICRC President and the Russian Minister of Foreign Affairs. It became the first – and tragically successful – attack on the neutrality of humanitarian aid in Ukraine. “The picture changed everything on the ground,” recalls Maksym Dotsenko, the Director General of the National Society, at a conference organized by the Ukrainian Red Cross Society. “Our staff stopped wearing the Red Cross vest because we were attacked at checkpoints. Everywhere across Ukraine, drivers in Red Cross vehicles were in danger.”

Strategically amplified false information framed the Red Cross as taking sides in the armed conflict, impacting trust, threatening staff safety and humanitarian work across the country. “Misinformation about the Red Cross and Red Crescent Movement affects everyone. People don’t distinguish between a National Society, the ICRC or the IFRC and false information takes only 15 minutes to spread everywhere on social media”, says Dotsenko.

Lack of access and trust followed. To counter this, the Ukrainian Red Cross Society launched a massive information campaign across all levels: “I had to prove to members of the parliament, to neighbours and friends that we are purely humanitarian and not influenced by Russia,” Dotsenko recalls. “I called every member of the government and explained our mandate. We basically knocked on every door.”

The hybrid war in the digital space forced the National Society to develop new communication competencies. Experts joined the team, a monitoring system was established to track public and social media discourse, and 24/7 readiness became the norm to comment, influence or clarify information, if necessary.

Every one of its 200+ branches developed a social media presence with tailored, localized content, while communications were woven into operational decision-making at headquarters. A media person always has a seat at the table when key decisions are made. Speed proved critical. “If you don’t provide answers, an information vacuum is formed and someone else will fill it,” says Dotsenko. Quick, clear and comprehensive communication across all available channels became life-saving. “Supporting people is most convincing. Every communications strategy fails if people don’t see the help. So, in our case, the work supported the communication and proved our mandate.”

The Ukrainian Red Cross Society also emphasized internal trust-building to reach volunteers and staff across the country with information on international humanitarian law and the fundamental principles. “The principle of neutrality is the most challenging one in a hybrid war with a former neighbour,” recalls Dotsenko. “We needed to build a solid basis of understanding for our principled approach with everyone. The fundamental principles provide us with the key to humanitarian access; this is the core. They also provide us with the responsibility to protect our volunteers and staff. The question is how we explain the principles, how we live them. You can’t hide behind humanitarian principles. Every day we must prove that we use them for achieving better access, better security.”

From the communication perspective, this demands openness and accessibility. “Every inquiry by the public needs to be answered and addressed,” stresses Lesia Oliinyk from the communications team. To ensure coherence and aligned messaging, they are using “a Q&A crisis matrix covering 80 topics, used by all 200 branches and volunteers when they are talking to people.”

The trust-building package also includes transparency and honesty as a rule. Oliinyk is convinced that “telling the truth in complex situations including about the finances of the National Society has helped to avoid communications crises.” Research on public perceptions of the Red Cross confirms progress: 96% of Ukrainians recognize the Red Cross emblem and half understand its mandate. To reach the rest, it has diversified its communication channels and set up its own media production room producing podcasts and analyses. “We have created a humanitarian media hub which is a website where we collect all news on the humanitarian sector”, explains Oliinyk. “This has developed into an important platform during the war; we provide humanitarian analyses and even government ministries publish content there, which is of course clearly marked as such.”

At the state level, the Ukrainian Red Cross Society has a constant dialogue with the government on humanitarian issues and partnered with the Ministry of Defence on international humanitarian law dissemination for the armed forces. Like the armed conflict itself, the war on information moves in waves. But every information attack on neutral humanitarian aid has also created the opportunity to reach out and communicate humanitarian values. “Our biggest success is when fake news is not even shared anymore because people are already informed and know more”, says Dotsenko.

In June 2025, the National Society convened a high-level conference to share its experiences and lessons from responding to the dual challenges of armed conflict and harmful information.

This text is a conversation between Gabriela Poller-Hartig of the Austrian Red Cross and Maksym Dotsenko of the Ukrainian Red Cross Society, reflecting on the National Society’s journey in safeguarding trust. The Austrian Red Cross supports the National Society in branch development.

Gabriela Poller-Hartig

Head of International Relations

Austrian Red Cross

Humanitarian organizations can speak out publicly without taking sides if this is done objectively, on the basis of principles and standards that apply equally to all. Neutrality does not mean silence in the face of suffering or violations. Criticism of humanitarian actors for not publicly condemning violations or suffering is often less about the principle of neutrality itself and more about judgement – specifically, a perceived over-reliance on confidential dialogue and persuasion or a reluctance to speak out due to a fear of backlash.¹⁴ In some cases, that criticism may be valid. In today’s polarized environment, this may require greater coordination among humanitarian actors in speaking out collectively and in carefully timing their communications.

Neutrality is also often burdened with connotations of distance and lack of empathy. Yet it does not imply an absence of feeling but rather **a means of conveying compassion for human suffering in a way that does not take sides**. Neutrality is also frequently misunderstood as requiring a perfect symmetry or equivalency in public statements or admonitions, even in situations of asymmetrical suffering or violations, particularly when there is mounting pressure for public condemnation. There are no easy answers and it takes courage to navigate this. Humanitarian organizations will continue to navigate neutrality differently depending on their mandates, modus operandi and the context. Avoiding the traps – especially in polarized environments – requires increased dialogue among principled humanitarian organizations, both to inform choices and to protect the space for action.¹⁵

Critiques of neutrality also often conflate all humanitarian organizations, overlooking that some are explicitly political while others are independent and guided by humanitarian principles. Failing to recognize these distinctions risks undermining trust in principled humanitarian organizations. Many humanitarian organizations address this by adhering to the *Code of Conduct for the International Red Cross and Red Crescent Movement and Non-Governmental Organizations (NGOs) in Disaster Relief* and other standards that reinforce the integrity of principled humanitarian action.

As humanitarian scholar Hugo Slim has observed: “the idea of political neutrality is not legally or semantically embedded in humanitarian aid, even if the last thirty years of international relations has tended to promote it as such. There are different ways to be humanitarian. Neutral humanitarians like the ICRC and the Red Cross and Red Crescent Movement ... and non-neutral humanitarians. People suffering in war usually need both, and, if the world moves into a new era of binary political commitments that set liberal democracy against illiberal authoritarianism, then we can expect many people’s political convictions to render their neutrality impossible, and we will see a surge in non-neutral humanitarians around the world. Non-neutral humanitarians will rightly be saving lives and protecting people within social movements that are politically positioned as democratic, authoritarian, socialist, Islamist, or green”.¹⁶

Moving forward, humanitarian space must be preserved for both neutral, impartial and independent humanitarian action and non-neutral humanitarianism – keeping people in need at the centre. Those in need should not be reduced to political arguments, but recognized as individuals entitled to respect for their humanity, rights, legal protection and agency, including humanitarian response based on need.

7.5 Patriotism and humanitarian principles

National Societies are deeply rooted in their countries. They carry the name of the nation, use the emblem with national identifiers and serve as auxiliaries to their public authorities in the humanitarian field. Patriotism can be a powerful force – motivating people to volunteer, mobilize resources and foster local trust.

Yet in humanitarian action, patriotism must always be balanced with the Fundamental principles, particularly humanity, neutrality, independence and impartiality. On one hand, joining a National Society to serve one’s community can rightly be seen as an act of patriotism. On the other hand, in situations of armed conflict, adhering to neutrality and impartiality may be perceived by some as undermining patriotic duty. This tension becomes especially acute when National Society staff or volunteers are present on both sides of a frontline: if they embrace nationalistic rhetoric or sentiment over the fundamental principles, the unity of the National Society – and indeed the cohesion of the Movement – risks splintering.

Upholding the fundamental principles – especially in armed conflicts – will never be easy, but this commitment is essential to protect humanitarian space and the Movement’s ability to serve people in need, regardless of nationality, political affiliation or other status. In a polarized information environment, adherence to these principles also acts as a powerful safeguard against harmful information – reinforcing trust, credibility and the collective identity of the Movement.

7.6 Independence as a condition for trust

Fundamental principle

INDEPENDENCE

The Movement is independent. The National Societies, while auxiliaries in the humanitarian services of their governments and subject to the laws of their respective countries, must always maintain their autonomy so that they may be able at all times to act in accordance with the principles of the Movement.

The principle of independence is closely linked to neutrality. It requires humanitarian organizations to remain detached from political, military, economic or religious influence and from the strategic agendas attached to them. Independence is, in many ways, the practical demonstration of neutrality – a condition that enables principled action in highly politicized environments. In today’s interconnected and interdependent world, however, independence is not always absolute but it must be safeguarded and demonstrated in context-specific ways. Doing so is essential both to ensure impartial action and to maintain neutrality.

Narratives that undermine perceived independence include:

- “They’re just an arm of the government.” Humanitarian organizations may be portrayed as extensions of national policy, particularly when their governance is closely linked to state actors or operations that rely on government funding or infrastructure.
- “They follow donor agendas.” This narrative implies that aid priorities are driven by political or strategic interests of donors rather than humanitarian needs, casting doubt on organizational autonomy.
- “They cooperate with the military, they can’t be independent.” In complex emergencies or civil–military settings, any coordination with armed actors can be interpreted as complicity or co-optation.
- “They changed their position after pressure.” When an organization revises public positions, especially in response to criticism, it may be interpreted as the result of external influence rather than principled adaptation.

7.7 Impartiality as a compass

Fundamental principle

IMPARTIALITY

It makes no discrimination as to nationality, race, religious beliefs, class or political opinions. It endeavours to relieve the suffering of individuals, being guided solely by their needs, and to give priority to the most urgent cases of distress.

The principle of impartiality, like neutrality, is frequently misunderstood and the two are often conflated, yet they are distinct. Impartiality requires that aid be provided solely according to need, with priority given to endeavours to relieve the suffering of individuals, putting the “most urgent cases of distress” first.

Its ethical foundation lies in the equal worth and rights of all human beings. In practice, impartiality translates into two operational obligations: aid must be distributed without discrimination and relief must be proportional to the level of need – the greater the need, the greater the response.

Barriers to impartiality arise for example when authorities or non-state actors obstruct assistance to populations under the control of another party – whether out of ideological conviction, political calculation or fear that it may benefit the adversary. Famines have been deliberately caused through such obstructions. When examining impartiality in the context of harmful information, several narratives emerge that undermine it. Whether rooted in harmful information or misunderstanding, these challenge the understanding of the core principle of providing aid solely based on need. For example:

- “They are helping the enemy.” Humanitarian actors are accused of favouring one side, especially when assisting populations under the control of an adversary. This delegitimizes aid and is often used to justify blocking access.
- “They only show one side of the suffering.” When humanitarian communication focuses on particular groups or geographies they can be interpreted or distorted as evidence of bias, fuelling claims of partiality. At the same time, efforts to appear “balanced” risk creating false equivalence between situations of suffering or responsibility, further complicating perceptions of neutrality and credibility.
- “Donors dictate who gets helped.” The perception that humanitarian aid is driven by political or donor interests or priorities rather than humanitarian needs feeds scepticism about whether aid is truly impartial and needs based.
- “They’re part of the West’s agenda.” In polarized or post-colonial contexts, humanitarian organizations can be portrayed as aligned with foreign political or ideological goals.

- **“If they were impartial, why aren’t they helping us?”** Communities not receiving aid may assume bias, especially in the absence of transparent needs assessments or clear communication about operational decisions and constraints.

To uphold the principle of impartiality in the face of harmful information, humanitarian organizations must ensure proactive and sustained measures, including anticipating instrumentalized narratives that exploit perceptions of bias or selective aid, and preparing clear, principled communication lines in advance (including prebunking). Transparent communication is essential, particularly around how needs are assessed and how decisions about aid allocation are made. Internally, building digital and media literacy among staff and volunteers is critical to enabling early detection and effective response to accusations of partiality. At the same time, organizations must actively monitor information ecosystems to identify and counter distorted portrayals before they gain traction and erode trust. Impartiality requires that assistance be provided solely on the basis of need, without discrimination or political consideration – a principle that safeguards humanity itself and ensures that no one is left behind because of who they are or where they stand.

7.8 **Organizational integrity: Internal alignment with principles**

Upholding humanitarian principles begins with internal organizational coherence and credibility. Transparency, accountability and consistency are essential anchors that reinforce trust and legitimacy – both within an organization and in the eyes of the communities it serves. At the heart of this is integrity: the alignment of words and actions with humanitarian values, even when doing so is inconvenient or costly. Integrity is not just about avoiding misconduct: it is about making principled choices, maintaining ethical standards under pressure and resisting the temptation to compromise values for short-term gain.

Leadership plays a central role in modelling integrity and principled behaviour, setting the tone for inclusive, respectful and values-driven practices across all levels of the organization. When integrity is embedded in decision-making, communication and daily operations, it creates a culture of trust – one that is resilient in the face of scrutiny, harmful information and political pressure.

Contributor Insight 7.5



Rebuilding trust: Lessons from an institutional crisis and recovery at the Uganda Red Cross Society

In 2013, the Uganda Red Cross Society experienced a major leadership crisis that became the most significant reputational challenge in its history. The incident demonstrated how quickly public trust in a humanitarian organization can be undermined when information – whether accurate, incomplete or exaggerated – spreads in the public domain, and how much effort is required to rebuild it.

The crisis emerged following allegations of irregularities related to the clearance of a truck of aid goods at the Malaba border. As media scrutiny intensified, concerns about governance of the National Society were amplified through national media and public speculation grew rapidly. The unfolding situation created widespread perceptions of leadership misconduct. As the matter progressed, the consequences escalated, eventually resulting in the arrest of the Secretary General at that time. This series of events severely eroded public trust in the institution.

Internally, the impact was profound. Staff and volunteers were demoralized and faced difficult questions from their communities and even their own families. Externally, partners raised serious concerns over governance and accountability, with some suspending or withdrawing funding commitments. National-level operations were disrupted, several staff members resigned and some volunteers began to disengage. The credibility of the entire organization was at risk.

Volunteers and staff working closest to communities were among the most affected. In several areas, they encountered hesitation and suspicion from community members who were uncertain about the institution's integrity. Media outlets reduced or stopped covering National Society activities, wary of providing visibility to an organization perceived to be undergoing governance challenges. Restoring confidence required deliberate, coordinated and sustained effort across all levels.

With support from the IFRC, the Uganda Red Cross Society leadership initiated a series of decisive reforms. Key staff positions were restructured and official statements were issued to clarify the situation and reassure both partners and the public. The National Society Governing Board initiated internal reviews and introduced interim governance measures to safeguard operational integrity. New management and Board members were appointed to guide the institution through a comprehensive recovery and change process. Over time, the National Society strengthened internal accountability mechanisms, improved financial oversight and introduced rigorous leadership vetting and performance systems. It also underwent an Organizational Capacity Assessment that resulted in the development of 11 recovery pillars that guided institutional rebuilding.

A key lesson from this experience is the importance of strong crisis communication, transparent governance and rapid institutional response when trust is at risk. In the years that followed, the Uganda Red Cross Society prioritized rebuilding public trust by implementing stricter accountability systems and maintaining close engagement with

partners, communities and other stakeholders. The crisis underscored how leadership-related reputational issues can quickly erode trust and how recovery requires transparency, structural reform and sustained commitment.

For a humanitarian institution whose effectiveness depends heavily on public goodwill, trust is essential and cannot be compromised. Staff, volunteers and governance bodies receive regular training on their roles, responsibilities and the policies that reinforce institutional accountability. Today, the Uganda Red Cross Society operates with strengthened systems at all levels, including an active Integrity and Compliance Committee, a revised Constitution and robust policies that uphold a zero-tolerance approach to corruption. The National Society's communications and public relations functions play an important role in building public trust and managing institutional relationships, which in turn has a significant impact on resource mobilization.

Irene Nakasiita

Director, Communications, Resource Mobilization and Partnerships

Uganda Red Cross Society

Contributor Insight 7.6



Q&A on principled action with the Russian Red Cross: Responding to a significant reputational challenge

1 What happened? What was the impact on trust and operations?

In February 2024, intense media scrutiny by foreign media led to accusations that the Russian Red Cross had violated the fundamental principles. Articles containing false allegations were published by media outlets across 11 European countries.

As a result, the IFRC had to evaluate the activities of the National Society, particularly its adherence to the fundamental principles.

2 Who was most affected? How did you respond?

Accusations against the Russian Red Cross – Russia's oldest humanitarian organization – threatened its resources, international reputation and ability to serve beneficiaries. The most affected were displaced people from Ukraine in Russia, the largest group receiving Russian Red Cross support.

After reviewing information provided by the National Society, the IFRC concluded that the Russian Red Cross' humanitarian services aligned with the principles and did not identify any violations. It identified some general IFRC policy standards that could be addressed and issued recommendations to strengthen these. The National Society adopted these reforms and adhered to these recommendations.

3 What was learned and changed?

The Russian Red Cross adopted reforms to safeguard its operations, reputation and reinforce accountability:

- a policy on good partnership to guide collaboration with others
- a policy on child safeguarding to conform to IFRC standards
- an updated Code of Ethics and complaint procedures, including protection for whistleblowers, victims and the option to involve external experts in investigations
- standardized training on partnerships and engagement for staff at headquarters and in regional branches.

These changes improved internal consultation and decision-making, with regional branches now proactively requesting guidance before launching partnerships or participating in public events.

This period also prompted reflection on applying the fundamental principles in times of crisis. The National Society had to carefully balance its auxiliary role to the government in the humanitarian field. This situation has taught us to think more and consider the consequences of each word and action from different perspectives. We had to explain that in times of crisis, activities such as providing assistance to families of military personnel and combat veterans and teaching first-aid skills to mobilized and military personnel are part of the Red Cross mandate.

Lessons learned: During crises, the Russian Red Cross recognizes we are under additional scrutiny and bias – so every word and action carries weight. It recognizes the need to anticipate reputational risks, communicate its mandate more clearly and adapt its operations to ensure that humanitarian principles are understood and upheld at all times.

Anastasia Teneta

Head of International Cooperation Department

Russian Red Cross

As information ecosystems continue to evolve, building digital literacy among staff and volunteers is increasingly vital. This includes the ability to identify and respond to harmful information, understand the dynamics of online narratives and engage responsibly on social media and in face-to-face engagements. When volunteers are equipped to navigate these spaces with confidence and integrity, principled humanitarian action is more likely to be consistently upheld – whether in person or online.

Maintaining volunteer motivation and trust requires clear, principled communication that resonates with shared humanitarian values. Volunteers are often the most visible representatives of an organization in their communities, thus how they communicate and behave directly shapes public perceptions of legitimacy and trust. Regular, open dialogue – grounded in the organization's mission and principles – reinforces their sense of purpose and belonging, particularly in complex or politicized environments.

Training, guidance and supportive leadership are crucial to ensure that all forms of engagement reflect neutrality, impartiality, independence and humanity, while also protecting staff and volunteers from becoming unintentional amplifiers of harmful information or targets of such campaigns. At the same time, there is an important duty of care when staff and volunteers are targeted by, or otherwise affected by, harmful information.

Contributor Insight 7.7



Q&A with the Honduran Red Cross: Fragile bonds between communities, volunteers and authorities

1 How does harmful information affect local relationships?

Harmful information represents a significant barrier that distorts community perceptions of humanitarian work, directly affecting relationships with local communities, volunteers and authorities. In the case of the Honduran Red Cross, we have encountered situations where rumours or misinformation have raised doubts about our impartiality and neutrality. When false information circulates linking our work to partisan interests or misinterpreting beneficiary selection, community trust is undermined and access to vulnerable areas becomes more difficult.

In response to these challenges, the Honduran Red Cross has strengthened its community communication approach by launching targeted campaigns through social and local media. These initiatives aim to reaffirm our humanitarian mandate, the fundamental principles and the neutral, impartial and independent nature of our work. Strengthening the Community Engagement and Accountability mechanism has also been essential for rebuilding trust with communities and maintaining the engagement of both volunteers and institutional partners.

2 What types of support, collaboration or guidance would help your organization address harmful information more effectively?

It is important to strengthen operational communication in communities and promote dialogue to clearly explain the scope and purpose of our actions. Greater investment is needed in communication and awareness-raising strategies. Local communities should be empowered to act in countering misinformation, reinforcing their role in protection and social cohesion. Active participation in local working groups is essential to ensure our actions are aligned with community needs.

3 What recommendations would you offer humanitarian organizations to address harmful information more effectively?

- Facilitate internal reflection spaces and training on strategic and operational communication within the Movement to share best practices for managing harmful information and humanitarian diplomacy.

- Involve community leaders, volunteers and trusted local figures as key actors in validating and disseminating reliable information, leveraging their legitimacy and close connection to the population.
- Implement ongoing and sustained communication campaigns – not only reactive ones – that highlight local impact stories and authentic community testimonies.

Honduran Red Cross

Contributor Insight 7.8



Volunteer voices: Volunteers as trust brokers

Across the Americas, volunteers face not only the visible effects of disasters but also the hidden crises triggered by harmful information. Misinformation about neutrality, conspiracy theories around health interventions and confusion about affiliations with governments have all endangered their safety. In extreme cases, volunteers have been harassed or excluded from communities, restricting access to the people most in need. In violent areas, some volunteers avoid entire zones for their own safety. This is not just a humanitarian access issue: it is about protection.

Volunteers often act as ‘trust brokers,’ mediating between communities and humanitarian organizations. But when public trust breaks down, this role becomes difficult and dangerous. Disinformation undermines morale. Some volunteers describe feeling treated as ‘free labour’ when excluded from recognition or decision-making structures. Many leave not from fatigue, but from feeling unsafe or undervalued. Recognition, protection and participation are therefore critical. In the 2023 annual Volunteer Management Survey (used for IFRC Americas baseline studies), 84% of National Societies reported having a recognition system, while 57% had inclusion mechanisms at all levels and 30% at some levels. Just over two-thirds (68%) reported having a Solidarity Fund, and 95% provide psychosocial support after security incidents. Most also equip their volunteers adequately. When these mechanisms are missing, the operational consequences are clear: reduced presence in high-risk areas, reluctance to engage on sensitive issues (such as vaccination), growing volunteer turnover and weakened community engagement. Addressing these gaps requires not only stronger protection and training, but also investment in strategic communication, local advocacy and organizational resilience.

Volunteers must be prepared with digital literacy, crisis communication and conflict-sensitive engagement skills. Encouragingly, the 2023 survey shows all National Societies offer continuous learning, either through the Stay Safe programme or their own initiatives. Yet inclusion remains uneven: while 89% of National Societies integrate volunteering into their organizational plans, only 63% have volunteer-specific strategies. Around 70% conduct research on volunteering, signalling a growing shift toward evidence-based approaches. Volunteers are more than service providers; they are agents of change and the Movement’s human face. When trust falters, their ability to serve is compromised. Rebuilding that trust

begins by listening to their voices, valuing their insights and equipping them to navigate today's complex information environment.

Andrés Morales

Thematic Lead, Volunteering Development and
Youth Engagement, Americas Regional Office

IFRC

7.9 **A framework for applying humanitarian principles in the information age**

Building on the 1979 Pictet commentary on the fundamental principles and the 2013 *World Disasters Report* matrix on the seven fundamental principles, the following updated framework highlights how these principles are upheld – or challenged – in the context of harmful information. It reflects the growing intersection between harmful information and humanitarian values in today's digital and information-saturated environment. The framework acknowledges both the evolving threats and opportunities posed by today's information landscape. It underscores the imperative for humanitarian actors not only to uphold the four humanitarian principles – and within the Movement the seven fundamental principles – but also to apply them in adaptive and context-specific ways that respond to the specific challenges posed by harmful information.

Table 7.1 Fundamental principles in the context of harmful information

Fundamental principle	Components	Contemporary relevance in harmful information age	Underlying humanitarian values
Humanity	<p>Alleviate and prevent suffering</p> <p>Protect life and dignity</p> <p>Respect and protect individuals</p>	<p>Harmful information can endanger lives and increase suffering by undermining trust in humanitarian action, humanitarian actors and emergency information.</p> <p>Combating it requires proactive, people-centred communication and engagement that protects dignity and well-being.</p>	<p>Human dignity</p> <p>Compassion</p> <p>Well-being</p> <p>Solidarity</p> <p>Mutual understanding</p> <p>Respect</p> <p>Sympathy</p> <p>Cooperation</p>
Impartiality	<p>Non-discrimination</p> <p>Needs-based action, proportional response to degree of suffering</p> <p>Prioritized on basis of urgency</p> <p>No individual action or decision on the basis of prejudice or personal preference</p>	<p>Harmful information can skew perceptions of who deserves humanitarian response, fuelling discrimination.</p> <p>May shape funding and media attention, potentially sidelining less visible crises. Countering this helps reinforce equal access and fairness.</p>	<p>Equality of rights</p> <p>Respect for diversity</p> <p>Objectivity</p> <p>Acceptance</p>

Fundamental principle	Components	Contemporary relevance in harmful information age	Underlying humanitarian values
Neutrality	<p>Not taking sides in armed conflicts</p> <p>No engagement in controversies of a political, racial, religious or ideological nature</p>	<p>Harmful information campaigns can portray humanitarian organizations as partisan or politically biased.</p> <p>Strategic silence or lack of response may be misinterpreted.</p> <p>Neutrality requires active narrative management and trusted community engagement.</p>	<p>Trust and confidence</p> <p>Self-control</p> <p>Discipline</p> <p>Freedom of action</p> <p>Objectivity</p>
Independence	<p>Autonomy from political, economic, social, religious, racial or ideological influence</p> <p>Autonomy from donor influence</p> <p>Auxiliary role to public authorities for National Red Cross and Red Crescent Societies but with operational freedom and independence</p> <p>Autonomy to act in accordance with the fundamental principles</p> <p>Autonomy to act in accordance with the humanitarian principles</p>	<p>Accusations of being too close to governments or too detached from communities are common in harmful narratives.</p> <p>Maintaining perceived and actual independence is essential for credibility.</p>	<p>Accountability</p> <p>Operational integrity</p> <p>Cooperation</p> <p>Freedom of action</p>

Fundamental principle	Components	Contemporary relevance in harmful information age	Underlying humanitarian values
Voluntary service	<p>Freely accepted commitment</p> <p>No desire for material gain</p> <p>Selflessness</p>	<p>Volunteers may be targeted by harmful information or face reputational risks due to their association with organizations.</p> <p>Support mechanisms to cope with and respond to harmful information must be built into digital volunteer engagement.</p>	<p>Altruism</p> <p>Spirit of service</p> <p>Solidarity</p> <p>Initiative</p> <p>Discipline</p>
Unity	<p>One National Society per country</p> <p>Open to all</p> <p>Active across the whole country</p>	<p>Harmful narratives may amplify internal divisions or portray humanitarian action as favouring certain groups, undermining collective identity and coherence in the Movement.</p> <p>Ensuring consistent messaging, transparent communication and inclusive engagement across all components helps preserve unity and reinforce shared purpose.</p>	<p>Cohesion</p> <p>Inclusivity</p> <p>Diversity</p> <p>Trust</p> <p>Common purpose</p> <p>Acceptance</p>
Universality	<p>Equal status of National Societies</p> <p>Solidarity and mutual support worldwide</p> <p>Universal vocation</p>	<p>Harmful information spreads rapidly across borders, shaping perceptions and reputations far beyond where a crisis occurs.</p> <p>Upholding the principle of universality requires coordinated responses, shared learning and solidarity across the Movement to maintain global trust and integrity.</p>	<p>International cooperation</p> <p>Mutual assistance</p> <p>Global solidarity</p> <p>Openness</p>

7.9.1

Decision-making framework: Responding to harmful information while upholding humanitarian principles

To respond effectively to the challenge of harmful information, organizations must be equipped to act in ways that minimize harm while upholding the humanitarian principles of humanity, neutrality, impartiality and independence. This decision-making framework offers a structured approach to guide analysis, reflection and principled, context-specific action.

Step 1: Pre-crisis preparedness

- Train teams in proactive monitoring, including rumour tracking, perception tracking and narrative mapping.
- Build trusted local partnerships before a crisis to enable credible community-led amplification and establish feedback loops.
- Identify trusted voices and credible messengers within communities, as well as potential threat actors or sources of harmful information, and map their influence.
- Develop internal protocols for responding to harmful information anchored in humanitarian principles, 'do no harm', accountability and context sensitivity, while recognizing the importance of timeliness.
- Ensure governance, leadership, staff and volunteers at international, national, district and branch levels have the knowledge, skills and tools to identify and promote principled humanitarian action.
- Identify priority contexts where serious risks or gaps in principled humanitarian action exist and integrate this awareness into preparedness planning.

Step 2: Assess the nature and impact of harmful information

Use guiding questions to understand the type of harmful information and its potential consequences (impacts and harms), for example:

- Is the information factually false, misleading or harmful by implication?
- Who is the primary audience (e.g., affected communities, state actors, donors, media)?
- Who are the key voices shaping the narrative – trusted messengers or potential threat actors – and what influence do they hold?
- What is the likely or observable impact (e.g., reduced access, threats to staff, rejection of aid, reputational damage)?

Output: Categorize the risk and/or threat level (e.g., low-risk rumour, operational disruption, challenge to access, legitimacy or safety).

Step 3: Evaluate the ethical and operational stakes

Assess the key tensions between action and inaction. For example:

- Would responding compromise neutrality or risk being perceived as politically motivated? How can this be mitigated?
- Would remaining silent undermine humanity or allow harm to escalate?
- What are the potential safety implications for staff, volunteers and communities?
- How might either response or lack thereof affect trust with communities, partners and stakeholders?

Table 7.2 Example matrix to map trade-offs between response and potential harm

Response	Principles it supports	Principles it risks being perceived as violating	Potential harm
Public response	Humanity Independence	Neutrality Impartiality	Political backlash Misperception Escalation of tensions Reputational damage
Silence or bilateral engagement	Neutrality	Humanity	Mistrust Disempowerment Perception of complicity
Community-led response	All principles (if done well)	Impartiality	Uneven messaging Potential fragmentation of trust

Step 4: Consider the messenger and the method

Guiding questions:

- Who should respond and how?
- Should the organization respond directly or should trusted local actors take the lead?
- What channel(s) or format(s) are most appropriate (e.g., social media, bilateral/in-person dialogue, third-party endorsement)?
- Could the response unintentionally amplify the harmful narrative?
- Would a third-party or coalition message reduce perceptions of partiality or strengthen the response?

Tactical options: Quiet correction through community liaisons; co-produced content with local influencers or respected community voices; fact-based public statements using neutral, non-confrontational language; silence (accompanied by internal documentation and monitoring) recognizing this may backfire.

Step 5: Align with the context and dynamics

Ensure responses are sensitive to the broader sociopolitical environment:

- What are the political sensitivities or risks of backlash in the context?
- Are there power asymmetries (e.g., post-colonial legacies, racial, gender-based dynamics) influencing how responses are perceived?
- Could engagement in information correction reinforce perceptions of organizational arrogance?

Principled reflection:

- Are the same standards and principles being applied equally based on need across different groups or communities?
- Is the response informed by local knowledge and trusted relationships in the community?

Step 6: Decide, act, document and reassess

Make an informed and principled decision:

- Choose the course of action that causes the least harm and best aligns with humanitarian principles.
- Act with consistency, humility and clarity.

- Document the decision-making process for accountability and institutional learning.
- Reassess regularly, especially if new actors emerge, narratives shift or risks escalate.
- Measure impact on trust.

To effectively counter harmful information, humanitarian actors must shift from a reactive to a proactive stance and also from defensive to offensive approaches, if appropriate. This shift begins by integrating information preparedness into operational preparedness – anticipating harmful narratives, pre-positioning factual content and empowering field teams and volunteers to communicate swiftly when needed. A key component is the decentralization of trusted voices: building the capacity of local volunteers and other actors and communities to speak first, faster and more credibly. These actors are often more trusted than international organizations and better positioned to respond in real time. At the same time, humanitarian organizations must define what constitutes ‘principled speed’ – ensuring that urgency and timeliness do not come at the expense of neutrality, accuracy or trust.

Establishing clear thresholds for when, where and how to engage can help ensure that rapid responses remain both defensible and ethical. Finally, lasting impact requires strengthening the understanding of principled humanitarian action in information ecosystems by engaging proactively with digital platforms, technology actors, media and civil society to shape the environments in which information circulates – rather than reacting only after harm has already occurred.

7.9.2

From principles to practical measures – community engagement and accountability

Upholding principled humanitarian action requires shaping how humanitarian organizations engage with communities, particularly in environments affected by harmful information. Drawing on the IFRC’s community engagement and accountability approach, the following measures are essential to translating principles into meaningful practice:

- **Contextualizing approaches:** Develop strategies to reflect local dynamics and information landscapes.
- **Ensuring inclusion, including women, youth and older people:** Involve local actors from the outset, invite participation and create safe spaces for dialogue in all stages of engaging with communities.
- **Prioritizing communication and transparency:** Plan how information will be shared. Share information regularly in accessible formats and languages, using channels trusted by the community.
- **Designing responsive feedback systems:** Provide timely, safe and confidential ways for people to share concerns or report abuse, including mechanisms for handling sensitive issues separately and confidentially. These should leverage existing local feedback mechanisms and be timely.

- **Embedding accountability:** Build community participation – including with the most marginalized and socially excluded groups – into programme design, monitoring and evaluation. Learn from feedback and allow space to process it and adjust policy and programming accordingly.
- **Reinforcing community engagement and accountability:** Make this a non-negotiable part of recruitment, performance, reporting and partnership standards.
- **Sharing information openly:** Affected populations should have access to clear details on how decisions are made, what resources are available and how aid is being delivered, including partnership arrangements, response actions, targeting criteria, funding levels and other issues that affect them.

7.10 Influencers: Connectors and dividers

Factors that influence communities can act as both connectors and dividers and must be carefully analysed as part of any engagement strategy. Understanding how these forces interact – and how community engagement can either reinforce or mitigate them – is essential to designing and adapting approaches that are context sensitive and ethically sound.¹⁷

A connector is an individual, group or structure that helps build bridges across societal divisions. Connectors contribute to local preparedness and response efforts by fostering relationships, promoting trust and generating positive interactions between different parts of a community. They enhance social cohesion and resilience and can amplify the reach and legitimacy of humanitarian efforts.

In contrast, a divider has a vested interest in preserving tension, exclusion or conflict. Dividers exploit or exacerbate societal fractures – such as those based on ethnicity, political affiliation, gender or religion – resulting in mistrust, harmful information and polarization. Their actions can pose serious risks to staff, volunteers and the integrity of humanitarian action.

Community engagement strategies should be designed – and continuously re-evaluated – based on a clear understanding of how connectors and dividers¹⁸ shape perceptions and behaviours. This is critical to applying the ‘do no harm’ principle in practice (Chapter 4: section 4.3 on page 150). Engagement that inadvertently strengthens dividers or bypasses connectors can undermine humanitarian objectives and cause harm to the very populations it seeks to support.

The connectors and dividers framing is especially relevant today in the context of influencers – whether online personalities, community leaders, media figures or local authority figures – because of their significant power to shape public perception, amplify narratives and influence behaviour at scale. In today’s hyper-connected information environment, influencers can function as either connectors or dividers, depending on their intent, messaging and the trust they command.

Influencers who promote social cohesion, accurate information, inclusion and empathy can serve as powerful connectors. When aligned with humanitarian values, they can:

- help bridge divides between humanitarian actors and communities
- dispel rumours and misinformation
- encourage trust in organizations and public health measures
- support local peacebuilding and solidarity efforts
- amplify marginalized voices and foster dialogue across communities.

For example: a local TikTok creator addressing harmful information about aid eligibility, a respected community leader countering hate speech online or a youth influencer explaining humanitarian principles in relatable terms.

Conversely, influencers who spread harmful narratives, reinforce stereotypes or exploit community tensions can act as dividers. They may:

- fuel mistrust toward aid agencies or certain groups
- amplify harmful information, xenophobia or conspiracy theories
- increase polarization or fear, especially in conflict-affected settings
- undermine community cohesion or cooperation with humanitarian efforts
- pose security risks to humanitarian workers or volunteers.

For example, a social media figure spreading conspiracy theories about humanitarian motives or a local community leader or official framing aid as biased or politically motivated.

In the age of harmful information, principled humanitarian action depends not just on what is done, but on how transparently, inclusively and accountably it is carried out. In contested spaces where narratives clash and principles are often misunderstood, trust must be earned through consistent demonstration of humanitarian values and effectiveness in practice.



Transforming humanity in Bolivia: Lessons from the IFRC-Disaster Response Emergency Fund Flood Project

At the beginning of the project activities, needs assessments were conducted in coordination with local authorities and community leaders. These assessments identified priority sectors for the distribution of cash and voucher assistance, kits, water purification tablets and jerry cans, in line with the humanitarian criteria established for the project's Disaster Response Emergency Fund (DREF) Action Plan. During project implementation, it became evident that community leaders had not provided sufficient information to the targeted population. Some residents reported that only families 'aligned' with certain neighbourhood organizations – referred to as affiliates – were participating.

In addition, part of the local population receiving assistance was unaware of the work of the Bolivian Red Cross. Lacking knowledge of our mission, principles and working methods, some community members suspected that the Bolivian Red Cross surveys were being conducted for political or economic gain – particularly given the climate of widespread speculation in the country. As a result, some individuals refused to take part in the surveys or to accept the assistance offered. During the cash and voucher assistance distribution process, a call centre was established. However, some people contacted to receive their codes feared the calls might be part of a telephone scam – an occurrence that is relatively common in the country.

Luis Juaniquina Navia

National DREF Coordinator

Bolivian Red Cross

7.11 Reaffirming humanitarian principles in the age of digital tools

The humanitarian principles have long served as an essential compass in navigating complex operational and political environments. Far from being static ideals, they provide a dynamic framework that has guided humanitarian organizations in their responses across diverse crises and contexts. In the digital age – particularly amid the rise of harmful information – these principles are more relevant than ever. The sector's credibility depends on effectiveness, transparency, clarity and visible adherence to humanitarian principles, underpinned by accountability to affected people and coherence across actors. Humanitarian actors must be able to explain the objectivity with which needs and responses are assessed and prioritized, as well as the proportionality of the response to existing needs.

Today, digital platforms are increasingly creating distance between humanitarians and the people they serve. In some cases, they are replacing direct human engagement altogether, which is especially troubling in settings where empathy, understanding and trust are essential. This tension is heightened with the use of AI which, by design, seeks to replicate or replace human judgement. As technology mediates more interactions, the risk increases that the fundamentally human connection of humanitarian work – listening, understanding and respecting dignity in complex realities – may diminish. **Empathy cannot be automated and should remain human, not machine driven.** As AI evolves to emulate the human, the human proximity must remain and not be sidelined. Dignity – central to the principle of humanity – is lived, subjective and recognized only through human presence and interaction. It cannot be predefined or interpreted by algorithms.¹⁹

This makes the responsible and principled use of digital tools all the more critical. Drawing on a mapping in the 2013 *World Disasters Report*, the benefits and risks of digital technology to the fundamental principles can be adapted to the specific challenges posed by harmful information. This would provide a practical framework for:

- **planning** and designing programmes aligned with the fundamental principles
- **advocacy** on principled digital transformation
- **policy-making and risk analysis** for reliable information
- **training** humanitarian actors on how to use digital tools to strengthen trust, reduce exposure to biased narratives and protect affected populations.

This provides insights for how humanitarian organizations can leverage digital tools to better identify needs, empower communities and enhance accountability – while actively mitigating the risks of harmful information.

Concluding remarks: Principled action requires more than declarations

Declaring adherence to the fundamental and humanitarian principles is not enough. Access to affected populations and acceptance of humanitarian actors must be continually earned. They depend on humility, principled behaviour, operational effectiveness and impact, and timely, transparent engagement and communication. Words must align with actions. Without this alignment, safe and sustained access – essential for principled humanitarian action – becomes increasingly difficult to secure.

Upholding humanitarian principles today means more than silent adherence. It requires actively promoting, explaining and embodying them in every context – including the digital sphere, where harmful information and politicized narratives quickly erode trust.

In an era of accelerating technology, human responses such as empathy, respect and moral restraint are more critical than ever. As philosopher Jonathan Glover observed, these moral resources help people exercise self-restraint, respect the dignity of others

and care for their suffering and well-being. He described them as “the tendency to respond to people with certain kinds of respect – as members of our community, as human beings” and as “sympathy – caring about the miseries and the happiness of others, and perhaps feeling a degree of identification with them.”²⁰

Today, harmful information is not a communication challenge; it is a threat to humanitarian action and requires a whole-of-society and whole-of-organization approach. Where humanitarian communicators once focused on facilitating information sharing with affected communities and on engaging donors, they are now increasingly tasked with detecting, responding to and trying to mitigate harmful content. Yet many organizations remain underprepared for the scale and sophistication of today’s information threats.

Technology can undoubtedly streamline humanitarian operations, but it also raises difficult questions about whether it risks dehumanizing humanitarian action – replacing listening with automation or judgement with algorithmic logic. Writing in 1999, long before today’s digital transformation, Glover presciently concluded in his book *Humanity*:



It is too late to stop the technology. It is to the psychology that we should now turn.”²¹

Glover is saying that the danger lies less in the machines themselves than in how human psychology responds to the power they give us. The survival of humanity depends not on stopping technology, but on understanding and strengthening the moral psychology that restrains its misuse. Just as Glover emphasizes that moral restraint and psychology are central once technology cannot be stopped, the Movement Appeal to States underscores that humanitarian principles – humanity, neutrality, impartiality and independence – serve as psychological and ethical safeguards in complex and high-pressure environments. In a polarized information landscape, where narratives can manipulate perception and inflame divisions, adherence to these principles acts as a moral and operational compass, guiding actors to resist pressures, make impartial decisions, and maintain trust. Essentially, **while we cannot control the speed or reach of harmful information (the ‘technology’), we can rely on principled humanitarian action (the ‘psychology’)** to ensure that responses remain ethical, impartial and focused on human need. In today’s landscape, where trust is fragile and harmful information can undo principled efforts in moments, Glover’s call is more urgent than ever.

The Movement has enormous capacity to act as a force amplifier for engagement through its staff and volunteers. National Societies seek to serve as trusted influencers who, by modelling the same positive behaviours online as they do offline in their communities, can help build healthier societies and shape norms of participation and response in the digital space.

Asks, aims and recommendations

Asks

Safeguard the fundamental and humanitarian principles – humanity, neutrality, impartiality and independence – across digital and offline spaces through collective action that protects access, safety and trust in humanitarian action, including the security of staff and populations in need. This includes respecting the legal and auxiliary roles of National Societies and ensuring that restrictive laws, sanctions or counter-terrorism measures do not undermine principled humanitarian action.

Aims

Defend humanitarian principles from erosion by harmful information or politicized narratives and societal or political polarization.

Protect humanitarian access, safety and security by visibly upholding impartiality and neutrality and by advocating for humanitarian exemptions in legal and regulatory frameworks.

Reinforce community trust by co-creating narratives that reflect humanitarian principles, values and ethical practice, and by demonstrating the proportionality and objectivity of responses.

Advocate for responsible digital ecosystems – including algorithms and AI – that promote tolerance, strengthen social cohesion and reduce harmful information.

Support leadership and staff capacity to uphold the fundamental principles through training, codes of conduct, social media guidance and peer-to-peer exchanges.

Strengthen humanitarian diplomacy and engagement with states, authorities, media and civil society to safeguard principled humanitarian action, counter harmful information and reinforce the credibility and operational effectiveness of the Movement.

Recommendations

States and policy-makers

- Safeguard humanitarian space in legislation, regulation and preparedness frameworks, ensuring the fundamental and humanitarian principles are respected in both digital and offline contexts.
- Avoid instrumentalizing humanitarian actors in partisan or political agendas.
- Publicly reaffirm and support the independence,

impartiality and neutrality of humanitarian organizations.

- Ensure humanitarian exemptions in sanctions, counter-terrorism and regulatory measures that might impede principled action.

Technology platforms

- Monitor and prevent harmful information targeting humanitarian action, with safeguards that respect fundamental rights.
- Ensure transparency, tools and crisis protocols that enable principled humanitarian engagement online.
- Act rapidly to address harassment and harmful information targeting humanitarian staff, volunteers and people in need.

Humanitarian actors

- Develop guidance for principled engagement in contested digital and online spaces, anticipating instrumentalized narratives and preparing clear communication lines in advance.
- Reaffirm and communicate fundamental and humanitarian principles across platforms, avoiding partisan or political alignment.
- Model adherence to the fundamental principles in words, behaviours and actions across all Movement components.
- Equip staff and volunteers with digital security training, psychosocial support and information and media literacy to detect and respond to harmful narratives early.
- Place renewed focus on addressing evolving challenges of intolerance²² across the Movement.
- Foster reflection and learning on communication and humanitarian diplomacy, embedding humanitarian principles in responses to harmful information.
- Involve community leaders and volunteers as validators and messengers, running sustained campaigns that highlight local impact and principled humanitarian action rather than relying only on reactive responses.

Communities and local leaders

- Act as trusted intermediaries reinforcing impartiality and bridging divides.
- Identify polarized spaces and harmful information that threaten humanitarian access, trust and safety.
- Co-create and amplify local narratives that reflect humanitarian values, ensuring humanitarian action remains responsive to community needs.

Endnotes

- 1 International Red Cross and Red Crescent Movement. Proclamation of the Fundamental Principles of the Red Cross. Adopted by the 20th International Conference of the Red Cross (1965). https://disasterlaw.ifrc.org/sites/default/files/media/disaster_law/2024-07/IC%2020%20%281965%29%20English.pdf
- 2 UN General Assembly. Resolution 46/182: Strengthening of the coordination of humanitarian emergency assistance of the United Nations (1991) <https://docs.un.org/en/A/RES/46/182>
- 3 Pictet, J. *The Fundamental Principles of the Red Cross: Commentary*. (ICRC, 1979)
- 4 Devidal, P. 'Lost in Digital Translation? The Humanitarian Principles in the Digital Age,' *International Review of the Red Cross* 2024:106(925), 120–154
- 5 Glover, J. *Humanity: A Moral History of the Twentieth Century*. (1999) p.22
- 6 *Ibid*, p.407–408
- 7 Deffenbaugh, N. 'De-Dehumanization: Practicing Humanity.' *International Review of the Red Cross* 2024:106(925) <https://international-review.icrc.org/articles/de-dehumanization-practicing-humanity-925>
- 8 *Ibid*, pp.56–89.
- 9 Council of Delegates of the International Red Cross and Red Crescent Movement. Resolution CD/24/R5: Call for Respect and Support for Principled Humanitarian Action, adopted 28 October 2024. <https://international-review.icrc.org/articles/reports-and-documents-selected-resolutions-of-the-2024-council-of-delegates-927>
- 10 Deffenbaugh, N. 'De-Dehumanization: Practicing Humanity.' *International Review of the Red Cross* 2024:106(925) <https://international-review.icrc.org/articles/de-dehumanization-practicing-humanity-925>
- 11 Council of Delegates of the International Red Cross and Red Crescent Movement. Resolution 9: Promote Respect for Diversity and Fight Discrimination and Intolerance, adopted 30 November 2003. www.icrc.org/sites/default/files/external/doc/en/assets/files/other/anglais-cd-2003-resolutions.pdf
- 12 International Red Cross and Red Crescent Movement. Proclamation of the Fundamental Principles of the Red Cross. Adopted by the 20th International Conference of the Red Cross (1965). https://disasterlaw.ifrc.org/sites/default/files/media/disaster_law/2024-07/IC%2020%20%281965%29%20English.pdf; International Conference of the Red Cross and Red Crescent. 32nd International Conference of the Red Cross and Red Crescent: Fundamental Principles in Action – 50th Anniversary of the Proclamation of the Fundamental Principles. Conference report (2015) <https://rcrcconference.org/about/previous-conferences/32nd-international-conference/>; International Red Cross and Red Crescent Movement. Resolution 5: Call for Respect and Support for Principled Humanitarian Action, adopted by the Council of Delegates, 28 October 2024, accompanied by an Appeal to States at the 34th International Conference of the Red Cross and Red Crescent (2024) <https://international-review.icrc.org/articles/reports-and-documents-selected-resolutions-of-the-2024-council-of-delegates-927>
- 13 One case which generated a flow of posts online is: ANSA. 'DIGOS probing anti-NGO graffiti.' 27 January 2025. www.ansa.it/english/news/politics/2025/01/27/digos-probing-anti-ngo-graffiti_d0aac4d7-5a6e-43f6-80ca-d6e2f589581c.html. Others were comments and messages to the Italian Red Cross posts accusing us or the ICRC of being partial and non-neutral. These have been removed. Italian journalists referred back to ICRC 'faults' at the time of the Holocaust, thus intending to cast a malicious light on the Red Cross as a whole.
- 14 Harroff-Tavel, M. 'Neutrality and Impartiality: The Importance of These Principles for the International Red Cross and Red Crescent Movement and the Difficulties Involved in Applying Them,' *International Review of the Red Cross* 1989:29(273), 536–552.
- 15 *Ibid*
- 16 Slim, H. *Solferino 21: Warfare, Civilians and Humanitarians in the Twenty-First Century* (Hurst Publishers, 2024), p.242.
- 17 IFRC. *Applying Better Programming Initiative – Do No Harm*. (2016) www.ifrc.org/sites/default/files/2021-08/2016_ApplyingBPI-DoNoHarm.pdf
- 18 The notion of connectors and dividers is referenced in IFRC. *Better Programming Initiative: How to Do Conflict-Sensitive Context Analysis* (2021) www.ifrc.org/document/better-programming-initiative-how-do-conflict-sensitive-context-analysis. This highlights that good programming and effective community engagement require a solid understanding of the local environment, as well as of the role – both actual and perceived – that organizations play. This applies whether work takes place in contexts marked by social instability, violence and conflict or in more stable and predictable settings.
- 19 Devidal, P. 'Lost in Digital Translation? The Humanitarian Principles in the Digital Age,' *International Review of the Red Cross* 2024:106(925), 120–154.
- 20 Glover, J. *Humanity: A Moral History of the Twentieth Century*. (Yale University Press, 1999) pp.22, 24–25.
- 21 *Ibid*, p.414
- 22 Upholding the commitments of Resolution 9 (2003) today requires addressing the digital dimensions of intolerance and integrating responses to harmful information into broader strategies for inclusion, protection and principled engagement – both online and offline.



Chapter 8

Truth, trust and resilience in the humanitarian sector





Chapter 8



Truth, trust and resilience in the humanitarian sector

Contents

	Introduction: The high stakes of the information crisis	303
8.1	Why connection matters in humanitarian response	307
8.2	What's ahead? Evolutions and known unknowns	309
8.3	Defining humanitarian resilience in the information age	313
8.4	Trust brokers – the human bridge to credibility	330
8.5	Humanitarian principles as a compass	332
8.6	Recommendations for resilience	339
8.7	Prerequisites and cross-cutting enablers	345
	Conclusion: Together, we can uphold and reclaim space for humanity	348
	Endnotes	350

Introduction: The high stakes of the information crisis

This chapter focuses on the shift required for the humanitarian sector to move from treating information as a support function to recognizing it as a core component of resilience. Governments, the International Red Cross and Red Crescent Movement, the broader humanitarian sector and wider society all have a role to play in strengthening resilience to harmful information. It draws lessons from the sector's evolution – from the top-down communication models that once dominated humanitarian action, through the period 2005–2015 when information began to be recognized as a form of aid in itself and participatory approaches such as community engagement and accountability¹ took root. These developments reflected an important commitment to more inclusive, transparent and accountable humanitarian action, yet they emerged at a time when the information ecosystem posed far fewer risks. Today, growing distrust, disillusionment, polarization and the increasingly complex operating environments shaped by harmful information have generated a new kind of crisis – one that strikes at the foundations of humanitarian action: an information crisis.

As UN Secretary-General António Guterres warned in *Our Common Agenda*, “disinformation is an existential threat to humanity,”² particularly when it undermines established scientific facts and erodes the social contract.

Over the past decade, the information landscape has undergone seismic change. The humanitarian sector now operates in an environment marked by eroding trust, the rapid spread of harmful information and increasingly polarized public attitudes. Trust – long the bedrock of humanitarian action – is under unprecedented strain, not only from harmful information but also a broader collapse in shared facts and institutional credibility, compounded in 2025 by an abrupt need for a humanitarian ‘renewal’ or ‘reset’ triggered by unprecedented deep funding cuts.

Harmful information during emergencies from the COVID-19 pandemic to inter- and intra-state armed conflicts and other humanitarian crises has exposed the vulnerability of both communities and humanitarian organizations. It exploits cognitive shortcuts, such as familiarity bias (the tendency to trust information that is frequently repeated) and availability bias (the tendency to favour emotionally charged or easily recalled narratives). These biases become especially potent in times of crises, when trustworthy and accurate information is most needed but least accessible.



As the saying goes, ‘It takes only a moment to spread a rumour, but a lifetime to refute it.’ The cost of refuting rumours is completely mismatched with that of spreading them. Moreover, there’s nothing we can do about some ordinary people – sometimes they are more willing to believe false information than the truth. This is a harsh reality.’

Community member, China

In digital spaces and connected offline spaces, the boundary between fact and falsehood has become dangerously blurred. The widespread availability of information online and the ease of expressing and amplifying opinions has created a “virtual world of information and misinformation cohabiting side by side ... one who will help you and the other who will hurt you”.³ The digital domain itself is far from neutral. As the authors of *LikeWar* observe, the internet has become a contested space – a modern battlefield, where harmful information is wielded as a tool of influence, power and control: “Battle on the internet is continuous, the battlefield is contiguous, and the information that it produces is contagious”.⁴

Tactics such as creation of false content, doxing (see [Annex I: Glossary, on page 353](#)), smear campaigns and coordinated harassment are increasingly used to silence, discredit and destabilize organizations, while deepening confusion, distrust and division among audiences. Harmful information erodes trust and misrepresents humanitarian action. The consequences are serious: they threaten staff and volunteer safety, community confidence and humanitarian access, challenging the ability to respond effectively to people in need.

New tools and platforms have lowered barriers to participation, but trust and safety in these digital spaces remain fragile, especially where harmful information spreads faster than verification and moderation systems can keep pace. Algorithms and digital platform dynamics amplify polarizing content, pushing users into echo chambers that reinforce harmful narratives and undermine humanitarian principles. The consequences are operational and immediate: reduced acceptance, increased hostility and limited access to vulnerable populations.

The rise of AI adds a new layer of complexity. AI can now generate credible convincing harmful information at scale – in visual, audio and text form, manipulate public opinion and enhance cyberattacks. Once the domain of a few highly resourced states and private actors, AI is now widely accessible, largely driven by rapid private sector innovation. This shift has lowered the threshold for malicious use, enabling the large-scale creation and spread of harmful content. AI technologies are already being used to manipulate public opinion, deepen societal divisions and erode institutional trust. In humanitarian contexts, **this raises important concerns around the protection of affected populations, the integrity of information environments and the responsible use of digital tools in crisis response.**⁵ AI also amplifies the potential for targeted exploitation. It can automate the identification of high-value individuals or vulnerable groups, increase the precision of cyberattacks and leverage large datasets to exploit people financially or psychologically. It also enables the production of deepfakes and synthetic narratives that distort public discourse and undermine informed decision-making. These capabilities present urgent challenges for humanitarian organizations, not only for community safety, data governance and digital risk management, but also for the ability to operate within polarized information environments. The responsible use of digital tools and AI – together with appropriate safeguards – is essential to maintaining the accessibility, reliability and trustworthiness of information and supporting principled humanitarian action.

The ‘MICE framework’ – standing for money, ideology, coercion (or compromise) and ego – has traditionally been used in intelligence and security studies to explain why individuals engage in espionage or other harmful acts. When applied to the information environment, it helps explain why individuals, groups and state actors create and spread harmful information. Today’s digital spaces add further motivation including status, revenge and attention-seeking, which can drive both personal and organizational behaviour. For political and state actors, these dynamics often play out at a larger scale:

ideology may advance political or strategic agendas; coercion and compromise can shape the actions of allied groups or populations; ego and status can drive the projection of influence domestically or internationally; and revenge or attention-seeking can be used to destabilize opponents or dominate media narratives. Altogether, these factors show that the spread of harmful information is rarely accidental – it is often guided by clear incentives and pressures, whether at the individual, organizational or state level.

Understanding these motivations is critical for humanitarian actors, as it informs strategies to protect staff, volunteers and communities, maintain trust, counter harmful narratives and mitigate challenges to principled access and response in complex information environments. Practical counter-strategies should be tailored to the specific drivers of harmful information while also **identifying opportunities for engagement and collaboration with states, technology platforms and other stakeholders**, since some measures extend beyond those that humanitarian actors can or should implement. Examples include disrupting financial incentives for criminal actors, providing alternative narratives and engagement pathways to counter ideology or status-driven spread, and establishing accountability mechanisms. Recognizing this spectrum of motivations strengthens the ability of humanitarian, policy and other actors to anticipate and respond effectively to harmful information in complex, high-stakes environments.

Importantly, harmful information rarely exists in isolation. It often amplifies the impact of other crises, intensifying the effects of geopolitical conflicts, climate-related and other emergencies. The World Economic Forum has identified this convergence of risks as a defining feature of today's global risk landscape,⁶ posing urgent questions for the humanitarian sector about how to adapt, respond and engage effectively.

In this fragmented landscape, trust in traditional sources of information has sharply declined. Society is increasingly replacing expert knowledge with personal belief, peer experience and influence-driven narratives – a trend described as the 'death of expertise' (Nichols, 2017). People now tend to trust 'people like me' – meaning individuals they perceive as similar to themselves in experience, values or identity – rather than relying on experts or institutions. According to the 2024 Edelman Trust Barometer,⁷ trust is becoming increasingly hyper-local, while media and social media continue to rank among the least trusted institutions in a range of countries. Alarming, two-thirds of people report difficulty distinguishing reliable news from disinformation. While NGOs are still viewed as ethical leaders, especially among populations with high levels of grievance, even this trust is beginning to erode.

As social theorist Niklas Luhmann noted, **trust reduces complexity and enables action in uncertainty**.⁸ In an environment where harmful information spreads rapidly, sustaining **trust, credibility and connection is central to protecting humanitarian space**, ensuring access, and maintaining the effectiveness of principled humanitarian action.

In this context, any humanitarian reset or renewal must prioritize not only explaining the principles of humanity, neutrality, impartiality and independence but also rebuilding both understanding of the impact and effectiveness of humanitarian action and trust in a world where credibility, access and legitimacy are seriously challenged. Understanding the impact of harmful information is essential for **building an evidence base** that informs policy interventions, strengthens operational strategies and guides effective risk strategies and responses. Addressing harmful information is not peripheral – it is central to safeguarding humanitarian space, protecting affected populations and ensuring humanitarian action remains possible.

Contributor Insight 8.1

Information as aid: 20 years on

In 2005, the *World Disasters Report* introduced a bold idea: information is as essential in emergencies as food, water or shelter. It popularized the phrase ‘information as aid’ and asked a provocative question: are humanitarian organizations using information to empower people – or to serve their own interests?

That idea took root. In 2009, it led to the creation of the Communicating with Disaster Affected Communities (CDAC) Network, which bridges media development and humanitarian organizations and champions two-way communication: not just broadcasting messages, but engaging in real dialogue with people affected by crisis.

Today, while ‘information as aid’ is widely accepted in principle, it is unevenly applied in practice. Too often, accountability initiatives fall back on one-way models where information is delivered and communities passively receive it. Another pitfall is treating information only as data collection – monitoring whether people received a message rather than asking what they need to know. Asking if someone received a red balloon – and whether they liked it – tells you little; you’ll never know if someone wanted a yellow one unless you let them steer the conversation. Two-way communication shifts power. It invites participation. It centres people’s knowledge and agency.

Today, the power of information is clearest where it is denied. From internet shutdowns to restrictions on or the targeting of journalists, communities face control over what they can access, say and share, as well as being confronted with a flood of overwhelming information platforms and channels. In this environment, digital technologies and AI are double-edged: they can amplify disinformation and bias, but also help filter noise, translate content and support communities in verifying and making sense of information. For CDAC, engaging with these technologies is now central to putting the principle of ‘information as aid’ into practice.

The crowded and contested informational environment demands that ‘information as aid’ is understood not as the one-way delivery of neutral facts, but as a relationship between equals. It’s not about broadcasting instructions or extracting data, it’s about building trust, exchanging knowledge and recognizing affected people as experts in their own lives. When approached this way, information becomes more than aid, it becomes the foundation for solidarity, accountability and shared decision-making.

Ila Schoop Rutten

Information Integrity Lead

CDAC Network

8.1 Why connection matters in humanitarian response

For humanitarian actors, truth and trust are not abstract ideals – they are essential for effective, principled action. Without trust, access can be denied, staff and volunteers face greater risks and affected communities may miss vital support. In today's digital environment, attention is the battleground, and stories – not just facts – shape perception. **Building meaningful connections with communities is now as important as conveying accurate information.** Messages must be culturally relevant, resonant and accessible across multiple languages and formats, while facilitating **two-way engagement** through listening, feedback and dialogue.

In a digital landscape where attention spans are measured in seconds, storytelling must be instant, visual and memorable. Simplicity matters, but so does resonance: narratives must connect to cultural frames and lived realities, allowing audiences to relate to people in need. This shift requires humanitarian organizations to ensure more than one-way broadcasting of information (one to all), to building relationships, facilitating true two-way communication through listening, multilingual engagement, moderation and feedback loops.⁹

Social media has made information-sharing participatory: people no longer just consume stories – they shape and amplify them, often in real time and with an audience. In this environment, responding with the nuance and care that humanitarian principles demand becomes increasingly difficult, yet no less essential. The concept of 'community' in humanitarian response has also expanded and blurred. It no longer refers solely to affected populations, field-based stakeholders and donors. Today, it includes policy-makers, digital influencers, diaspora networks and global publics – often all at once. As a result, the traditional model of 'one message to many', which persists, is likely to land where the boundaries of who is listening – and who is reshaping and amplifying the message – are increasingly unclear. Humanitarian organizations are not only trying to inform, but also to influence, engage and retain the trust of multiple, overlapping audiences, each with different expectations and varying proximity to different crises. Responding effectively means not only knowing what to say, but also understanding *who* is hearing it, *how* they are interpreting it, and *what* they might do with it.

The challenge, then, goes beyond managing harmful content. It is about restoring and maintaining credibility, connection and trust in an increasingly volatile information environment. Yet, most humanitarian organizations lack the structures and resources for continuous, localized dialogue in real time. Those who fail to adapt, whether through hesitation, fragmentation, caution or inertia, risk being left behind. Institutions unable to communicate their story clearly or to respond effectively when harmful narratives take hold lose relevance and public trust.

Contributor Insight 8.2

Spotlight: IFRC workshop on AI and harmful information – ‘Communities are networks of trust’

An IFRC workshop in July 2025 brought together humanitarian practitioners and technologists to explore the complex intersection of AI, trust and harmful information. Some key insights and recommendations from the discussion were:

Top insights

- **Trust first:** Rather than attempting to counter every piece of misinformation, efforts should prioritize amplifying trusted sources already valued by communities.
- **Timing matters:** Delayed communication – often caused by over-validation processes – can create information vacuums filled by harmful content. Early, transparent engagement is essential.
- **Misinformation is systemic:** It doesn't just affect communities but also confuses decision-makers. Its impact spans every level of the system.
- **Offline trust is foundational:** Digital solutions must be grounded in offline trust-building efforts, such as working with community figures.
- **Silence is not neutral and can be harmful:** The absence of reliable information can itself be a form of harmful information.

Recommendations for humanitarian organizations

- **Integrate community trust indices** into existing assessments and monitoring frameworks.
- **Establish pipelines for ‘good information’ production**, focusing on quality, frequency and relevance.
- **Re-examine humanitarian risk tolerance** in addressing harmful information, acknowledging that inaction or silence can carry its own dangers.
- **Incorporate AI literacy, capacity building and critical thinking** for humanitarian staff and volunteers in training and tools.

This workshop emphasized the need to move from reactive to proactive strategies, embedding human-centred design for trust, transparency and local legitimacy into every layer of AI deployment and information management. By fostering system-wide

alignment and bold experimentation, the humanitarian sector can better navigate the evolving landscape of harmful information.

IFRC, Geneva

8.2 What's ahead? Evolutions and known unknowns

Trust in institutions will likely continue to decline. People increasingly rely on influencers, peer networks and local intermediaries to judge credibility. For humanitarian organizations, credibility will hinge on the ability to engage through these trusted channels, rather than solely on institutional identity. At the same time, gaps in local journalism and media deserts leave communities vulnerable, reducing access to reliable information and amplifying harmful narratives. Strengthening the information ecosystem requires investment in local media capacity, journalism and community-based mechanisms.

Harmful information will increasingly spread in closed, encrypted and algorithmically governed spaces such as WhatsApp, Telegram and social media platforms. These spaces make detection more difficult, but they also offer opportunities for peer-to-peer communication during crises. At the same time, algorithmic moderation and ranking systems can amplify or suppress humanitarian voices, shaping what information people see and highlighting the need for transparency, safeguards and investment by technology platforms, policy-makers and in local media. Further research is still needed to understand the impacts of harmful information on humanitarian operations and on the people they aim to assist and protect.

What is certain is that the information environment will continue to evolve, including the growing use of AI in military operations. It will likely become more volatile, localized and shaped by technology, with AI a defining factor. The rise of AI adds both opportunities and risks. On one hand, generative AI (see [Annex I: Glossary, on page 353](#)) is likely to supercharge harmful information – enabling the scaling of the production of convincing deepfakes, synthetic voices and hyper-targeted narratives that can spread faster than verification systems can respond. It also offers tools for translation, fact-checking and multilingual communication when applied responsibly. The development of agentic AI – systems capable of autonomous decision-making, long-term goal pursuit and multi-step task execution – introduces a new phase in the harmful information ecosystem. Unlike current generative models that largely respond to prompts, agentic AI systems can plan, adapt and act strategically in ways that may accelerate the creation, targeting and amplification of harmful narratives. Humanized AI – such as companion bots – risks eroding trust in real human interaction. As AI systems increasingly emulate human traits and communication styles, they risk eroding the intrinsic value of human judgement and interaction. Meanwhile, the deliberate humanization of AI in design – particularly in the form of companion bots and emotionally responsive systems that elicit sentiment, feelings and projections – may further blur boundaries and subtly shape or manipulate human behaviour.

The future of harmful information and cognitive manipulation – meaning increasingly subtle, data-driven techniques that influence how people think, feel or make decisions

without their full awareness – will likely involve highly personalized, AI-enabled tactics that blur the line between persuasion and deception, making stronger societal, technological and regulatory safeguards essential to protect public trust and cognitive autonomy.

In humanitarian crises, the humanization of AI, by mimicking empathy, care or authority, may undermine trust in genuine human interaction at a time when it is most vital. Crisis-affected populations could begin to attribute human-like intentions or credibility to AI systems, projecting emotions and dependencies onto tools that are not capable of moral judgement or accountability. Such systems may also manipulate sentiment in vulnerable populations. For example, a companion bot that ‘comforts’ individuals could be exploited to spread harmful information, distort perceptions of humanitarian actors or redirect trust away from legitimate support channels. In fragile contexts marked by trauma, displacement or social fracture, this can deepen confusion, exacerbate mistrust and escalate harm. Moreover, the illusion of empathy created by humanized AI risks devaluing authentic human connection in crisis response. **Humanitarian aid relies not only on material assistance but also on empathy, dignity, compassion and trust-building.** If affected communities experience AI-mediated interactions as substitutes for real human engagement, the intrinsic value of human-to-human solidarity may be eroded, weakening the relational foundations of humanitarian action.

On the other hand, AI also brings opportunities: new tools for fact-checking, content provenance and translation, for example, providing multilingual crisis communication. In highly resource-constrained crises, these systems might help bridge gaps – provided they are transparent, well-regulated and carefully integrated into existing humanitarian services. Overall, the challenge lies in ensuring that AI supports rather than replaces human care, reinforcing rather than weakening the trust and solidarity on which humanitarian action depends.

AI holds potential for humanitarian action, but there is a growing risk that cost-driven, unregulated use could harm vulnerable communities. *Building a Responsible Humanitarian Approach: The ICRC’s Policy on Artificial Intelligence*¹⁰ provides an overarching framework to guide the organization’s exploration and use of AI in ways that align with its humanitarian mission and principles. The SAFE AI project,¹¹ led by CDAC Network, The Alan Turing Institute and Humanitarian AI Advisory, with support from the UK Foreign, Commonwealth and Development Office, aims to create practical standards, tools and community-driven frameworks to ensure AI is used responsibly and ethically in humanitarian settings. Monitoring the impact of AI on crisis-affected populations will be essential to ensure that its use remains safe, effective and principled.

The state’s central role in defining and enforcing the boundaries of legitimate information control and response¹² may be reinforced. Regulation and governance related to harmful information remain fragmented. Some frameworks, like the European Union’s Digital Services Act (2024), require major platforms to assess and mitigate systemic risks such as disinformation and to provide transparency on content moderation. Other regulations may restrict freedom of expression or civic space, for example, laws which criminalize vaguely defined ‘false information’.

Global norms on responsible behaviour in cyberspace, including discussions at the UN Open-ended working group on information communication technologies (ICTs) and its successor the Global Mechanism on Developments in the Field of ICTs in the Context of International Security and Advancing Responsible State Behaviour in the Use of ICTs, should provide a forum to address harmful information deployed through cyber means.

This would mirror what was achieved with the protection of critical infrastructure and the explicit references to health care during the COVID-19 pandemic.¹³

The UN Security Council Resolution 2730 (2024) on the protection of humanitarian and UN personnel is an important step toward safeguarding humanitarian action – and a foundation that must be built upon. It calls on member states to take appropriate action to address the increasing threat of disinformation campaigns and misinformation that undermine trust in the UN and humanitarian organizations and put humanitarian personnel at risk. Building on this, global cyber governance forums offer a further opportunity to advance calls for greater protections for humanitarian organizations against harmful information deployed through cyber means. The objective is to ensure that digital threats do not compromise access, safety or principled humanitarian action. Voluntary initiatives like the UN Global Principles for Information Integrity seek to establish shared norms for states, platforms and civil society, but their implementation remains voluntary.

In short, the **information environment ahead will be more sophisticated, pervasive, deeply embedded in daily life and locally fragmented.** The key question is whether humanitarian organizations can collaborate and build the resilience to withstand and counter these pressures; and crucially, whether they can move quickly enough to adapt – embedding harmful information analysis, investing in information and media literacy, community engagement and trust-building as a central element of humanitarian action.

Contributor Insight 8.3



The humanitarian imperative to combat misinformation in disasters

Trust is fundamental to effective disaster management. At every stage – preparedness, response, and recovery – communities rely on clear, accurate information to make decisions that reduce risks and support recovery. However, the growing presence of misinformation and disinformation poses a serious humanitarian threat in Australia, compromising access to life-saving information, undermining trust in institutions and exacerbating harm, particularly for already marginalized populations.

False claims about the existence, severity or cause of disasters – such as the widely circulated hashtag #ArsonEmergency narrative during Australia's 2019–20 bushfires which falsely attributed the bushfires to arson rather than climate-related hazards – undermine public confidence and distract from coordinated action. During the COVID-19 pandemic, misinformation competed with and contradicted official health advice in Australia, fuelling vaccine hesitancy and deepening existing health inequities. In many cases, disasters have been opportunistically used to advance unrelated agendas or intensify existing societal divisions, capitalizing on fear and uncertainty.

While regulatory efforts such as the Australian Code of Practice on Disinformation and Misinformation have sought to address this challenge, the scale and speed of misinformation continue to outpace institutional responses. A recently published [disinformation playbook](#) outlines practical countermeasures, including:

- **Pre-emption and early detection:** Anticipating false narratives by understanding local tensions and building trusted information-sharing networks.
- **Prebunking (see Annex I: Glossary, on page 353) and spread prevention:** Proactively equipping communities to recognize and reject common disinformation themes before disasters strike.
- **Debunking and recovery:** Correcting falsehoods using clear, repeated communication.

The human cost of inaction is clear. Identifying misinformation risks, recognizing that the emblem of the Australian Red Cross helps to reinforce trust when distributing accurate information and addressing misinformation are now core components of day-to-day operations. Tackling misinformation must be treated as a humanitarian priority. This requires equitable, inclusive and culturally informed communication systems. Community-led strategies, trusted messengers and sustained education are critical to protecting lives, strengthening social cohesion and ensuring all communities are supported during times of crisis.

Jenny Gillett

Senior Manager – Policy and Research, External Engagement

Australian Red Cross

Contributor Insight 8.4

The spread of harmful information in situations of conflict and violence

In situations of armed conflict, information can be a lifeline. But when it is distorted, preventing people from accessing life-saving services or leading them to make inadequate decisions about their safety, it can have serious consequences. When information fuels hatred between communities or incites violence, it exacerbates the suffering of people already affected by the harsh realities of conflict.

There is neither an international legal nor agreed-upon definition of harmful information. The ICRC defines it as “information that can potentially cause or contribute to harm, either physically, psychologically, economically or socially”.¹⁴ This includes misinformation, disinformation, malinformation, hate speech and other narratives that spread in violation of international humanitarian law even if they do not fall entirely within those categories.

Harmful information increases people’s exposure to risks and deepens vulnerabilities during armed conflict. For example, when displaced people are intentionally given misleading information about life-saving services and resources, they may be diverted away from help and toward harm. In situations where hostilities are ongoing, false warnings could lead people to make decisions that put them in harm’s way. Speech and narratives that encourage hatred and violence can endanger lives and livelihoods. For example, online

calls for violence against minority groups can trigger acts of violence against individuals and inflict psychological and social harm through harassment, defamation and intimidation. Harmful information also undermines trust in humanitarian organizations' ability to operate, potentially limiting their access to the people they intend to serve.

Addressing harmful information is complex and requires a conflict-sensitive approach. This involves considering the potential harm that the spread of harmful information can have on affected populations. Responses often need to occur at local, national and global levels and may be preventive or reactive. In some cases, humanitarian actors may engage in dialogue with states or non-state actors, while in others or in parallel, they may focus on improving the availability of accurate, timely and reliable information for people. In other cases, humanitarian actors may partner with local and community groups to strengthen preparedness and prevention efforts.

Such responses should be sensitive to existing risks and conflict dynamics and not trigger any unintended harm; this may mean that the focus is not on the veracity of the information or on refuting or correcting it, but rather on the potential for escalation and harm should certain information spread in that context. They should aim to improve the protection of people affected by conflict and violence by both engaging belligerents and other relevant actors, including technology and social media companies, and by mitigating psychological and societal harms. At the same time, it is necessary to strengthen the resilience and agency of people and communities, including diaspora groups, to better navigate information related to risks. This means improving access to knowledge and reliable information and enabling people to identify and understand harmful information and navigate its effects. Equally important is strengthening principled humanitarian action by reinforcing the ability of humanitarian actors to operate and access the people they intend to serve.

Joëlle Rizk

Digital Risks Adviser, Protection Division

International Committee of the Red Cross

8.3 Defining humanitarian resilience in the information age

In the context of harmful information, humanitarian resilience means far more than managing reputational risk. It is the capacity to anticipate, absorb and adapt to information threats – while remaining grounded in humanitarian principles and accountable to affected communities. In an era where digital platforms and algorithmic tools can amplify falsehoods at scale, reactive communication strategies are no longer sufficient.

A truly resilient approach begins with an understanding of harm, not just to organizations, but to communities. Although research remains limited, recent initiatives are shedding light on the damaging effects of harm including at the local level. For example, work by Grand Challenges Canada, Fondation Hirondelle and Internews has highlighted serious consequences: incitement to violence through social media, erosion of social cohesion and deepening mistrust of humanitarian actors.

Research from the University of Melbourne on disinformation during disasters further focuses on how false narratives can fragment communities and inflame tensions, especially in fragile or polarized contexts. Its *City Playbook for Countering Disinformation* underscores the importance of local, context-specific community-led responses that prioritize community trust, civil society engagement and early detection and prevention.¹⁵

To address these harms, humanitarian organizations must go beyond ad hoc responses. This includes investing in systems that monitor digital threats, assess the impact of interventions and identify structural triggers and vulnerabilities that allow harmful information to spread. **Central to this is community engagement: local actors are not only the first to be affected but often the best positioned to create and sustain effective solutions.**

Building resilience requires understanding the information ecosystem and how people access information and their needs, creating space for feedback, building surge capacity to respond to information crises, supporting digital and information literacy and embedding simulations and anticipatory tools into preparedness planning. It also demands a willingness to confront legitimate criticism with transparency and humility, recognizing that trust is not built through messaging alone, but through sustained, principled and impactful humanitarian action.

The following pages showcase examples from different organizations, illustrating how they are confronting harmful information in practice and the lessons that can help guide the wider humanitarian sector.

Contributor Insight 8.5

Crisis communication preparedness at South Sudan Red Cross

The South Sudan Red Cross has established a Crisis Communication Plan to anticipate and respond to the spread of harmful information during emergencies. This plan sets out clear objectives, key messages, target audiences and communication channels to be used in a crisis. To operationalize the plan, the National Society formed a Crisis Communication Committee comprising the Secretary-General, Communications Manager, Safety and Field Coordinator, Partnerships Coordinator and the Emergency Operations Centre. The Communications Manager conducts regular environmental scanning to detect any mention of the South Sudan Red Cross in public discourse and flags negative narratives for immediate action. The Safety and Field Coordinator monitors the safety and security of staff and volunteers, while the Partnerships Coordinator ensures effective coordination with partners. The Emergency Operations Centre includes departmental managers and

three regional coordinators who facilitate two-way communication between headquarters and branches.

Pascal Ladu

Communications Manager

South Sudan Red Cross

Contributor Insight 8.6

UNHCR's Information Integrity Toolkit

Drawing on extensive testing of activities and responses from pilot projects in Asia and the Americas – and in collaboration with a wide range of stakeholders (including UN agencies, humanitarian partners, civil society organizations and NGOs such as digital rights groups, academia, governments and the private sector) – UNHCR released the **Information Integrity Toolkit** in April 2025. Designed as a resource for the entire humanitarian sector, the toolkit provides a structured, four-step response framework with practical tools and guidelines to address misinformation, disinformation and hate speech. It is adaptable to different contexts and operational mandates, enabling humanitarian actors to respond quickly and effectively to threats to information integrity.

Gisella Lomax

Senior Advisor, Information Integrity

UN High Commissioner for Refugees

Contributor Insight 8.7

Strengthening information integrity: From principles to practice

Theory of change for information integrity

Emerging normative frameworks on information integrity are grounded in 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.

Yet, rapid transformations in the global information ecosystem have heightened risks and vulnerabilities. No single actor can address these challenges alone – effective responses require multi-stakeholder collaboration and action. While some stakeholders, like states

and major technology companies, hold the greatest power, resources and responsibilities, others provide vital perspectives and lived experiences that must inform solutions.

Cross-cutting challenges

Strengthening information integrity requires addressing interconnected challenges that create both obstacles and opportunities for building a more resilient global information ecosystem. Key cross-cutting challenges include:

Sectoral silos: Information risks cut across thematic and geographic boundaries, yet responses typically remain compartmentalized in distinct sectors such as elections, public health, climate and conflict. More effective approaches must apply lessons about tackling adversarial behaviour across these domains rather than treating each area in isolation.

Cross-domain information manipulation: Information – and the actors who spread it – move seamlessly between digital and physical spaces. Information risks therefore transcend the artificial boundary between 'online' and 'offline' environments. As trust in digital platforms erodes, particularly amid uncertainty around emerging AI technologies – offline spaces may gain new importance for individuals and communities seeking reliable information. Meanwhile, adversarial actors exploit both domains in tandem to shape public perceptions and influence policy outcomes.

Systematic targeting of information defenders: Researchers, journalists, fact-checkers and civil society activists face strategic harassment and coordinated campaigns designed to undermine their credibility and silence their contributions. These harassment campaigns – frequently gendered and sexualized – create lasting deterrent effects, systematically eroding research capacity precisely when it is most needed.

Research limitations and methodological bias: Data availability varies widely across platforms, pushing researchers to over-rely on single sources and producing skewed assessments that can lead to flawed responses. Geographic and linguistic biases compound these challenges, with most research concentrated in English-language contexts while vast areas of the global information ecosystem remain under-examined.

AI and information integrity: AI is fundamentally transforming how people access information and how adversarial actors generate false content, effectively making societies involuntary participants in a large-scale information experiment with far-reaching consequences. Generative AI tools are proliferating without adequate safeguards, lowering the barriers to producing hate speech and convincing disinformation at scale. In doing so, they undermine every pillar of information integrity – from societal trust to independent media.

The 3R operational framework

Strengthening information integrity requires systematic approaches that move beyond reactive responses to build proactive resilience against evolving information threats. The Research-Risk Assessment-Response (3R) operational model offers a structured framework for organizations to understand and address a range of information risks, and is particularly valuable in resource-constrained settings. Information risks can be understood as actions, conditions or factors that undermine the integrity of information environments and weaken public access to, and understanding of, evidence-based information, informed decision-making, societal trust and cohesion. Socioeconomic and political factors can enable and exacerbate these risks.

Research

Effective interventions begin with rigorous research into information ecosystems, using cost-effective methods including desk reviews, situational analysis and examinations of influence operations or disinformation campaigns. Organizations can leverage existing expertise, partnerships and open-source techniques to identify emerging risks, uncover policy gaps and assess potential solutions. Such research should address critical questions: What risks are present? Who are the drivers? Which tactics make them effective? Which audiences are targeted and with what impacts?

Risk assessment

Risk assessment bridges research and action by setting clear criteria to prioritize responses according to severity, credibility and potential scope of harm. The process assesses factors such as source authenticity, behavioural patterns, narrative content, distribution levels and impacts on target audiences across social, political, operational and human rights domains. Standardized assessment practices classify risks from very low to high, helping determine both the urgency of response and resource allocation.

Response

Response strategies operate across multiple timeframes with four core objectives:

- 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.
- 3 **Mitigation:** Contain risk escalation and reduce impacts in real time.
- 4 **Recovery:** Restore disrupted capabilities while rebuilding trust and resilience.

Implementation faces significant obstacles, including data limitations, systematic targeting of information defenders and researchers, and persistent research biases. Effective response requires skilled personnel, strong coordination mechanisms and adherence to human rights-based professional standards.

Charlotte Scaddan

Senior Adviser on Information Integrity

UN Global Communications

Contributor Insight 8.8

WHO: Building a resilient emergency system (part 2 of 2)

Common challenges

As funding for crisis responses continues to shrink, risk communication, community engagement and infodemic management (RCCE-IM) is too often considered a 'nice-to-have' element rather than a critical component of emergency preparedness and response. Despite growing awareness and progress, countries and agencies still face persistent challenges including:

- **Resource constraints:** Many health authorities lack stable funding and dedicated specialist capacity for RCCE-IM, hampering both preparedness and response efforts.
- **Fragmented coordination:** Siloed working and absence of well-tested, cross-sector coordination mechanisms, such as simulations or joint emergency planning, can delay unified messaging and confuse or frustrate the public.
- **Gaps in co-creation and localization:** Limited experience in co-designing interventions and messages with communities and insufficient attention to local contexts or language diversity can reduce their relevance and impact.
- **Expertise deficits:** Critical skills in infodemic management, social listening and message testing are often lacking. Without proactive communication, these gaps allow voids in which misinformation and disinformation can thrive.
- **Influencer readiness:** Trusted actors such as health workers and community representatives often lack the motivation, training or resources to fulfil their vital communication roles. Thus, RCCE-IM is too often not implemented.
- **Limited research-to-practice collaboration:** RCCE-IM aims to translate behavioural science, evaluation data and the evidence base into real-world application. However, a lack of training and support in applying RCCE-IM can lead to ad hoc interventions that may be neither effective nor sustainable.

Recommendations for humanitarian organizations

Given these challenges and lessons learned, moving forward requires strategic, systemic action:

- **Embed RCCE-IM in emergency planning:** Make RCCE-IM a statutory and operational priority at all levels of emergency preparedness. Ensure that human resources and regular training are funded to support response.

- **Professionalize and train RCCE-IM teams:** Build multidisciplinary teams with technical expertise in RCCE-IM and behaviour change. Invest in capacity building and mentorship for all practitioners, especially at local levels.
- **Institutionalize coordination:** Create robust, pre-tested systems for multi-agency, multilevel collaboration with clear roles and shared protocols. Allocate resources to test coordination mechanisms through simulations and retrospective reviews of past responses.
- **Forge community partnerships:** Proactively engage civil society organizations, community groups and trusted influencers as co-designers and co-implementers of interventions and messaging.
- **Integrate research and evidence:** Bridge the gap between research and practice by ensuring monitoring and evaluation data form an operational basis for rapid response. Establish partnerships and knowledge-sharing mechanisms that enable continuous learning.
- **Systematize feedback and adaptation:** Embed formal systems for listening, collecting public feedback and adapting RCCE-IM strategies dynamically before, during and after crises.

Nancy Claxton
RCCE-IM Technical Officer
WHO

Leonardo Palumbo
Community Engagement
Technical Officer
WHO

Cristiana Salvi
Regional Technical Advisor
for Community Resilience
and Protection
WHO

Contributor Insight 8.9

WikiRumours: Crowdsourcing verified information for community safety and prosperity (part 1 of 4)

The Sentinel Project employed a combination of community-based monitoring, digital verification and two-way communication platforms to address harmful information in South Sudan and the Democratic Republic of the Congo (DRC). Key tools and strategies included:

- **WikiRumours platform:** An open-source system where users submitted rumours via SMS or voice calls. Trained moderators verified the submitted information using a standard rubric and shared responses back within 24 hours through the same channels.
- **Community ambassadors:** Hundreds of locally trained individuals collected, verified and distributed information. These ambassadors also operated

'peace spaces' where community members could engage in dialogue and fact-check rumours in real time.

- **Preventive messaging:** During high-risk periods (e.g., elections), the team proactively disseminated verified updates to pre-empt misinformation. For example, after a polling station in Bunia was burned, verified updates helped calm the community before rumours spread.
- **Collaborative verification groups:** In the DRC, local authorities such as religious leaders and police joined discussion groups to verify and co-disseminate information. This significantly boosted community trust.
- **Local media partnerships:** Radio programmes, town hall meetings and posters were used to deliver verified information in familiar, trusted formats.

What worked well and why?

- Two-way communication via SMS/voice ensured broad reach, especially in low-connectivity areas.
- Community ownership increased credibility – local people trusted messages verified by people they knew personally.
- Targeted counter-messaging limited the spread of misinformation and addressed rumours at their source.
- Real-time verification improved situational awareness for civilians and humanitarian actors.

What did not work well and why?

- Over-reliance on community ambassadors in the early stages proved insufficient. Residents placed greater trust in authority figures like local chiefs or religious leaders, prompting a shift to include them in verification chains.
- Surveillance of the WikiRumours platform by armed groups posed operational risks in some areas, requiring the adoption of more secure communication methods.
- High resource demands – including logistical costs (especially in South Sudan) and time-consuming verification – challenged scalability without sustained funding.

Unintended consequences and lessons learned

- Early trust-building with community leaders is essential for system adoption and long-term credibility.
- Anticipatory messaging, when deployed quickly, can reduce the risk of rumour-driven panic.

- Localized verification mechanisms, such as peace spaces, promote both information accuracy and community resilience.
- Judicial use of verified data: In Beni, system-generated reports were used in court proceedings, demonstrating the value of verified reports as credible records in conflict settings.
- Family reunification: An unexpected outcome was WikiRumours' role in facilitating the reunification of families separated by conflict and displacement. This is a result of the Sentinel Project's efforts to transmit information across different communities. Subscribers used the WikiRumours platform to request information about their missing family members, while community ambassadors worked to identify and locate them. These efforts successfully reunited 82 children with their families.

Anahi Ayala Iacucci, Nabeel Chudasama,
Nabeela Jivraj, Zainah Alsamman
Grand Challenges Canada

Christopher Tuckwood
The Sentinel Project

Contributor Insight 8.10

WikiRumours: Crowdsourcing verified information for community safety and prosperity (part 2 of 4): gaps and support needs

Despite strong results, the Sentinel Project's experience revealed critical gaps in capacity, coordination and specialized support that limited the scale and sustainability of its response to harmful information during the project period.

Identified gaps

- **Technical capacity:** While the WikiRumours system was highly effective offline, the team lacked the capacity to systematically monitor online platforms such as Facebook and WhatsApp, where misinformation also circulated – especially in urban areas.
- **Analytical tools:** The project initially lacked advanced tools for data visualization, rumour-trend mapping and predictive analytics, which could have enabled faster identification of emerging narratives or misinformation spikes.
- **Monitoring and evaluation expertise:** Limitations in designing impact frameworks hindered the ability to capture both behavioural and

perception-level changes across large populations, making it harder to measure long-term shifts in trust or information resilience.

- **Security protocols:** In areas where armed groups surveilled communication channels, the team faced risks without dedicated guidance on secure data handling and risk mitigation in information operations.
- **Sustainable funding:** High operational costs in contexts like South Sudan strained project budgets. The absence of long-term funding commitments made it difficult to retain trained ambassadors or invest in infrastructure (e.g., solar chargers, community radio) beyond initial grant periods.

Opportunities for partnerships, technical support or external guidance

- **Strategic partnerships with tech platforms and digital rights organizations** to enhance online monitoring, detect misinformation patterns and develop secure communication tools tailored for humanitarian contexts.
- **Guidance and toolkits from humanitarian coordination bodies** (e.g., IFRC, CDAC Network) to standardize rumour-verification protocols, establish safety measures for community reporters and integrate responses to harmful information into broader humanitarian coordination systems.
- **Cross-sector alliances with media organizations and trusted local influencers** (e.g., radio stations, religious leaders, teachers) to co-deliver counter-messages and strengthen long-term media literacy.
- **Pooled funding mechanisms or donor consortia to provide multi-year, flexible financing** for community-driven harmful information initiatives, ensuring continuity and enabling scale-up.

To counter harmful information more effectively, system-level strategies must prioritize local engagement, trust-building and cross-sector collaboration. Based on lessons from the Sentinel Project's implementation in South Sudan and the DRC, the following recommendations for organizations are proposed:

Prioritize community-led verification mechanisms

- Empower trusted local actors (e.g., chiefs, faith leaders, teachers, health workers) to act as information stewards, not just recipients.
- Integrate community feedback loops and 'peace spaces' into humanitarian programming as standing structures for rumour tracking, counter-messaging and dialogue.

Develop policy guidance and coordination protocols for harmful information

- Co-develop clear guidance on responding to harmful information in conflict settings, including defined roles, verification standards and escalation pathways.
- Embed coordination within humanitarian cluster systems to avoid siloed responses or purely reactive responses.

Formalize collaboration with local media and radio networks

- Support community radio stations and local journalists as frontline responders to misinformation.
- Establish shared content platforms or templates for rapid, localized dissemination of verified, culturally contextualized information.

Advocate for digital inclusion and responsible platform governance

- Engage social media companies to provide localized misinformation mitigation tools (e.g., flagging features, WhatsApp fact-check bots) in low-bandwidth, multilingual contexts.
- Advocate for AI transparency, stronger content moderation investments and data-sharing frameworks adapted to humanitarian needs in fragile settings.

Promote harmful information literacy across the ecosystem

- Embed media and rumour literacy into school curricula, health outreach and aid distribution channels.
- Train volunteers, government field staff and humanitarian workers – not just technical staff – on identifying, documenting and responding to harmful information.

Invest in anticipatory communication systems

- Deploy proactive messaging protocols before elections, vaccine campaigns or anticipated flashpoints, using SMS, posters and radio to pre-empt likely rumours before they spread.
- Use predictive analysis models, drawing on existing rumour databases such as WikiRumours, to identify potential harmful information hotspots in advance.

Anahi Ayala Iacucci, Nabeel Chudasama,
Nabeela Jivraj, Zainah Alsamman
Grand Challenges Canada

Christopher Tuckwood
The Sentinel Project

Contributor Insight 8.11

Centre for Humanitarian Dialogue's disinformation diplomacy: Mediating the digital frontlines of conflict

Formal diplomacy is under pressure. We are witnessing more armed conflicts today than at any time since World War II, alongside a global struggle over truth, trust and technology. Online spaces are increasingly weaponized, fuelling violence, eroding social cohesion, undermining peace efforts and development gains, and hardening the positions of conflict parties – making dialogue and compromise harder to achieve. The rise of AI-powered tools and the rollback of platform moderation are accelerating these risks. Automated online manipulation enables the mass production of synthetic content, bot-driven amplification and coordinated influence operations, while lowering the barriers to 'disinformation-for-hire' services.

Vulnerable communities in conflict zones will continue to bear the brunt of these dynamics, as global diplomacy and international norms struggle to keep pace with rapid technological disruption and geopolitical competition.

Current diplomatic responses remain largely reactive, with an emphasis on counter-disinformation, regulation and digital literacy. While these approaches are essential for addressing the 'disinformation supply chain' (production, distribution, consumption), they are not sufficient on their own. Social media risks must also be addressed at the source: by influencing the behaviour of armed conflict actors not just through sanctions but through broader strategies that reduce the incentives and impacts of information manipulation. Lasting solutions require a comprehensive approach, one that tackles all elements of the chain, from origin to outlet and from platforms to people.

Over the past five years, the Centre for Humanitarian Dialogue, a Swiss foundation with a proven track record in preventing and resolving armed conflicts through discreet diplomacy, has engaged in at least 70% of the world's most violent conflicts,¹⁶ working to reduce suffering, foster dialogue and open pathways to stability and development.

During this period, the centre has expanded its mediation mandate to address the negative impact of digital technologies, including social media, on armed conflicts and peace processes, piloting innovative approaches in more than 15 countries to promote online restraint, including:

- **Facilitating seven social media codes of conduct** in **Nigeria** (2021), **Kosovo** (2021), **Bosnia and Herzegovina** (2022), **Thailand** (2023) and beyond.
- **Establishing disinformation de-escalation channels** between political actors in the Caucasus.
- **Running mediation workshops** on managing digital harms for local mediators in the Horn of Africa.

- **Building capacity for diplomats and peacebuilders** to address online threats in crises.
- **Creating discreet backchannels** with major social media platforms to flag risks and discuss prevention and mitigation measures.

The Centre for Humanitarian Dialogue's approach blends discreet diplomacy with multi-stakeholder dialogue, engaging conflict actors, civil society, regulators, social media platforms and 'unconventional' conflict stakeholders, such as social media influencers, to identify shared concerns and foster solutions that encourage online restraint and mitigate social media's harmful impact in conflict contexts.

From our work, we have identified at least three key lessons that form the foundation of our 'disinformation diplomacy' – a set of dialogue-based approaches designed to complement, not replace, existing initiatives to address evolving information threats and to strengthen traditional peace and security efforts:

- **Online restraint requires offline dialogue.** Progress becomes possible when conflict parties recognize that disinformation can escalate into risks none can control or are willing to bear. At this stage, conflict actors may agree on 'red lines' and voluntary codes of conduct or norms for responsible behaviour, sometimes even before formal peace talks commence. Establishing formal or informal channels to de-escalate disinformation – similar to cyber diplomacy protocols – remains rare but is increasingly necessary. Such initiatives can also act as confidence-building measures in broader peace processes.
- **Codes of conduct matter but require greater collaboration and accountability.** The Centre for Humanitarian Dialogue has facilitated seven voluntary social media codes of conduct, often linked to elections. While useful in weakly regulated contexts, turning words into action demands institutional accountability, civil society oversight and active collaboration with platforms. Early, sustained engagement with platforms is essential, as their policies and algorithms can inadvertently exacerbate tensions and conflict, positioning them as actors in the conflict and diplomatic space.
- **Local ownership is essential:** The most effective norms are those developed, demanded and defended by those most affected by online harm. Local actors – often best placed to influence community dynamics – must play a central role in shaping and upholding standards for responsible online behaviour. Yet they are too often excluded from formal dialogue and mediation processes.

Today, conflict mediation is no longer solely about persuading parties to lay down physical weapons. It also involves negotiating restraint in the use of digital arsenals and building partnerships with platforms and local actors to prevent, mitigate and resolve armed conflicts. **Cognitive warfare is a race to the bottom.** In this era of fragmentation and multipolarity, the international community must back disinformation diplomacy efforts with both real political will and adequate resources.

Jacobo Quintanilla

Programme Manager, Social Media and Conflict Mediation

Centre for Humanitarian Dialogue

Contributor Insight 8.12

Strengthening information integrity: Towards a preventive and inclusive approach to disinformation

Five key recommendations

Prevent the effects of disinformation by strengthening the professionalism, independence and viability of local media.

- Enhancing these qualities enables local media to respond effectively to community information needs and to create trusted spaces for inclusive dialogue – before rumours and disinformation fill the vacuum and fuel tensions.
- This journalistic approach includes, but precedes, fact-checking, which reacts only after disinformation is already circulating. It also goes beyond prebunking, which focuses narrowly on pre-identified disinformation topics.
- Crucially, support to local media must remain distinct from strategic communication and public diplomacy, which aim to promote the views of funders. Blurring this line risks undermining media credibility and eroding public trust.

Strengthen local-level interaction between media and diverse population segments to understand information needs.

- Digital tools and AI can enhance the media's ability to gauge public sentiment at scale. However, in-person engagement – such as face-to-face meetings, focus groups and field research – remains essential to avoid conclusions distorted by the digital divide or algorithmic bias.

Develop a hybrid (offline and online) inclusive multimedia offer that meets the needs of diverse and specific audiences.

- When selecting formats, broadcast methods, languages, topics and the journalists best suited to address them, it is essential to consider social inequalities and promote inclusivity. This includes taking into account factors such as gender, class, culture, language, religion, urban or rural location, educational background and the digital divide.

Strengthen journalists' and their networks' knowledge of topics that are frequent targets of disinformation.

- Thematic training should cover areas such as conflict dynamics, politics, elections, the environment, gender, the economy, justice and public health. These sessions should be offered as complementary modules – only after

journalists have acquired a solid grounding in the core principles and practices of journalism.

Improve journalists' and audiences' understanding of how digital platforms and social media algorithms work and AI functions.

- Training should address both the opportunities and risks of using these technologies when producing and disseminating public interest media.
- Media outlets should develop clear charters on the use of these technologies – especially for generative AI – and ensure that their application is transparent to audiences.
- Media literacy should be strengthened through accessible and engaging programming that raises awareness of the risks posed by evolving information ecosystems, particularly among marginalized groups.

Sacha Meuter

Head of Research and Policy

Fondation Hironnelle

Contributor Insight 8.13

Rebuilding local information ecosystems: A critical pillar of post-crisis recovery

ICT infrastructure is often overlooked or excluded from post-disaster and post-conflict reconstruction plans. It is frequently excluded from negotiations due to the normalization of its deliberate targeting by political actors or warring parties, often under a veil of limiting transparency and accountability. As a result, millions of people remain in a state of prolonged or permanent disconnection from the internet for months or years after a disaster has struck or a conflict has ended. This not only impacts resilience and recovery but also increases vulnerability to future crises.

In response to the persistent reluctance of the international community to address this issue, civil society is stepping up with coordinated action. For example, 7amleh – the Arab Center for Social Media Advancement, together with the Palestinian Digital Rights Coalition and dozens of international organizations, has launched #ReconnectGaza, a global campaign calling for the rebuilding of Gaza's telecommunications network and the recognition of access to communication as a fundamental human right.¹⁷

Addressing this challenge requires a holistic approach that includes both short-term emergency connectivity solutions, such as eSIM cards, satellite internet access and mobile communication hubs, to restore basic services, as well as long-term investment in modern telecommunications infrastructure, including fibre optics and renewable energy-powered

networks. Such infrastructure is essential not only for communication but also for the delivery of education, healthcare and economic recovery.

Giulio Coppi

Senior Humanitarian Officer

[Access Now](#)

Marwa Fatafta

Middle East and North Africa Policy
and Advocacy Director

[Access Now](#)



Explainer: Offensive, defensive and integrated responses to harmful information

Defensive responses aim to contain, correct or minimize the damage caused by harmful information. These are typically reactive and focus on protecting an organization's reputation, staff safety and operational access. Key approaches include:

- rapid fact-checking and myth-busting
- activation of crisis communication protocols
- quiet corrections and bilateral engagement with relevant stakeholders
- clarifying harmful information through trusted channels
- strengthening internal alignment on messaging
- ensuring decision-making authority as close to the operational context as possible and escalated only when necessary.

Offensive responses take a proactive approach, aiming to influence the information environment before harmful narratives take hold. These strategies focus on amplifying credible voices, building public trust and pre-empting harmful information by occupying the narrative space early and intentionally. Key approaches include:

- strategic storytelling and values-based campaigns
- partnerships with local influencers and media
- narrative inoculation ('prebunking' tactics)
- community co-creation and distribution of messages
- digital monitoring to anticipate and counter emerging harmful information.

Integrated responses blend defensive and offensive tactics to provide both immediate protection and longer-term influence. These approaches ensure that rapid reaction is

linked to ongoing narrative shaping and community engagement, creating a continuous cycle of protection and trust-building. Key approaches include:

- embedding rapid response capacity within long-term communication strategies
- conducting real-time monitoring to inform both corrections and proactive messaging
- coordinating cross-functional teams to align operational updates with storytelling
- linking incident management to broader reputation and trust-building goals
- sharing lessons from past incidents to strengthen future preparedness.

All three approaches are essential. Defensive responses manage acute incidents. Offensive responses strengthen long-term resilience and trust. Integrated strategies connect the two – ensuring the proactive and reactive shaping of the information ecosystem.

Contributor Insight 8.14

Recognizing the systemic risk

Harmful information is undermining not just communications strategies, but also access, safety and trust. Online narratives – whether grounded in fact, misperception or disinformation – have triggered real-world consequences including operational restrictions, reputational crises, funding freezes and security threats. Yet the humanitarian sector continues to treat harmful information primarily as a communication challenge, rather than a systemic risk. In contexts such as the Sahel, even neutral updates are being reframed as evidence of espionage or political bias, while silence is interpreted as complicity. In Syria and Myanmar, communities turn to platforms like Facebook not only to express anger, but also to request aid.

Proactive monitoring is no longer optional. AI-powered tools, like those employed by Insecurity Insight, have proven essential for detecting sentiment shifts, identifying disinformation surges and flagging rising hostility before it escalates. These tools have captured moments when harmful narratives were not only widespread, but were actively shaping public understanding of humanitarian principles, particularly neutrality.

Four shifts are now needed:

- **Coordinating sector-wide efforts:** Harmful narratives rarely stop at one organization's door. A collective response – including coordinated messaging, data-sharing and joint digital risk assessments – is essential to protect humanitarian space.

- **Investing in digital literacy and local partnerships:** From community influencers to local media, trusted local actors are essential allies in countering misinformation and amplifying accurate information.
- **Reframing digital engagement as essential to humanitarian access:** Social media must be recognized not only as an outreach tool but as a space where acceptance is won or lost. As community-driven content moderation becomes more common, aid actors must actively participate in these spaces, not retreat from them.
- **Developing a humanitarian language for the digital space:** Social media has developed its own language and norms. Traditional humanitarian language that evolved in diplomatic corridors needs to be translated into communication that resonates on social media: clear, sharp and unambiguous. The use of diplomatic phrases is just as out of place on social media as an emoji would be in a UN Security Council resolution.

Christina Wille

Director

Insecurity Insight

Clara de Solages

Researcher

Insecurity Insight

8.4 Trust brokers – the human bridge to credibility

Access to trusted information depends not only on what is communicated – the message – but also on who delivers the message – the messenger. In the Red Cross and Red Crescent Movement, local staff and volunteers will increasingly serve as the ‘trust brokers’: individuals who play a vital intermediary role in building, maintaining and restoring trust between communities and organizations. Trust brokers help translate institutional intent into locally meaningful terms, bridge gaps in power, knowledge and access and vouch for the credibility of both information and actors. Their effectiveness lies in being perceived as independent, culturally grounded and aligned with community needs. As members of the communities they serve, they are uniquely positioned to navigate sensitive dynamics, defuse tensions and foster dialogue.

In contexts marked by harmful information or low institutional trust, **trust brokers often form the first and most credible line of communication.** Supporting them with timely, accurate messaging, digital access and tools, and clear guidance is essential – not only to counter harmful information but to sustain humanitarian access and reinforce the legitimacy of humanitarian action.

A sustainable presence depends not only on continuity on the ground but on trusted relationships, principled engagement and the ability to adapt to evolving community needs over time.

Humanitarian organizations have a duty of care toward their staff and volunteers to help them cope with the personal and professional impact of harmful information. Exposure to online harassment and harmful information can cause stress, reputational

risk and moral distress. Providing psychosocial support, digital safety training and clear organizational guidance is essential to safeguard their well-being and maintain operational effectiveness.

Who are the influencers in and on the humanitarian sector?

- **Volunteers and frontline staff:** Often the **true bridge** to communities, they are the most trusted voices because they are embedded locally, speak the language(s) and share lived realities.
- **Affected communities:** Displaced people, refugees and disaster survivors often become influencers by sharing lived experiences that reframe global perceptions.
- **Humanitarian organization leadership:** Leadership figures in international organizations and NGOs shape high-level narratives, policy influence and donor agendas. However, they are currently less visible.
- **States** are among the most powerful influencers through funding, legislation, public messaging and control over access. Their endorsement can safeguard humanitarian space, while their rhetoric or laws can just as easily shrink it. States shape narratives, and they can amplify principled humanitarian action and trust or fuel harmful information that delegitimizes organizations and stigmatizes affected communities.
- **Local authorities and officials** play a critical role in shaping perceptions of humanitarian action. Their statements, policies and engagement with communities can either enable or constrain access, influence trust and affect the safety of humanitarian personnel.
- **Community leaders and faith-based actors:** They command local trust and legitimacy and their endorsement can make or break acceptance of humanitarian action.
- **Media and journalists:** They play a major role in shaping public understanding of crises and humanitarian needs.
- **Digital influencers:** Increasingly, popular online voices (YouTubers, TikTokers, diaspora bloggers) can spread humanitarian narratives far faster than official channels.

Does the humanitarian sector ‘need’ influencers? Yes, for reach and resonance: in a crowded information ecosystem, humanitarian organizations cannot rely only on institutional voices. Influencers – whether local, digital or community-based – help amplify messages, contextualize them and connect them with people’s values and concerns. But selectively: not all influencers align with humanitarian principles thus due diligence is required to avoid co-optation, politicization or loss of neutrality. They can provide:

- **Scale:** Influencers can rapidly reach audiences that humanitarian agencies struggle to access.
- **Localization:** Community or local leaders can ensure information is trusted and culturally and contextually relevant.

- **Storytelling:** Personal voices can humanize crises, fostering empathy and solidarity in ways official reports cannot.
- **Countering harmful information:** Influencers can debunk rumours or redirect harmful narratives.

However, there are risks to neutrality and impartiality, of over-reliance and of undermining authenticity or exploiting humanitarian messages for profit. The humanitarian sector *may* need influencers, but it must approach them differently than consumer brands – prioritizing trust, impartiality and long-term community relationships over reach at any cost.

8.5 Humanitarian principles as a compass

The humanitarian principles rely on trust, credibility and clarity of purpose: qualities that can be drowned out in noisy, emotionally charged environments. In a world where **perceived authenticity is the currency of influence**, humanitarian organizations must not only communicate in principled ways but also visibly align their messaging with their actions. The risk is not only that falsehoods outpace facts, but that neutrality, impartiality and independence are misunderstood – or worse, mistrusted – when they fail to align with the prevailing emotional or political narratives of the moment. Humanitarian principles are not only a moral compass but also a vital operational safeguard. The principles serve a dual purpose:

- **aspirational:** reaffirming the humanitarian ideal to alleviate suffering, protect dignity and assist solely on the basis of need
- **practical:** providing a tested framework for maintaining access, navigating contested environments and sustaining trust.

To remain effective, the principles must be actively demonstrated: visible in how principled humanitarian actors engage, how they listen and how they respond. Upholding humanity today means countering dehumanizing narratives, reinforcing dignity through action and communicating with clarity, humility and consistency. Ultimately, principled humanitarian action in the digital age demands more than operational competence. It requires widespread ethical clarity, collective discipline and the courage to resist expedient or reactive narratives. In a world shaped by emotional velocity and harmful information, the **humanitarian principles are more than a moral compass – they are a critical operational safeguard and one of the last defences against the erosion of trust, access and the humanitarian space.**

An emotionally reactive information ecosystem threatens the very conditions on which principled humanitarian action depends – dialogue, trust and space for reasoned engagement. In today's volatile information environment, humanitarian communication is rarely perceived as a neutral act. When handled poorly, it can cross red lines – amplifying harmful narratives, oversimplifying complex realities or becoming co-opted for political or ideological ends. Humanitarian actors face a delicate balance: maintaining transparency while safeguarding operations and security, and upholding neutrality without appearing detached or indifferent. This tension is most pronounced in contexts

where trust has already been eroded. In such environments, even accurate and well-intentioned communication can be met with scepticism, suspicion or outright hostility.

In response, the sector must reaffirm the humanitarian principles not only as a compass, but as a practical framework for navigating contested information spaces. Humanity, neutrality, impartiality and independence are not abstract ideals; they are operational standards, demonstrated through consistent action and credible engagement. In fractured environments, visibly adhering to these principles needs to be reinforced, especially when harmful narratives seek to politicize or delegitimize humanitarian action.

At the same time, the broader information economy is increasingly emotion driven. Surprise, anger and disgust dominate digital platforms, amplified by algorithms and now by AI. Tactical interventions such as digital 'circuit-breakers' and sentiment-based analysis show promise in interrupting the viral spread of emotionally charged harmful information.¹⁸ These measures aim to protect the public sphere by slowing virality, not silencing dissent.



Circuit-breakers

In the digital context, **circuit-breakers**¹⁹ are interventions designed to slow or disrupt the spread of harmful information before it becomes viral. Much like in financial markets, where circuit-breakers halt trading during volatility, these mechanisms temporarily limit the amplification of content that exhibits signs of coordinated manipulation, emotional extremity or rapid spread. This can include platform-triggered slowdowns, content throttling or requiring fact-checking before further distribution. The goal is not to censor, but to create space for verification, reduce emotional escalation and protect public discourse.



Sentiment-based analysis

Sentiment-based analysis uses **natural language processing and machine learning** to detect the emotional tone behind digital content. It categorizes messages as positive, negative or neutral and can further identify specific emotions such as anger, fear or empathy. In humanitarian contexts, this analysis can help organizations understand public mood, track shifts in community perception and anticipate narrative escalation. It supports early warning and communication strategies by highlighting emerging risks, sentiment hotspots or emotional manipulation. It is both what is said – and what is left unsaid – and by whom.

Contributor Insight 8.15

WikiRumours: Crowdsourcing verified information for community safety and prosperity (part 3 of 4) – measuring impact and harm

How was the impact of harmful information assessed?

- Baseline and follow-up surveys measured self-reported changes in access to reliable information, trust in sources and perceived safety.
- Focus group discussions provided context-rich insights into how misinformation influenced decisions, such as whether to flee, return home or accept aid.
- Monitoring of offline and SMS-based channels tracked rumour types, frequency, response time and the effectiveness of efforts to counter misinformation.

Reach: Over 27,000 direct subscribers across South Sudan and DRC received verified updates, with indirect reach exceeding 2 million people through community sharing, radio and posters.

Behavioural change: Surveys found that 85% of users believed the platform (WikiRumours) helped prevent rumours. Communities reported exercising more cautious behaviour – verifying claims before reacting or sharing information.

Service uptake: After verified messages were shared during Ebola/COVID-19 outbreaks, communities showed greater willingness to engage with health services, reflecting improved trust in both humanitarian responders and the information they provided.

Offline feedback: Community ambassadors regularly gathered anecdotal feedback in 'peace spaces' and town hall meetings, which informed strategy adaptations.

Methods used

- **Baseline and follow-up surveys:** These measured changes in self-reported access to accurate information, trust in sources and perceptions of safety.
- **Focus group discussions:** These provided qualitative insights into how misinformation shaped community behaviour and attitudes toward humanitarian aid and services.
- **Rumour report analytics:** The WikiRumours system automatically logged rumours submitted via SMS and voice calls, enabling the team to assess frequency, response time and rumour resolution rates.

- **Community feedback loops:** Regular in-person ‘peace spaces’ and dialogue sessions gathered real-time insights from residents and local stakeholders.
- **Primarily offline monitoring:** The project relied on SMS short codes, toll-free voice lines and community ambassador reports. These offline channels were most relevant in low-connectivity, rural environments.
- **Limited online monitoring:** In areas like Bunia and Beni with some digital access, project staff informally tracked social media trends, especially during elections. However, systematic online monitoring was not a core component due to infrastructure limitations.

Changes tracked

- **Behaviour:** Community members increasingly verified information before acting. For example, after receiving counter-messaging about a false attack, people chose not to flee, avoiding unnecessary displacement.
- **Trust:** 85% of users surveyed believed that WikiRumours helped prevent the spread of false information. Humanitarian actors also reported improved coordination with communities that received verified updates.
- **Access to services:** In areas where rumours initially discouraged vaccine uptake or use of health services, sharing corrected information led people to attend and engage more with both humanitarian and health services.

Data gaps or challenges in measurement

- **Limited online monitoring:** Low internet penetration meant the project focused on SMS, radio and word-of-mouth channels rather than systematic tracking of online platforms.
- **Pandemic restrictions and logistical barriers:** In rural or volatile areas, these factors reduced the frequency of data collection.
- **Attribution challenges:** Without larger-scale studies, it was difficult to isolate the specific impact of misinformation from other conflict drivers such as insecurity or displacement.

Anahi Ayala Iacucci, Nabeel Chudasama,
Nabeela Jivraj, Zainah Alsamman
Grand Challenges Canada

Christopher Tuckwood
The Sentinel Project

Contributor Insight 8.16

WikiRumours: Crowdsourcing verified information for community safety and prosperity (part 4 of 4) – policy and framework gaps

- 1 Lack of formal misinformation, disinformation and hate speech policy:** While operational tools exist, there is no dedicated organizational policy on harmful information that includes a standard template and operating procedures across regions and scenarios. Such a policy would make this work easier and more standardized.
- 2 Absence of a universal safeguarding policy for digital engagement:** With the growing use of SMS and digital rumour platforms, there are no specific protocols addressing data protection, digital surveillance risks or safe engagement in contested online spaces.
- 3 No escalation framework:** There is no formal process for escalating misinformation that threatens aid operations or community safety to local authorities, humanitarian clusters or social media platforms.
- 4 No application of a standardized impact measurement framework:** Although monitoring tools are in use, Grand Challenges Canada's impact measurement framework (developed as part of our [misinformation, disinformation and hate speech scoping study](#) and designed to track changes in behaviour, trust and rumour prevalence over time) was not yet developed during the project period.

Suggested frameworks and tools that would be beneficial

- 1** A comprehensive harmful information response protocol integrating rumour tracking, rapid verification, community engagement and staff safety measures.
- 2** Cross-sector toolkit for humanitarian responders with templates for counter-messaging, local risk assessments and training curricula on harmful information.
- 3** Safeguarding and risk mitigation guidance for digital rumour collection and verification in high-risk areas, including protocols for staff and volunteer protection.

- 4 A harmonized harmful information impact measurement framework aligned with broader humanitarian indicators to evaluate changes in trust, access and behavioural response.

Anahi Ayala Iacucci, Nabeel Chudasama,
Nabeela Jivraj, Zainah Alsamman,
Grand Challenges Canada

Christopher Tuckwood
The Sentinel Project

Contributor Insight 8.17

International Red Cross and Red Crescent Movement Initiative on Harmful Information

To address the negative impact of harmful information on trust and acceptance in the Red Cross and Red Crescent Movement, the ICRC, IFRC and Swiss Red Cross launched the Movement Initiative on Harmful Information. The initiative's overarching objective is to strengthen the Movement's ability to address harmful information while leveraging its unique strengths, values and global network to safeguard the space for principled humanitarian action. It aims to:

- **Build capacities** of all Movement components to address harmful information through a multidisciplinary approach.
- **Establish a coordination mechanism** for crisis management, enabling early detection, in-depth analysis and collective responses to harmful information incidents.
- **Contribute to global knowledge, humanitarian diplomacy and advocacy** efforts.

Its governance structure includes a Steering Committee, Permanent Secretariat and Coordination Group overseeing four interconnected workstreams: crisis management, training, thematic, and external engagement. The initiative also undertakes to engage in humanitarian diplomacy activities to influence global decision-making in this field. As of June 2025, the Swiss Red Cross is hosting the Movement Initiative on Harmful Information Hub.

Swiss Red Cross

Contributor Insight 8.18

The importance of research and evidence-based practices to counter harmful information and support trust in the humanitarian sector

In an environment where trust is fragile and the consequences of misinformation and disinformation can be immediate and severe, research and evidence-based practices are essential for generating actionable insights that reinforce credibility, resilience and accountability.

Robust research helps humanitarian organizations design interventions that are responsive to local realities and grounded in verified information – reducing the risk of perpetuating inaccuracies and often resulting in more innovative and cost-effective practices. It provides the foundation for understanding the sources, patterns and impacts of harmful information and allows humanitarian actors to move beyond reactive responses and instead develop proactive strategies informed by data, context and community insights.

Evidence-driven communication also fosters transparency and trust with affected populations. When communities see that humanitarian organizations are guided by reliable data and open about both what is known and what remains uncertain, trust grows. This trust is essential not only for countering harmful narratives but also for ensuring that humanitarian responses are accepted, relevant and effective.

As a distributed network of research, academic and scientific entities and initiatives within the Movement, the Red Cross Red Crescent Research Consortium (RC3) supports continued commitment to research and the integration of evidence-driven practices and policies. It aims to ensure the Movement remains trusted and responsive in meeting the needs of affected communities worldwide.

In a sector defined by urgency, constraints and complexity, research is sometimes viewed as a luxury. But in today's volatile information environment, it is a necessity. Evidence-based practice is not just a matter of accuracy – it is a matter of ethics, impact, trust and, ultimately, saving lives. Looking ahead, evidence must become more than a retrospective tool – it must serve as a compass for anticipating challenges, adapting to evolving contexts, informing policy and mitigating the effects of harmful information. This requires investing in predictive research, real-time data analysis and community-based monitoring systems that can detect emerging narratives and inform timely, context-specific responses.

Red Cross Red Crescent Research Consortium (RC3)

8.6 Recommendations for resilience

Building resilience against harmful information and safeguarding principled humanitarian action requires more than isolated interventions. It demands a coordinated, values-driven roadmap embedded across humanitarian diplomacy, community engagement and accountability, safer access initiatives, programme design, preparedness, risk management and communication strategies. The following eight pillars offer actionable steps across short-, medium- and longer-term horizons.

8.6.1

Trust as a strategic asset

Trust is central to humanitarian action – supporting access, operational effectiveness, delivery and legitimacy. Harmful information seldom creates mistrust in humanitarian action on its own: it amplifies existing tensions, inconsistencies and perceived shortcomings. Trust is not static or binary. It exists on a spectrum shaped by cultural experience, power dynamics and exposure to harmful narratives. It evolves along a continuum: **tell me** → **show me** → **prove it** → **keep proving it**. At each stage of this continuum, trust can be weakened or strengthened, but never assumed; in humanitarian crises, it must be continually earned and safeguarded against the risks and corrosive effects of harmful information.

The messenger matters: volunteers, local staff and local leaders often serve as trust brokers. Supporting them to share accurate, timely and accessible information builds credibility from the ground up. Engagement and community information networks are central to resilience and to reduce the vacuum in which harmful information thrives. Leadership accountability and principled consistency across humanitarian organizations and operations are non-negotiable. People assess institutions not only by what they say, but by what they do and whether the two align.

Two dimensions of trust are critical:

- **Operational trust**, grounded in presence and (human) proximity in interactions with affected communities, authorities, armed actors, media and peers. This trust must be personal and embodied: every staff member and volunteer must carry the organization's humanitarian integrity in their behaviour, grounded in principles, standards and professionalism.
- **Institutional trust**, built through principled behaviour, ethical conduct, accountability and regulatory compliance so that stakeholders believe the organization's stewardship, competence, effectiveness and values.

Action points

- **Short term:** Strengthen internal messaging and training that emphasize trust-building behaviours and responsibilities of individuals as ambassadors of principled humanitarian action. Map trusted community and authority figures (e.g., chiefs, religious leaders, elders) and include them in credibility chains; formalize partnerships to co-verify and disseminate information.

- **Medium term:** Monitor and analyse community perceptions of trust linked to the operational environment (not only reputation) and impact of harmful information. (The IFRC's Community Trust Index could be extended to measure harmful information.)
- **Long term:** Integrate trust-based performance metrics that capture both relational and technical dimensions of humanitarian action. Institutionalize multi-layered trust networks as part of preparedness and accountability systems.

8.6.2

'Right-touch' compliance in a digital age

Accountability in the information space is essential to sustaining trust and legitimacy in humanitarian action. Humanitarian organizations must apply the same standards of transparency, responsibility and protection to their communication as in their operations – verifying information, mitigating against harmful content and addressing unintended impacts. Feedback mechanisms should enable communities to question and influence how information about them is used.

Humanitarian actors should promote norms and accountability across the information ecosystem – engaging with stakeholders from technology, media and states to uphold humanitarian principles and protect people in need from harm.

While strong compliance systems support credibility and accountability, excessive bureaucracy and overly rigid procedures can erode trust, add unnecessary burden and weaken the human proximity that is central to humanitarian action. A 'right-touch' (i.e., striking the right balance) approach balances safeguards with flexibility, ensuring compliance reinforces – not replaces – principled and ethical judgement and humanitarian integrity.

States and other donors can support **right-touch compliance** by promoting due diligence that reinforces ethical judgement and humanitarian integrity, rather than imposing overly rigid procedures that slow responses or erode trust. Flexible frameworks ensure accountability while enabling principled, timely decision-making in complex operational and information environments.

Action point

- **Continuous:** Support compliance approaches that balance accountability with flexibility, enabling humanitarian actors to exercise principled and anticipatory decision-making.

8.6.3

Informational and digital literacy and capacities

Effective responses to harmful information require confidence and competence in navigating the information ecosystem and digital space, supported by strong internal capacities, technology access, strategic partnerships and shared standards. This shifts the focus from countering individual messages to understanding broader dynamics of how information is created, shared and trusted within communities, while emphasizing local media ecosystems, social trust and inclusive access to reliable information.

Staff and volunteers should be equipped to navigate digital environments responsibly, recognize emerging threats and engage constructively. Collaborative partnerships with technology actors, civil society and media organizations help promote safe and principled digital practices and advocate for accountability where harm occurs. Community-based digital literacy should be viewed as a protection strategy – empowering people to assess, challenge and contextualize information. Coordination across the Movement and with external actors is essential to share insights, align standards and amplify principled voices, as no single actor can address this challenge alone.

Co-creation with communities ensures they are active partners, not merely recipients, in shaping credible, context-appropriate responses. This approach considers five pillars: the **message** (content and framing), the **medium** (channels of transmission), the **audience**, the **actors** creating and circulating information, and the **impact** on people and systems.

In a post-truth era, staying principled is both an ethical and operational necessity, and digital literacy and access are core enablers of trust and risk management.

Information and digital literacy and access are now core enablers of trust and risk management. Limited literacy, especially at decision-making levels, undermines the ability to anticipate and respond effectively to harmful information. Without the skills, support and infrastructure to navigate today's complex digital information environment, even the most principled strategies risk being reactive rather than anticipatory.

Action points

- **Short term:** Integrate digital literacy, access and harmful information considerations into programme planning tools, risk matrices and community engagement strategies.
- **Medium term:** Establish cross-functional teams to design anticipatory strategies and pre-emptive messaging, and to embed digital monitoring skills.
- **Long term:** Promote sector-wide dialogue to align standards, share insights and amplify trust-based approaches to communication and engagement, backed by sustained investment in information and digital literacy and inclusive access.

8.6.4

Embed risk management in core systems

Risk assessments should address not only physical and reputational risks but also trust-related vulnerabilities, including perception gaps, harmful narratives and community backlash. Humanitarian diplomacy must reaffirm the relevance of humanity, neutrality, impartiality and independence, especially in contested environments.

Actions points

- **Short term:** Integrate trust and harmful information management and monitoring into risk frameworks and preparedness planning. Invest in systematic audience research, combining digital tools with in-person

engagement to ensure diverse perspectives shape humanitarian engagement and communication strategies.

- **Medium term:** Establish cross-functional teams to design anticipatory strategies to address potential information risks before they escalate and coordinate principled, pre-emptive messaging.
- **Long term:** Ensure sector and system-wide dialogue to align standards, share insights and amplify trust-based approaches to humanitarian risk frameworks.

8.6.5

Anticipation and integrated response strategies

A proactive response begins with anticipation and foresight. Humanitarian actors must invest in tools and practices that help them understand and anticipate harmful information, not just react to it. This includes scenario planning, risk mapping and early warning systems to help identify potential sources, narratives and impacts on access, trust and safety. Understanding triggers and enablers – such as grievances, power dynamics and moments of societal stress – supports more targeted preparedness and response. Frameworks like the ABCDE²⁰ approach can help map actors, messages, distribution mechanisms and effects, though they require analytical capacity and timely data access.

Anticipatory strategies must be community led or guided, locally relevant and adaptive to evolving digital threats.

Resilience demands more than reaction: it requires offensive, defensive and integrated proactive strategies connected in a cycle of protection and influence.

Actions points

- **Short term – defensive:** Rapid fact-checking, crisis communication, quiet corrections and escalation protocols. Develop standard operating procedures for safe use of rumour-tracking and verification platforms. Deploy two-way communication systems (SMS, voice, radio call-ins) for immediate rumour verification. Set up a sector-wide crisis communication taskforce on harmful information and rapid information-sharing protocols on harmful information in risk assessments and cluster mechanisms.
- **Medium term – offensive:** Prebunking, strategic storytelling, values-based campaigns, partnerships with local influencers and community co-creation. Train staff and volunteers on digital security and establish incident reporting protocols for platform misuse. Embed anticipatory messaging protocols into crisis communication plans to pre-empt harmful narratives.
- **Long term – integrated:** Link rapid response to narrative shaping, embed real-time monitoring and coordinate cross-functional teams so defensive and offensive approaches reinforce each other. Embed digital safeguarding and secure communication standards into organizational policies and donor frameworks. Scale and systematize real-time, two-way verification platforms as core infrastructure for community resilience.

8.6.6

Information diplomacy and norm setting

Addressing harmful information requires more than technical fixes – it also demands dialogue, restraint and shared norms. Just as ceasefire agreements limit the use of physical weapons, **harmful information diplomacy** seeks to create voluntary guardrails around the instrumentalization of information. This complements existing humanitarian and peacebuilding efforts by preventing escalation, fostering trust and creating space for principled humanitarian action.

Action points

- **Short term:** Facilitate offline dialogue mechanisms where conflict parties acknowledge the risks of harmful information and agree to red lines or voluntary codes of conduct. Use these as confidence-building measures in peace and mediation processes. Build awareness that ICT networks form part of critical civilian infrastructure and are protected under international humanitarian law. Advocacy efforts include awareness of connectivity gaps that disproportionately affect marginalized groups, including women, displaced persons and people with disabilities, who rely on communication networks for safety, services and participation.
- **Medium term:** Support the development and monitoring of voluntary codes of conduct for digital behaviour in weakly regulated contexts backed by civil society oversight and early engagement with platforms. Advocate for states to recognize ICT restoration in reconstruction plans and include digital access in humanitarian negotiations. Integrate digital safety and harmful information safeguards into all emergency connectivity responses.
- **Long term:** Strengthen local ownership of digital norms by ensuring communities most affected by harmful information are central to shaping, demanding and defending standards for responsible digital and offline behaviour.

8.6.7

Policy and governance for information resilience

Policy gaps remain a barrier to systematic action. While tools exist, most organizations lack dedicated harmful information policies and safeguarding standards for digital engagement. Escalation frameworks for harmful information, protocols for volunteer and staff protection, and standardized understanding of harms and impacts are urgently needed.

Action points

- **Short term:** Develop organizational harmful information policies with clear procedures for prevention, escalation and response. Identify and document impacts and harms through a standard taxonomy or framework of harms.
- **Medium term:** Establish safeguarding and risk mitigation standards for digital engagement, including data protection and safe use of SMS and rumour-tracking platforms. Encourage media outlets to adopt transparent charters on AI use, ensuring audiences understand how content is generated or assisted by technology.

- **Long term:** Adopt and institutionalize a harmonized harmful information impact measurement framework aligned with broader humanitarian indicators.

8.6.8

Research, evidence and partnerships

Harmful information is borderless and adaptive; tackling it requires evidence, innovation and collaboration. At present, most humanitarian actors document incidents only anecdotally or as part of broader communication or access challenges. This leaves significant gaps in evidence: the human, social and operational impacts and harms²¹ of harmful information remain under-measured compared to physical damage to lives, infrastructure or livelihoods. Without this evidence base, policy responses risk being reactive, fragmented or misaligned with humanitarian principles.

Greater investment is therefore needed in research, tools and partnerships that strengthen resilience at scale. Evidence turns anecdote into accountability. A robust understanding of the impacts of different forms of harmful information enables more effective advocacy with states, regulators and platforms, and helps ensure that humanitarian concerns are embedded in emerging governance frameworks. It also strengthens internal accountability by ensuring responses are data driven, anticipatory and principled.

Action points

- **Short term:** Map existing incident reporting systems across the sector and align them with harmful information categories (e.g., emblem misuse). Establish rapid reporting channels that capture not only what content spreads, but how it affects safety, access and community trust. Develop and share a cross-sector toolkit for humanitarian responders, including templates for counter-messaging, local risk assessments and training curricula on harmful information. Document and share evidence of its impacts and unintended positive outcomes. Provide thematic training for journalists on issues frequently targeted by harmful information, complementing core journalism standards. Ensure systematic engagement with local media and journalism.
- **Medium term:** Develop shared metrics to quantify the impact of harmful information on humanitarian outcomes (e.g., delays in aid delivery, reduced health service uptake, safety incidents). Pilot integration of harmful information indicators into needs assessments, early warning systems and programme evaluations. Invest in research and evidence on harmful information as a humanitarian risk, including its impact on trust, behaviour and access. Identify and pilot sustainable funding models for collaboration on monitoring, verification and moderation systems, including cross-sector partnerships.
- **Long term:** Institutionalize a global evidence base for harmful information, feeding into humanitarian diplomacy and policy advocacy. This could include an inter-agency repository of cases, impact studies and lessons learned to inform regulation, funding support and norms on information integrity. Build multi-stakeholder partnerships with states, platforms, media actors and community influencers to ensure approaches are locally

grounded but globally coordinated. Develop measurement frameworks that capture co-benefits (e.g., peacebuilding) alongside humanitarian outcomes.

8.7 Prerequisites and cross-cutting enablers

Building resilience to harmful information depends not only on trust, compliance, policy and partnerships but also on a set of enabling conditions that determine whether recommendations translate into practice. Three enablers stand out:

8.7.1

Crisis communication preparedness

Effective responses to harmful information depend on preparedness before a crisis hits with established crisis communication structures with clear roles, escalation protocols and links to staff and volunteers. Regular environmental scanning and decision-making frameworks (when to engage publicly, prioritizing internal communication) ensure rapid, coordinated responses that balance transparency with risk management.

In today's complex information environment, communication is not merely a support function – it is a critical enabler of principled humanitarian action. Moving from reactive to strategic communication enables actors to shape the information environment through transparency, inclusion and dialogue – reinforcing trust and community resilience. Strategic and context-sensitive communication helps to safeguard humanitarian space, sustain access and build trust with communities. When grounded in sound analysis and principled practice, it can prevent the escalation of tensions and reduce harm through effective message framing, audience engagement and dialogue. The aim is to promote alternative narratives, enhance community resilience to harmful information and foster information and media literacy. This goes beyond just correcting falsehoods and aims to change behaviours.

This requires an understanding of local drivers and triggers of harm, as well as investing in trusted, locally anchored engagement. Volunteers, including digital volunteers, can serve as early responders in the information ecosystem to detect and respond to emerging narratives, offering scalable, community-based interventions. This requires solid engagement, internal communication and support. Mapping influencers and narrative dynamics can build understanding of who shapes opinion in specific contexts and how that influence can be used constructively in support of humanitarian response.

Timely, inclusive and principled communication is essential. While information is not water or shelter, it often determines how – and whether – those needs are met and the basis on which people make decisions.

8.7.2

Standardized tools and frameworks

Fragmented responses increase vulnerability. Information Resilience or Information Integrity Toolkits (e.g., Movement Safer Access Framework (being updated), UNHCR

Information Integrity Toolkit (2025),²² ICRC Framework²³ (2025), IFRC Organizational Capacity Assessment and Certification²⁴) demonstrate how structured resources can provide a common approach to prevention, detection, escalation and response. Standardized, adaptable tools which could be socialized cross-sector and with digital rights groups, academia and the private sector to help humanitarians act faster and more consistently across contexts.

8.7.3

Resourcing and professionalizing RCCE

Risk communication and community engagement (RCCE) is not just an add-on service for responses to disasters and emergencies – it enables and drives community-centred and evidence-informed responses by placing the needs, feedback and realities of communities at the forefront.

In times of crisis, RCCE has proven to be not just a tool but a lifeline to protect the health and well-being of communities. Placing the needs, feedback and realities of communities at the forefront of responses strengthens accountability, improves public confidence in response efforts and minimizes the impact of risks on the lives, livelihoods and well-being of those affected. Failing to integrate RCCE and wider community protection technical areas from the outset of a response contributes to miscommunication, mistrust, poor compliance with public health guidance – all of which ultimately lead to a slower and less effective response.

RCCE capacities must be embedded and invested in as a foundational element of disaster and emergency preparedness, so that the systems, capacities and expertise can be leveraged when crisis strikes. This includes:

- Trained, multidisciplinary teams capable of collecting, analysing and applying social science evidence and feedback in real time.
- Proactive community partnerships that build trust, foster co-creation of solutions and ensure marginalized voices are heard.
- Robust feedback loops so that community concerns and insights are systematically captured, analysed and visibly acted upon.

By institutionalizing these elements, responses move from being reactive to systematic, evidence-driven and locally anchored, ultimately strengthening both the effectiveness and legitimacy of emergency response.

Without these enablers, the eight roadmap pillars risk being applied unevenly or only after harm is done. With them, humanitarian actors can move from fragmented, reactive responses to system-wide resilience: prepared, equipped and embedded in communities before, during and after crises.

8.7.4

Harnessing influence for humanitarian resilience

Resilience to harmful information depends on enabling trusted voices to carry principled narratives. Volunteers and frontline staff remain the strongest enablers, embedded in communities and building trust locally. Community leaders, youth networks, etc. amplify this trust and diaspora groups extend influence across borders. Independent media and

journalists provide credibility through accurate reporting, and digital influencers and activists offer reach into fragmented digital spaces, though with risks for neutrality. Above all, technology platforms act as 'meta-influencers', shaping which voices are amplified or suppressed. Advocacy for transparency and rights-respecting governance are vital.

Ultimately, it is **trust, not reach, that turns influence into resilience**, safeguarding humanitarian space and ensuring that communities can act on accurate information.

The IFRC's **Community Trust Index** offers an evidence-based framework to measure, track and enhance trust between humanitarian organizations and the communities they serve. It assesses community perceptions across two dimensions: competence (technical skills, effectiveness, relevance) and values (integrity, transparency, participation) while also identifying enablers and barriers to trust in areas such as early warning systems, climate resilience, migration and public health. Other organizations and governments could adopt or adapt the index to generate structured, standardized data that complements real-time community feedback. This combination provides a reliable baseline and trend analysis to inform strategic decision-making.

Looking ahead, the Community Trust Index could be further strengthened by integrating targeted focus on harmful information. Doing so would allow organizations to: track the spread and impact of rumours systematically, identify community segments most affected, and bridge qualitative feedback with quantitative insights. By evolving in this way, the index can help anticipate challenges and co-create tailored solutions with communities. Ultimately, the Community Trust Index is then more than a measurement tool – it is a call to action. By diagnosing trust gaps and empowering communities as partners, it equips organizations to rebuild and sustain trust well beyond crises.

8.7.5

Systemic support needs

Expertise and experiences highlight that addressing harmful information cannot rely on isolated projects or short-term fixes. Sustained resilience requires system-level investment, coordination and safeguards. Several gaps stand out:

- **Technical and analytical tools:** Humanitarian actors need the ability to monitor both online and offline spaces, supported by data visualization, narrative and rumour trend mapping and predictive analytics to anticipate harmful narratives before they spread.
- **Monitoring and evaluation frameworks:** Impact must be measured not only in outputs (e.g., messages delivered) but also in behavioural shifts, perception change and trust dynamics at community level.
- **Security protocols:** Dedicated standards are needed for secure data handling and staff safety in contexts where armed groups or political actors monitor or exploit information systems.
- **Sustainable financing:** Rumour-tracking and verification systems are resource-intensive. Without multi-year, flexible donor support, community capacity and infrastructure cannot be maintained or scaled. There is real potential for cross-sector collaboration to benefit from economies of scale.

- **Policy and coordination guidance:** Clear roles, standards and escalation pathways are required within humanitarian coordination mechanisms to avoid fragmented or reactive responses.
- **Integration of local media:** Community radio stations and local journalists should be actively engaged as frontline responders to harmful information, co-producing and distributing trusted, localized counter-messages.
- **Digital inclusion and platform accountability:** Platforms must adapt their tools to humanitarian realities, providing low-bandwidth, multilingual and locally contextualized solutions while ensuring transparency and safeguards.
- **Ecosystem-wide literacy:** Media and rumour literacy should be extended beyond technical staff to volunteers, health workers, teachers and community members, embedding resilience across whole populations.
- **Anticipatory systems:** Predictive modelling and pre-bunking strategies – delivered via SMS, posters or radio ahead of flashpoints such as elections or vaccination drives – can reduce the space for harmful rumours to take hold.

Together, these support needs point to a critical shift: from isolated interventions to systemic resilience, where humanitarian actors, governments, donors, platforms and communities co-invest in shared infrastructure for trustworthy information.

Conclusion: Together, we can uphold and reclaim space for humanity

Trust is not assured; it is built, reinforced and renewed. In the face of harmful information, it remains the most powerful safeguard for humanitarian space. By investing in trust, embedding right-touch compliance, addressing policy gaps and advancing evidence-based partnerships, the humanitarian sector can shift from reactive counter-narratives toward systemic resilience.

Harmful information cannot be addressed piecemeal. Building resilience requires trust at the centre, compliance that enables rather than obstructs, integrated risk management, stronger policies, deeper evidence and partnerships – and above all, proactive strategies that connect defensive and offensive efforts. Silence and delay carry their own dangers; early, transparent and trusted communication and engagement is the most powerful safeguard for humanitarian space. Governments need to act to preserve this humanitarian space.

The humanitarian sector cannot afford to cede the information space. While the speed, scale and sophistication of harmful information poses significant challenges, disengagement is not a viable option. To maintain access, credibility and principled impact, humanitarian actors must engage deliberately grounded in the humanitarian principles,

supported by collaboration and partnerships, and informed by a clear understanding of the social dynamics shaping today's contested narratives.

Reclaiming narrative space requires more than correcting harmful information. It calls for a reframing of communication as connection, rooted in listening, empathy, proximity, humility and consistency. It also requires organizations to understand how harmful information spreads and why people believe or accept it, addressing not just the content of falsehoods but the emotions, fears and identities that give them power.

In times of crisis, harmful information thrives by reducing complex realities into simplistic explanations and easily identifiable enemies. Principled humanitarian action must not only speak truth, it must also understand fear, identity and belonging. In this fragmented, emotionally charged landscape, reclaiming narrative space is not about controlling the story. It is about restoring trust, rebuilding connection and reasserting the relevance of humanitarian principles and action in the eyes of affected communities and the broader public. Governments have an important role in this regard to reinforce the importance of principled humanitarian action and preserving the space for humanitarian organizations to operate.

The Movement's Resolution on Tolerance offers a valuable foundation to reinvigorate efforts against harmful information – reminding us that respect, diversity and non-discrimination are not abstract ideals but practical tools to reduce polarization, counter dehumanizing narratives and preserve humanitarian space. Tolerance online also has its limits: “what is often framed as a fight over *speech* is actually a fight over *reach*” (DiResta²⁵) – the algorithmic amplification that determines which voices are elevated, repeated and made unavoidable. The humanitarian sector must advocate for changes that reduce the reach of hate speech and malicious content that imperil humanitarian action and endanger staff, volunteers and crisis-affected populations.

The scale of the challenge may feel overwhelming, but it need not paralyse us. By acting collectively – rooted in principles, grounded in trust and united across communities, states, platforms and humanitarian organizations – the sector can move forward with courage and clarity.

The intersection of resilience, principled communication and institutional credibility is now a critical arena for action. Humanitarian actors must shift from a defensive posture to proactive, adaptive and systematic approaches that prepare for and mitigate the impact of harmful information, particularly where it threatens access, undermines trust or endangers lives. Delivering this requires a cross-functional approach in collaboration with communities. **The right to information is contested, but in humanitarian crises, the need to know elevates information to a core element of response.**

Together, we can uphold and reclaim space for humanity.

Endnotes

- 1 IFRC. *Community Engagement and Accountability Toolkit*. (2020) <https://communityengagementhub.org/resource/cea-toolkit>
- 2 UN. *Our Common Agenda: Report of the Secretary-General*, UN Doc. A/75/982 (2021) p.27 www.un.org/en/content/common-agenda-report/
- 3 Levitin, DJ. *A Field Guide To Lies: Critical Thinking in the Information Age*. (2016) He highlights that we need to use just some of the time we saved in information acquisition to perform proper information verification, p.253
- 4 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) p.261
- 5 Kreps, S. *Democratizing Harm: Artificial Intelligence in the Hands of Non-state Actors*. Foreign Policy at Brookings, (2021) pp.2–5 www.brookings.edu/articles/democratizing-harm-artificial-intelligence-in-the-hands-of-non-state-actors
- 6 World Economic Forum. *Global Risks Report 2025*. (2025) www.weforum.org/publications/global-risks-report-2025/digest. The report identifies the convergence of threats – including geopolitical, environmental, societal and technological risks – as a defining feature of today's global risk landscape.
- 7 Edelman. *2024 Edelman Trust Barometer: Global Report*. (2024) The report notes that media became the least trusted institution in 2020, while social media has been the least trusted source of news and information since 2016. www.edelman.com/trust-barometer
- 8 Luhmann, L. *Trust and Power*. (John Wiley and Sons, 1979)
- 9 Singer, PW. and Brooking, ET. *LikeWar: The Weaponization of Social Media*. (2018) p.186
- 10 ICRC. *Building a Responsible Humanitarian Approach: The ICRC's Policy on Artificial Intelligence*. (2024) www.icrc.org/en/publication/building-responsible-humanitarian-approach-icrcs-policy-artificial-intelligence
- 11 CDAC Network, The Alan Turing Institute and Humanitarian AI Advisory. *SAFE AI: Standards and Assurance Framework for Ethical Artificial Intelligence in Humanitarian Action*. (2025) www.cdacnetwork.org/safe-ai
- 12 Kaspersen, A. and Cavelti, MD. Digitalisation of Conflict, GESDA Science Breakthrough Radar. p.3
- 13 UN. *Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security: Final Substantive Report*. UN Doc. A/75/816 (2021) <https://digitallibrary.un.org/record/3906060>
- 14 ICRC. *Addressing Harmful Information in Conflict Settings: A Response Framework for Humanitarian Organizations* (2024) <https://shop.icrc.org/addressing-harmful-information-in-conflict-settings-a-response-framework-for-humanitarian-organizations-pdf-en.html>
- 15 University of Melbourne. *City Playbook for Countering Disinformation*. Melbourne Centre for Cities and Centre for Artificial Intelligence and Digital Ethics. (2022) www.unimelb.edu.au/cities/research/projects/current-projects/disinformation-in-the-city/disinformation-in-the-city-response-playbook
- 16 Defined as 'wars' or 'limited wars' by the Heidelberg Institute for International Conflict Research.
- 17 Zameh. '#ReconnectGaza: A global campaign to restore connectivity in Gaza.' News. Association for Progressive Communications. 26 February 2025. www.apc.org/en/news/reconnectgaza-global-campaign-restore-connectivity-gaza; York, JC. 'Connectivity is a Lifeline, Not a Luxury: Telecom Blackouts in Gaza Threaten Lives and Digital Rights.' Electronic Frontier Foundation. 16 June 2025. www.eff.org/deeplinks/2025/06/connectivity-lifeline-not-luxury-telecom-blackouts-gaza-threaten-lives-and-digital
- 18 Johns Hopkins University, Imperial College London and Georgia Institute of Technology. *Countering Disinformation: Improving the Alliance's Digital Resilience*. NATO Review. (2021) ch.1 <https://archives.nato.int/nato-review-countering-disinformation-improving-the-alliance-s-digital-resilience>. The article discusses AI-based sentiment analysis, emotional predictors, and automated 'circuit-breakers' to slow the virality of harmful content while protecting freedom of expression.
- 19 Ibid
- 20 A – Actor, B – Behaviour, C – Content, D – Degree, E – Effect. Pamment, J. *The EU's Role in Fighting Disinformation: Crafting A Disinformation Framework*. (2020)
- 21 ICRC. *Addressing Harmful Information in Conflict Settings: A Response Framework for Humanitarian Organizations*. (2025). The ICRC determines that decisions about if, when and how to respond to harmful information are based on the harm potential, spread potential of information and associated risk indicators, p.12; Lindsey, C. and Glasser, G. *Report of Second Expert Meeting on the Development of a Harms Methodology*. CyberPeace Institute. (2024) pp.5–6 defines harm as: "an impairment or disruption of an entity's capacity or ability to function and exist as it otherwise would have in its usual context". This definition identifies four levels of harm affecting individuals, organizations, societies and international peace and security. See also 'UN High-Level Advisory Body on Artificial Intelligence. *Governing AI for Humanity: Final Report*. (2024) cited in Paris Peace Forum. *Forging Global Cooperation on AI Risks: Cyber Policy as a Governance Blueprint*. (2025). This highlights that a strong emphasis on risks and harms should be at the centre of AI governance "that focuses on who is at risk and accountable, and not just what is at risk".
- 22 UNHCR. *Information Integrity Toolkit*. (2025) www.unhcr.org/handbooks/informationintegrity
- 23 ICRC. *Addressing Harmful Information in Conflict Settings: A Response Framework for Humanitarian Organizations*. (2024) www.icrc.org/en/publication/addressing-harmful-information-conflict-settings-response-framework-humanitarian
- 24 The OCAC is a comprehensive process that helps National Societies review their capacity and performance. It enables them to identify strengths and weaknesses, focus efforts to become strong and sustainable service providers, and measure themselves against the minimum standards expected of modern humanitarian and development organizations. IFRC. *Organizational Capacity Assessment and Certification*. (2019) <https://data.ifrc.org/en/ocac>. See also Policies and key commitments. www.ifrc.org/our-promise/trust-and-accountability/policies-and-key-commitments
- 25 DiResta, R. *Invisible Rulers: The People Who Turn Lies into Reality*. (2024) p.318





Annex I



Glossary

Algorithmic amplification

The process by which recommendation systems, ranking algorithms or automated curation tools on digital platforms prioritize and boost the visibility of certain content over others. This amplification is typically based on engagement metrics (e.g., clicks, shares, watch time), personalization or platform incentives, rather than the accuracy or trustworthiness of the information.

AI-generated fake content

Audio, text or visual content, produced by generative AI (see below), that depicts people or events in a malicious or deceptive way that differs from reality. Examples include showing people doing things they did not do, saying things they did not say, changing the location of real events, or depicting events that did not happen.

AI-generated 'slop'

Refers to low-quality media, such as text, images or videos, created using generative artificial intelligence tools and characterized by mass production, minimal effort and lack of meaningful substance or artistic integrity. The term is a pejorative implication similar to 'spam,' highlighting superficial or banal AI outputs often produced solely for volume or engagement.

Artificial intelligence (AI)

The capability of machines, computer systems or software to perform tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, natural language understanding and decision-making.

Cheapfake

Manipulated or misleading media created using simple, low-cost or readily available editing tools.

Cognitive warfare

The use of information, psychological and technological tools to influence, disrupt or control the perceptions, beliefs and decision-making of individuals or groups. It targets the human mind as the contested domain, aiming to weaken trust, sow confusion and manipulate behaviour in ways that serve strategic or political objectives.

Community engagement and accountability (CEA)

A way of working that recognizes and values all community members as equal partners, whose diverse needs, priorities and preferences guide the IFRC's work.

This includes integrating meaningful community participation, open and honest communication, and mechanisms to listen to and act on feedback.

Confirmation bias

A cognitive bias whereby people tend to seek, interpret and recall information in ways that confirm their pre-existing beliefs, while discounting or overlooking contradictory evidence.

Conspiracy theory

An explanatory belief that attributes the cause of an event, phenomenon or social condition to the secret actions of powerful groups acting in concert for malevolent purposes. Conspiracy theories are typically resistant to falsification (evidence against them does not disprove them, instead, it often gets absorbed into the theory), simplify complex realities into intentional plots, and often portray the world as controlled by hidden actors.

Content moderation

The process by which platforms, organizations and institutions monitor, review and manage user-generated content to ensure it complies with legal requirements, community standards and safety guidelines. Moderation can involve the removal, labelling or restriction of harmful, illegal or policy-violating material, as well as the promotion of accurate and trustworthy information.

Coordinated inauthentic behaviour

A term introduced by Meta (Facebook) to describe organized efforts where groups of accounts or pages work together to mislead people about who they are and what they are doing. It is not defined by the content of the messages but by the deceptive, coordinated and concealed nature of the actors and their activities. It often involves fake accounts, covert networks or hidden sponsorship, designed to manipulate public debate, amplify narratives or influence political outcomes.

Cybercriminals

Individuals or organized groups who use computers, networks, or digital technologies to commit crimes. Their activities include unauthorized access to systems, theft of data, identity fraud, financial scams, ransomware attacks and the distribution of malicious software. Cybercriminals may act independently, as part of loosely connected online networks, or within highly organized transnational criminal enterprises.

Debunking

A reactive strategy that addresses misinformation after it has begun to

circulate. Its effectiveness depends on being timely, clear and delivered by trusted messengers. Effective debunking not only identifies a claim as false but also explains why it is false. Communication techniques such as the 'truth sandwich' – which begins and ends with accurate information while addressing the falsehood in the middle – help to minimize the risk of reinforcing the myth.

Decontextualization

The practice of presenting accurate information, images or quotes outside their original context in a way that alters their meaning or implications. By stripping away crucial details such as time, place, source or intent, decontextualization can mislead audiences, distort events and reinforce false narratives, even without fabricating new content.

Deepfake

Synthetic media – most often video, audio or images – created using AI techniques, particularly deep learning, to realistically manipulate or generate content that portrays events or people in ways that did not actually occur.

Demystification

The process of unpacking how misinformation operates. It entails explaining its forms and functions, as well as the psychological, social and technological triggers that drive its spread. Unlike fact-checking, which focuses on verifying the truth or falsehood of individual claims, demystification highlights the structural, emotional, algorithmic and contextual factors that influence the information people encounter, thereby fostering a deeper understanding of why misinformation is persuasive and persistent.

Disinformation

False information that is deliberately created or spread with the intention to deceive or cause harm.

Doxing (or 'doxxing')

The intentional gathering and publishing of someone's private or personally identifiable information without their consent, typically carried out to shame, embarrass, harass, intimidate, threaten or cause harm.

Fact-checking

The process of verifying the factual accuracy of information, claims or statements, usually in journalism, public communication or online content. It involves systematically evaluating evidence, consulting credible sources

and providing transparent corrections when false or misleading claims are identified. Fact-checking can be proactive (anticipating claims) or reactive (debunking after claims spread).

False dichotomy

A logical fallacy that presents a situation as having only two opposing options or outcomes, when in reality there are additional possibilities. By reducing complex issues to a simplistic 'either/or' choice, false dichotomies distort reasoning, polarize debate and limit consideration of alternative perspectives. For example: "You're either with us or against us."

Farm

A 'farm' in the misinformation ecosystem refers to an organized operation that systematically produces, amplifies or manipulates content at scale. **Content farms** typically pursue profit by generating click-driven misinformation, while **troll farms** are politically or strategically motivated, aiming to distort public debate and influence opinion.

Filter bubble

A state of intellectual or informational isolation created by personalized algorithms that curate online content based on a user's past behaviour, interests and preferences. Within a filter bubble, individuals are primarily exposed to information and viewpoints that reinforce their existing beliefs, while alternative perspectives are downplayed or excluded. This phenomenon can limit diversity of information, strengthen confirmation bias and contribute to polarization.

Foreign information manipulation and interference

A pattern of behaviour that threatens or has the potential to negatively impact values, procedures and political processes. Such activity is manipulative in character, conducted in an intentional and coordinated manner, by state or non-state actors, including their proxies in and outside their own territory. (Also known as FIMI.)

Generative AI

A class of AI systems that can create new content, such as text, images, audio, video or code, based on patterns learned from large datasets. Unlike traditional AI systems designed primarily for classification or prediction, generative AI models are capable of producing novel outputs that resemble human-created work, often through deep learning architectures such as generative adversarial networks or large language models.

Hacktivists

A hacktivist is an individual or collective that uses hacking techniques as a form of political or social activism. Hacktivism combines 'hacking' and 'activism' to pursue goals such as promoting free information, protesting government or corporate actions, or drawing attention to social causes. Tactics may include website defacement, denial-of-service attacks, data leaks and hijacking online accounts.

Harmful information

Information that has the potential to cause, contribute to or result in harm to an individual or entity.

Hashtag hijacking

A reactive or strategic misuse of a trending or branded social media hashtag. It occurs when users adopt a hashtag intended for a specific topic or campaign and repurpose it to promote a substantially different or conflicting message – often involving spam, trolling, political agendas or conspiracies. This diversion can disrupt the original conversation and obscure or distort the intended content.

Hate speech

Any kind of communication in speech, writing or behaviour that attacks or uses pejorative or discriminatory language with reference to a person or group based on their identity, such as religion, ethnicity, nationality, race, colour, descent, gender or other factor.

Illusory truth effect or familiarity effect

A cognitive bias in which repeated exposure to a statement increases the likelihood that it will be perceived as true, regardless of its factual accuracy. This effect arises because familiarity enhances processing fluency – the ease with which information is mentally processed – which people often mistake for accuracy.

Infodemic

An overabundance of information, including accurate, false and misleading content, circulating in digital and physical environments during a crisis, which makes it difficult for people to find trustworthy sources and reliable guidance.

Infodemic management

The systematic response to these challenges – using tools like social listening, debunking and prebunking to manage the information environment and support public health decision-making.

Information integrity

Refers to the reliability, accuracy and trustworthiness of information within an ecosystem. It encompasses efforts to safeguard information environments from manipulation, misinformation, disinformation and other harmful content, ensuring that people can access factual, contextual and credible information to make informed decisions.

Information operations

Coordinated activities that use information, communication and influence tactics to affect the perceptions, decisions and behaviours of target audiences in support of political, military or organizational objectives. They often combine psychological operations, cyber activities, propaganda and disinformation campaigns to achieve strategic effects, both online and offline.

Inoculation theory

This proposes that people can build resistance to persuasion in a manner similar to medical inoculation: by being exposed to a weakened form of an argument, combined with refutations, they develop 'mental antibodies' that help them resist stronger attacks later. In the context of harmful information, inoculation theory underpins prebunking strategies – educational interventions that expose people to examples of manipulation techniques (such as scapegoating, conspiracy framing or emotional appeals) before they encounter them in real life. Research shows that inoculation can improve people's ability to detect and reject misinformation in public health, humanitarian and political contexts.

Lone wolf

An individual who plans and carries out violent or disruptive actions independently, without direct operational support from an organization, even if they are ideologically inspired by one. In the information domain, the term also extends to individuals who act alone in spreading extremist or conspiratorial content online, amplifying harmful narratives without belonging to a coordinated network.

Malinformation

Genuine information shared with the intent to cause harm – for example, by taking something true out of context or using it to discredit someone.

Misinformation, disinformation and hate speech

An umbrella term that refers to different types of false, misleading or damaging information. It includes misinformation (false information shared without

intent to harm), disinformation (false information deliberately created or spread to deceive or cause harm) and hate speech. (Also known as MDH.)

Misinformation

False or inaccurate information that is shared without intent to deceive.

Open-source intelligence

The collection, analysis and use of information that is publicly available and legally accessible. Sources include traditional media, academic publications, government reports, social media, satellite imagery and other open data. Open-source intelligence is widely used by governments, law enforcement, journalists, researchers and civil society to support decision-making, investigations and situational awareness.

Political manipulation

The deliberate use of deceptive, coercive or manipulative tactics by political actors to shape public perceptions, attitudes and behaviours in ways that serve their own interests, often at the expense of informed democratic decision-making. It can involve controlling narratives, spreading misinformation, exploiting emotional appeals, suppressing dissent or engineering consent through covert influence operations.

Prebunking

A proactive method aimed at building resilience to misinformation by exposing people – in advance – to weakened examples of misleading strategies or false claims, coupled with refutations. By leveraging inoculation theory, it prepares individuals to better recognize and resist manipulative content when they encounter it later. There are two types of prebunking: **narrative inoculation** counters broader storylines or persuasive frames that might be used to mislead, while **tactical inoculation** focuses on techniques of manipulation rather than the storyline itself, e.g., teaching people how tactics like emotional manipulation, decontextualization of data or scapegoating can make them more resistant when they encounter them. Together, these build ‘mental immunity’ – the cognitive resilience that helps people identify, resist and discount misinformation before it takes root.

Pre-emption

Refers to proactive strategies that anticipate misleading claims, narratives or harmful information before they spread, with the aim of reducing their impact.

Propaganda

The systematic and deliberate use of communication – through symbols, messages or media – to shape perceptions, manipulate cognition and direct behaviour to achieve the objectives of its sponsor. It is typically selective in its presentation of facts and may appeal to emotions rather than rational analysis.

Proxy

An intermediary actor – such as an organization, media outlet or individual – used by a state or group to conduct activities indirectly on its behalf. In information operations, proxies amplify or disseminate narratives while concealing the true origin or sponsor of the content. Proxies may act knowingly (aligned actors) or unknowingly (‘useful idiots’), providing plausible deniability for the real orchestrators.

Resilience

The ability of individuals, communities, organizations or countries exposed to disasters, crises and underlying vulnerabilities to anticipate, reduce the impact of, cope with and recover from the effects of shocks and stresses without compromising their long-term prospects.

Rumours

Unverified information that spreads quickly within communities. Rumours may later prove to be true, false or partially true, but they can still cause confusion or mistrust while circulating.

Satire-as-truth

Refers to situations where parody or satirical content is mistaken for genuine information. When satire circulates without its original humorous or ironic context, such as memes, parody news articles or comedy sketches, audiences may interpret it literally and spread it as fact.

Scams and fraud (related to aid)

Deliberate acts of deception aimed at gaining money, resources or sensitive

information, often targeting vulnerable populations in crises. These can include fake aid offers, fraudulent fundraising or misrepresentation by individuals or organizations. Disinformation often amplifies these schemes, creating confusion, urgency or mistrust among communities and donors.

Social listening

The systematic process of monitoring, collecting and analysing conversations, trends and mentions across digital and social media platforms to gain insights into public perceptions, behaviours and narratives. Unlike basic social media monitoring, which tracks metrics (likes, shares, mentions), social listening emphasizes qualitative analysis of context, sentiment and emerging issues.

Sock puppet account

A false online identity created and controlled by a person or organization to deceive others.

Synthetic content

Refers to text, images, audio, video or other media that is generated, manipulated or modified using AI or algorithmic systems, rather than being directly created or recorded by humans.

Troll

An individual who deliberately disrupts online conversations, communities or platforms by posting inflammatory, off-topic or deceptive content with the intent to provoke, mislead or manipulate others. In misinformation contexts, trolls may operate independently or as part of organized campaigns (‘**troll farms**’), amplifying divisive narratives or spreading false information for political, financial or ideological purposes.

Visual manipulation

Changing or misusing photos and videos – by editing, cropping or adding false captions – to trick people or make them believe something that is not true.

Watermarking

A technique used to embed a marker (visible or invisible) during digital content creation, such as text, images, audio, video or AI-generated media, to indicate its origin, authenticity or ownership.





Annex II



The role of data in humanitarian decision-making

A2.1 Why data matters: Informing decisions, guiding action

The first questions asked when assessing a disaster or crisis are ‘how big/bad’, ‘where’ and ‘to whom’ the impacts will be felt. To answer these questions, we need data. In recent years, the volume of data produced to help answer these questions has mushroomed. As a result, the ability to build systematic methods for allocating resources, targeting assistance and anticipating crises has improved rapidly. Such methods enable humanitarian organizations to leverage evidence to guide our principles and rules with the aim of providing assistance where it is needed most.

Building an evidence base for such principled decision-making requires transparency on what the data does – and does not – show; how it has been collected, produced or transformed; and how timely and reliable it is, based on previous experience. The data presented and analysed in this report has been collected into a database of databases – the Montandon Global Crisis Data Bank – which is designed as the world’s largest open source, open access repository for data on hazards, their impacts and the responses mobilized to respond to them.

One of the key motivations behind the development of the Montandon is to serve as a reference point for global trends, as presented below. In addition, we collect such a wide variety of data to help us triangulate, establish how much confidence we should have in forecasts, and curate the most useful information for our IFRC network and beyond.

We are grateful to many partners who have collaborated to build the Montandon, and for their tireless work in collecting data in the field and at home to serve as inputs into our collective intelligence.

A2.2 Recent trends

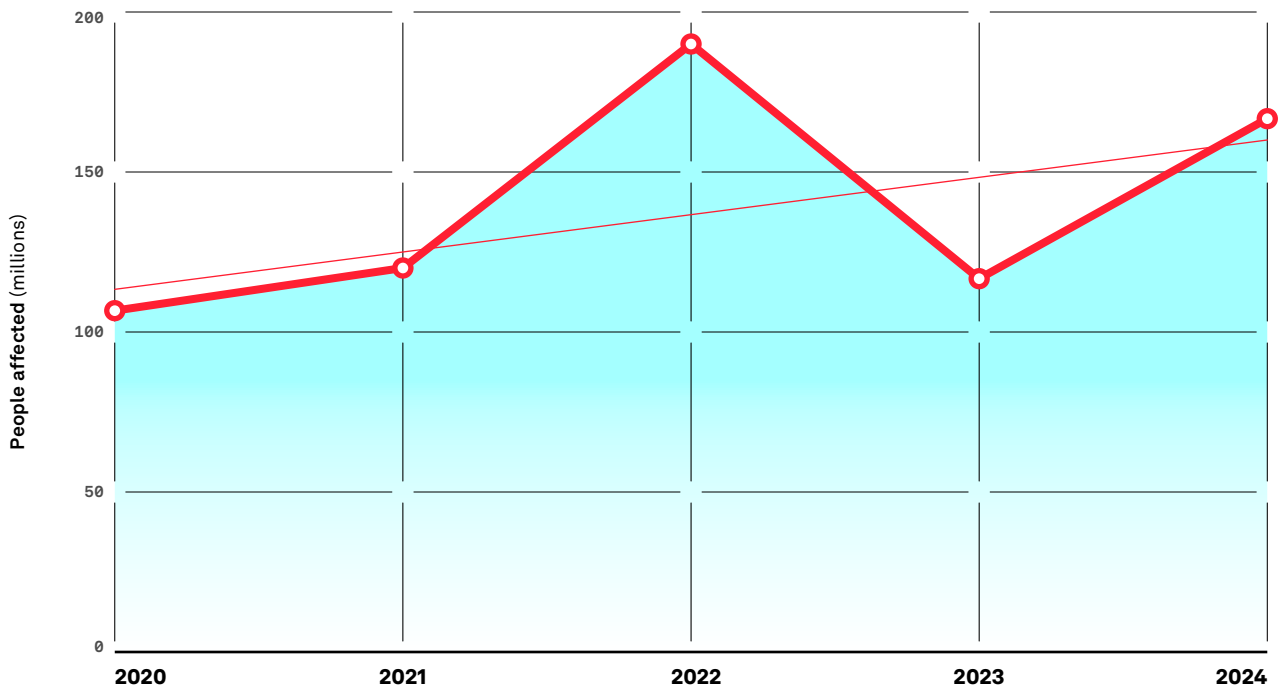
The data from 2020 to 2024 provides an illustrative snapshot of what we know, how we know it, and how data is being used to shift towards more anticipatory approaches to humanitarian crises.¹

Looking across five years of crises and impact data, we focused on answering these key questions:

- How many events are occurring?
- Where are disasters occurring?
- How severe are they?
- How do communities respond – or take anticipatory action – with or without support from international actors?
- Which of these questions can we answer confidently and with what data?

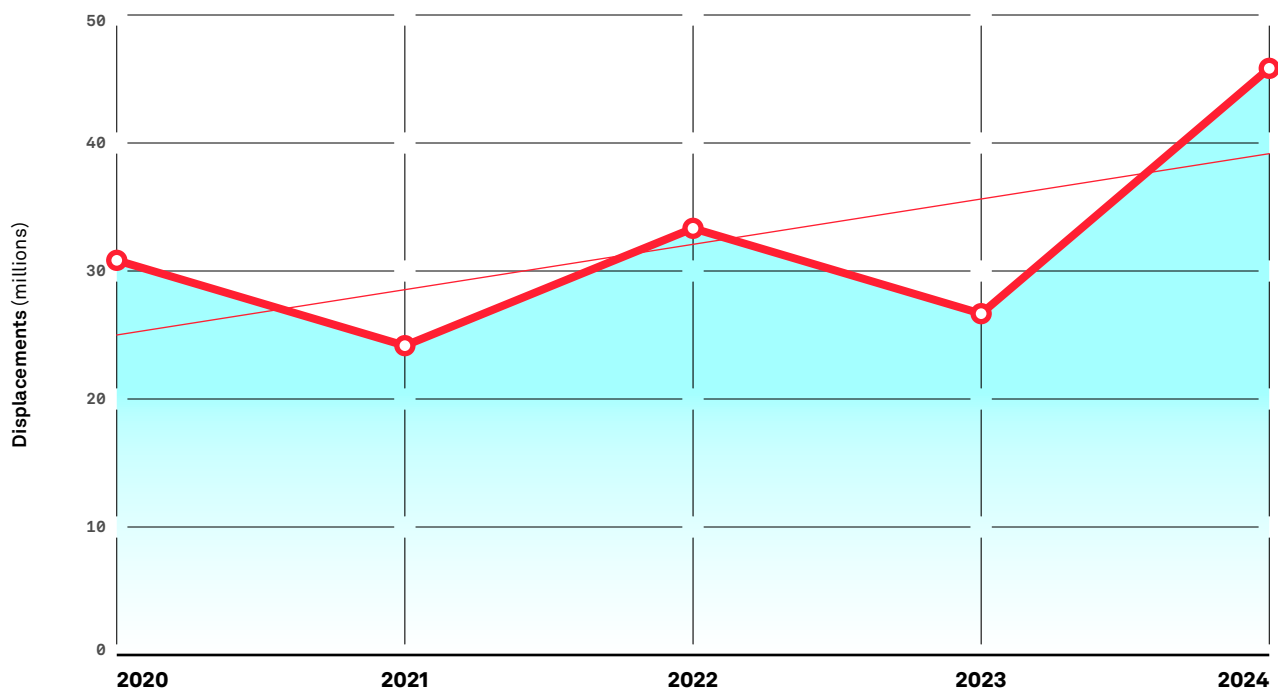
The number of people affected by disasters has steadily increased between 2020 and 2024 across two metrics: people affected (Figure A2.1.a) and displacements (Figure A2.1.b).

Fig A2.1.a Number of people affected by disasters per year, 2020–2024



Source: Centre for Research on the Epidemiology of Disasters (CRED) Emergency Events Database (EM-DAT)

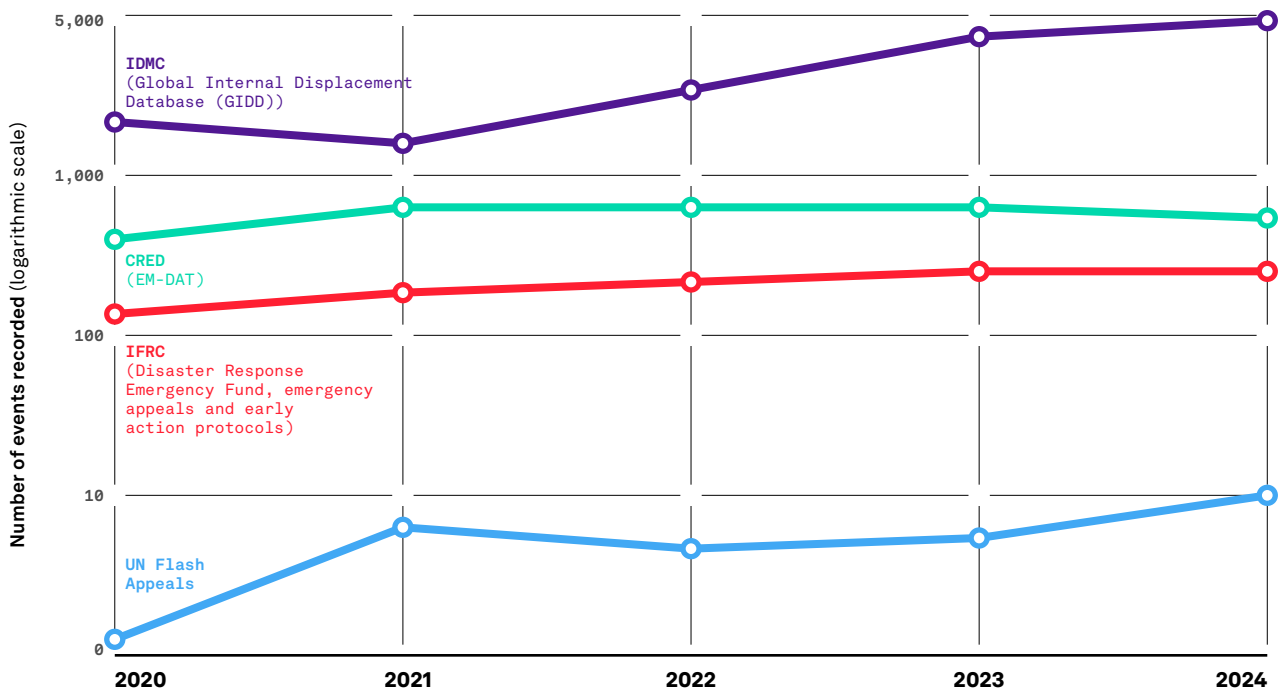
Fig A2.1.b **Number of disaster displacements recorded per year, 2020–2024**



Source: IDMC

These increases reflect both a rise in the number of events (disasters) and improvements in how they are recorded. The Internal Displacement Monitoring Centre (IDMC), for example, recorded 14,348 individual disasters between 2020 and 2024 (inclusive) (Figure A2.2).²

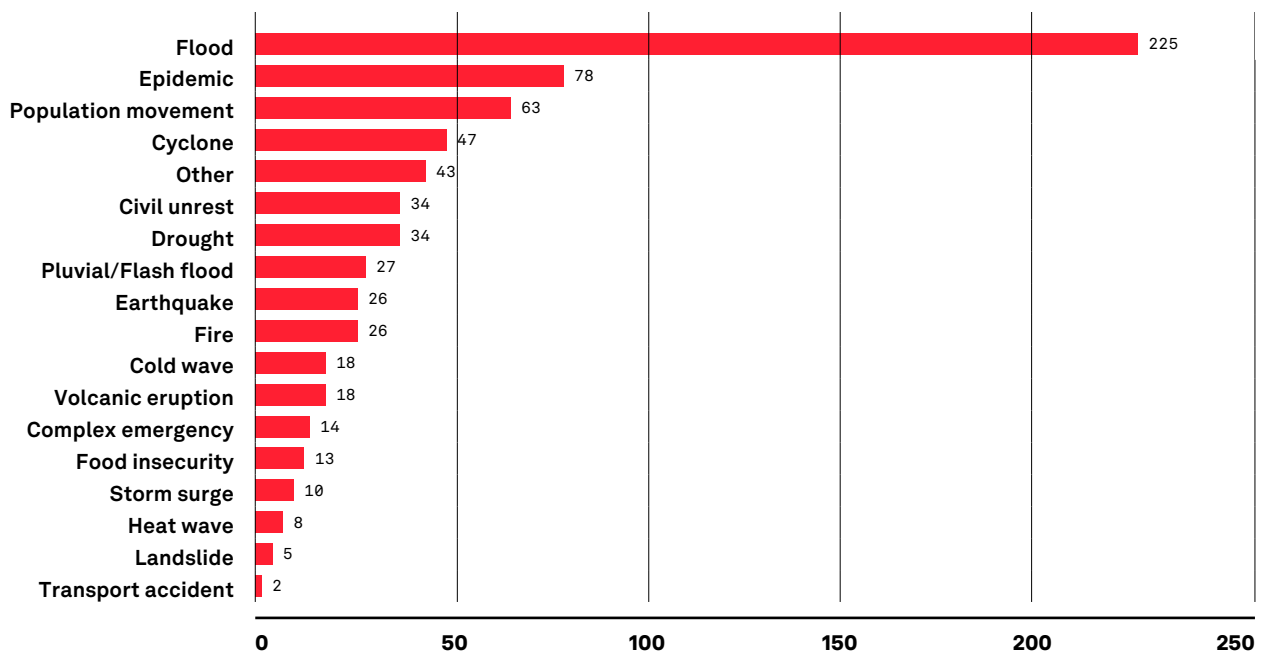
Fig II.2 Number of disasters recorded per year, 2020–2024



Source: Montandon and OCHA Financial Tracking Service

The data also shows that most disasters are being managed through the capacities of national and local actors. Comparing the largest estimated total number of events (IDMC) with those that triggered an international humanitarian response (as recorded by IFRC and the UN Office for the Coordination of Humanitarian Affairs), reveals that in the vast majority of cases, national governments and local communities responded without requesting, directly or indirectly, international assistance.

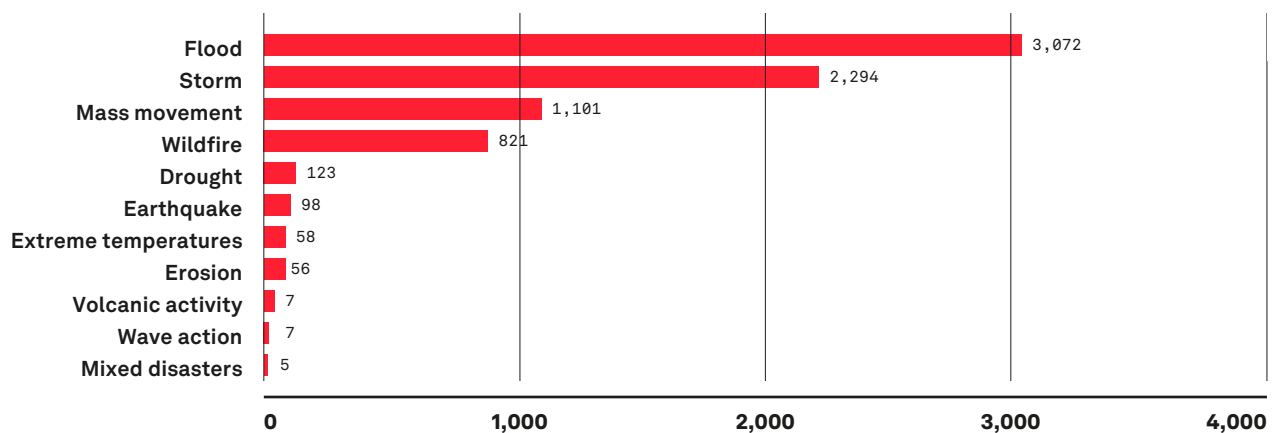
Fig II.3 **Number of requests by National Societies for international assistance, 2020-2024**



Source: IFRC GO (DREF and Emergency Appeals)

Figure A2.3 shows the number of requests received by the IFRC from National Societies for international assistance, the majority of which relate to floods. The fact that most disasters are managed predominantly nationally and locally is both appropriate and a sign of resilience. This is because the majority of disasters are caused by frequently occurring, low-intensity hazards that affect a relatively small number of people and communities. For example, of the 14,348 displacement-causing disasters recorded by IDMC, more than half (7,642 events, or 53%) displaced 100 or fewer people – predominantly due to localized floods, storms and landslides (Figure A2.4).

Fig II.4 **Number of events that displaced 100 or fewer people, by hazard type, 2020–2024**

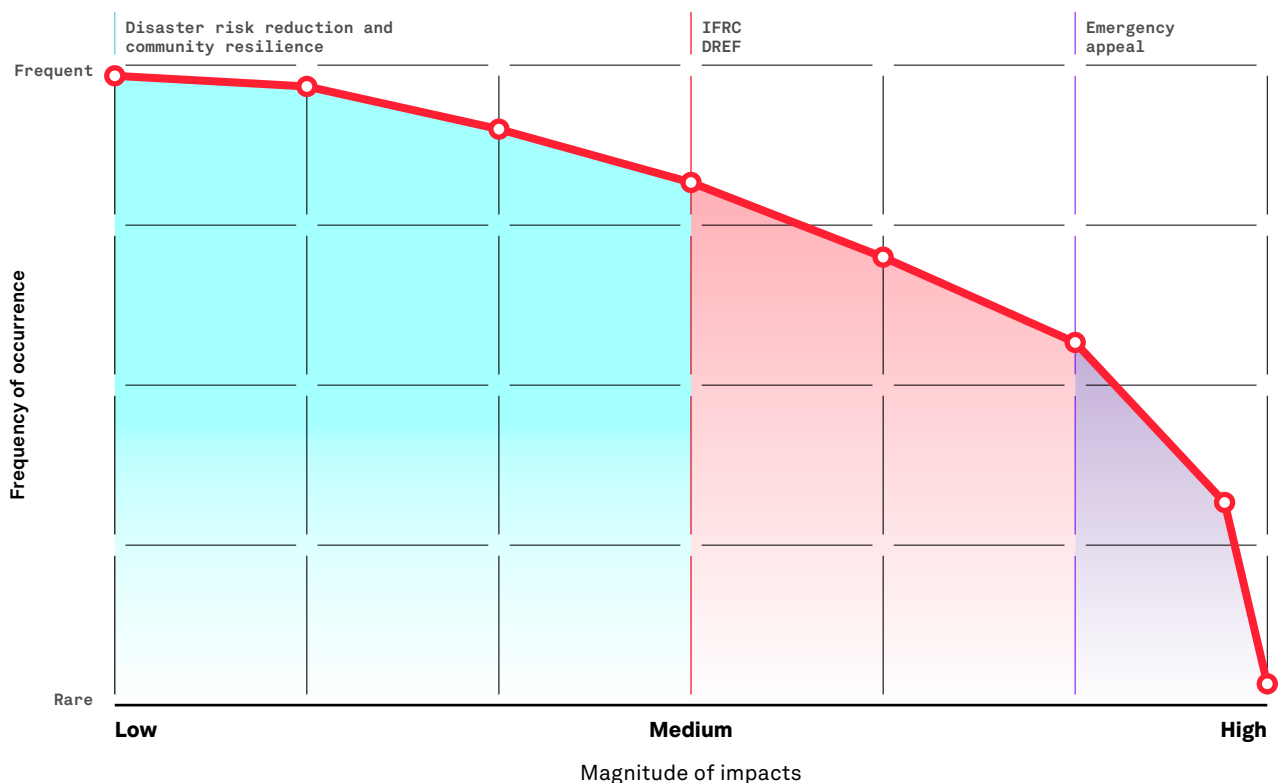


Source: Montandon/IDMC-GIDD

Disasters have always occurred but we are becoming better able to detect and understand this trend due to greater investment and improvements in data collection, particularly for smaller-scale events.

For decades, the IFRC has allocated its funds and mobilized international humanitarian assistance in line with the scale of disasters: strengthening National Society capacities in disaster risk reduction, preparedness and response to manage ‘smaller’ disasters, using the Disaster Response Emergency Fund (DREF) for medium and large disasters, and launching emergency appeals for the more infrequent catastrophic events (Figure A2.5).

Fig II.5 How IFRC and other humanitarian actors manage risk across the sector

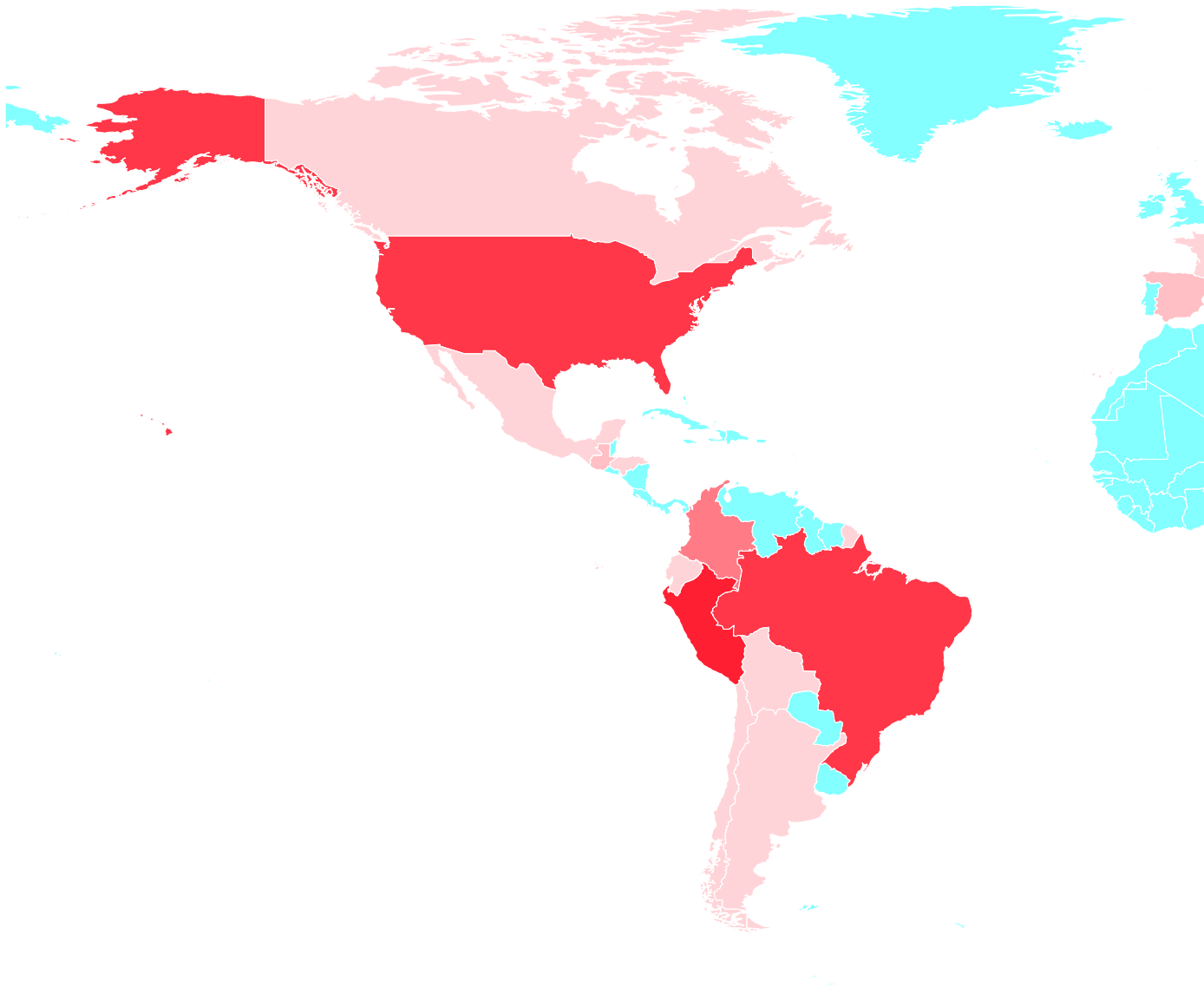


A2.2.1

Where are disasters happening?

The geographical distribution of recorded disasters reveals some striking patterns: 84% of events (disasters) occurred in just 37 countries (Figure A2.6). At first glance, these countries appear diverse, spanning high-, upper-middle-, lower-middle- and low-income countries. What they share, however, is both hazard exposure and the capacity to systematically record the impacts of small, medium and large disasters, often in settings with large populations. A comparison of recorded impacts in Africa with disaster risk (Figures A2.6 and A2.7) indicates that there is likely under-reporting across many countries.

Fig II.6 Countries where most disasters are recorded: 2020–2024



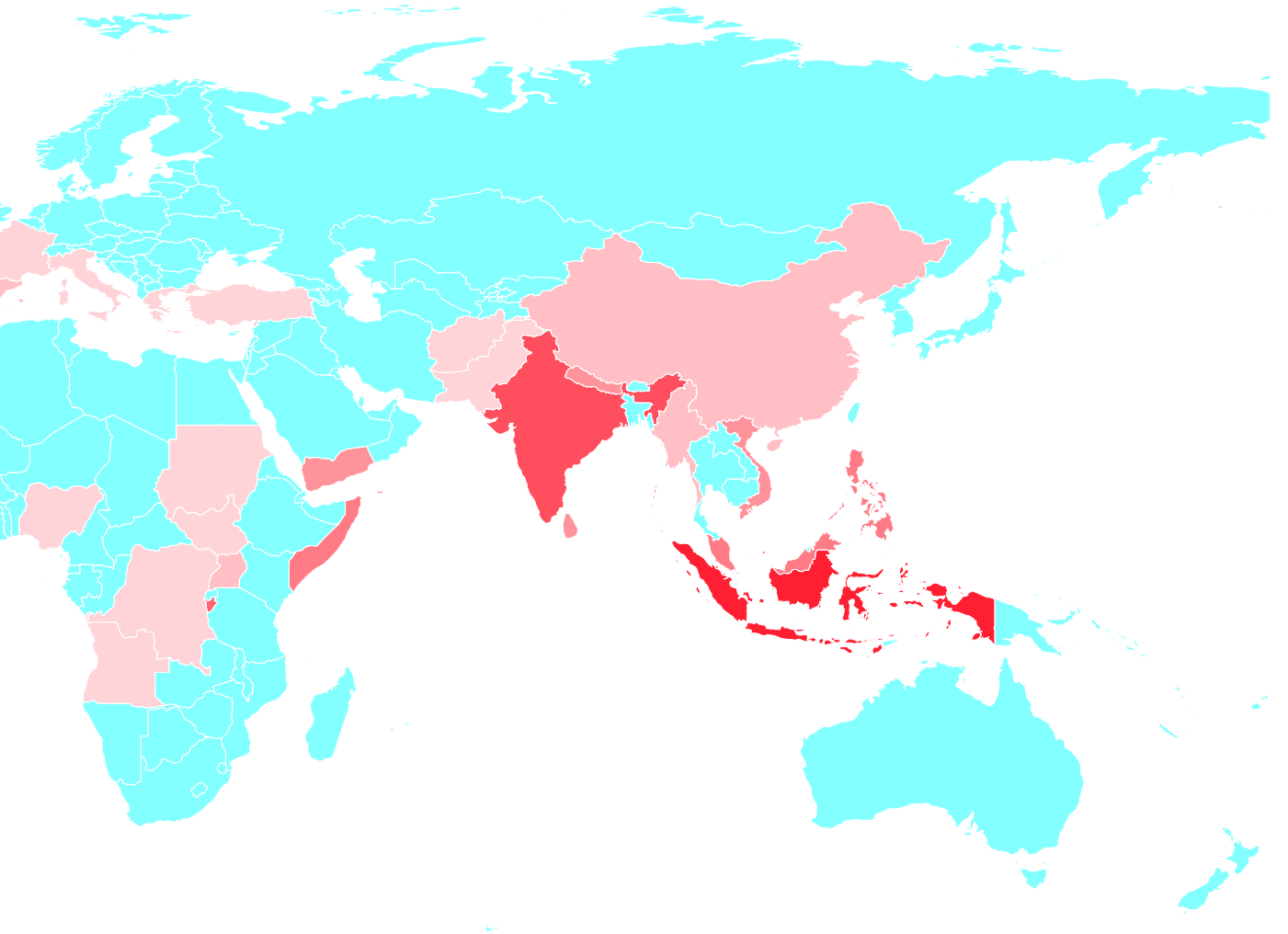
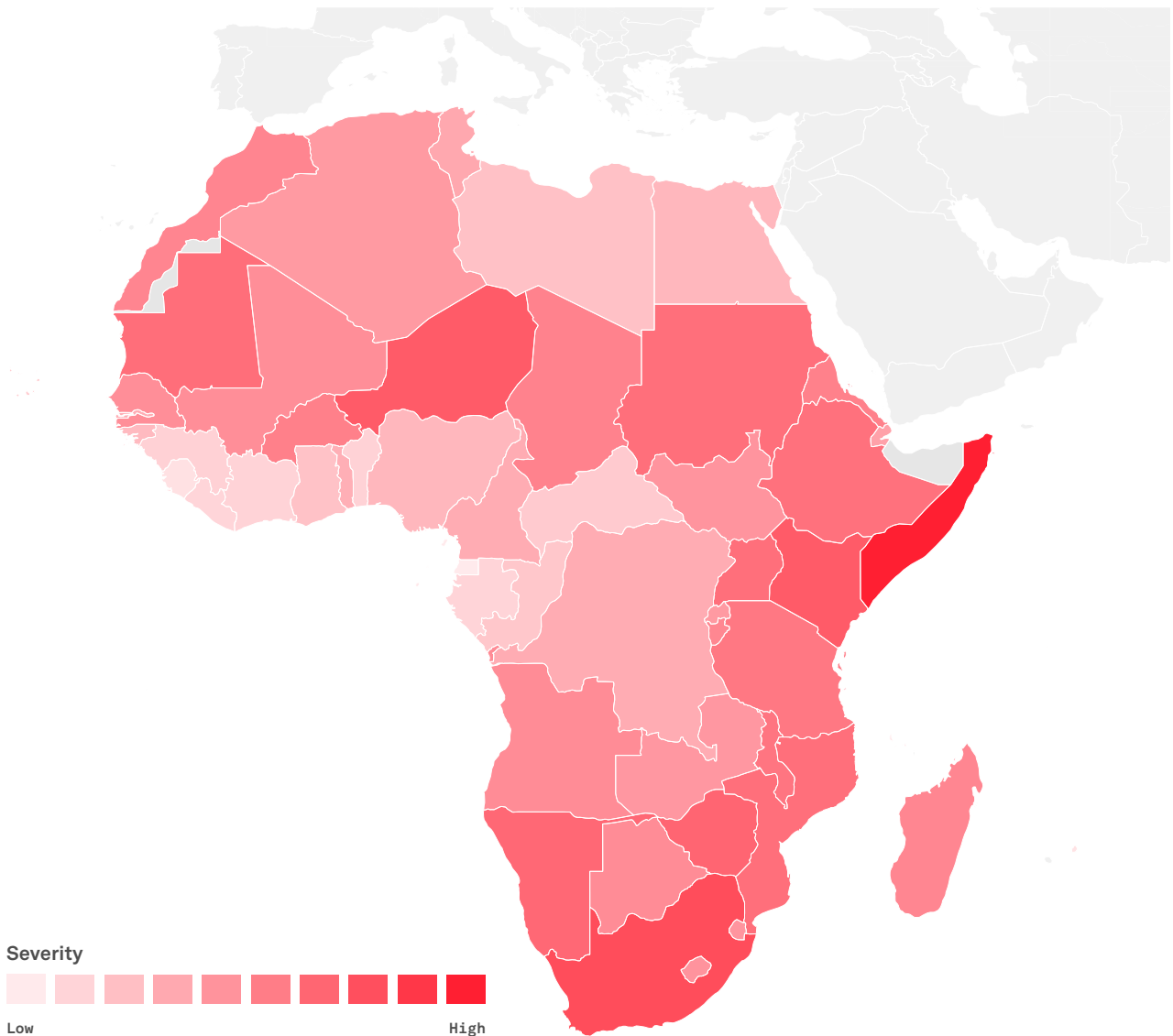


Fig II.7 Countries in Africa by risk of flooding and cyclones



Source: IFRC GO and INFORM Risk Index

A2.2.2

How severe are they?

Between 2020 and 2024, disasters affected some 697 million people,³ caused 105 million displacements and killed 271,000 people.⁴ The severity, as indicated by the number of people in need of humanitarian assistance, also increased during this time, from 192 million people in January 2020 to more than 433 million people at the end of 2024.⁵ According to the Intergovernmental Panel on Climate Change (IPCC), “Human-induced climate change, including more frequent and intense extreme events, has caused widespread adverse impacts and related losses and damages to nature and people, beyond natural climate variability.”⁶ Given that hydrometeorological hazards drive most of this

risk, as shown in Figures A2.3 and A2.4 above, it is expected that the number of people killed, displaced or otherwise affected by disasters will grow over the coming years.

A2.2.3

Where do we still struggle to answer these questions due to a lack of data?

The data paints a fairly comprehensive picture in some parts of the world (North and South America, Europe, Asia), but not others (Africa), and the impacts of certain hazards are significantly under-reported, particularly metrics related to extreme heat and diseases. Given that both of those types of hazards are influenced by weather and climate, it means our understanding of these risks and how to manage them need more resources and international attention.

Our ability to collect data on disasters has improved in recent years, but recent defunding of many, often small and independent, data collection organizations threatens that ecosystem of evidence building. The evidence base on which principled and effective humanitarian interventions are prioritized and implemented is, as a result, much poorer.

Concluding remarks: Lessons and looking ahead

- How many events are occurring? Many of the key sources show an increasing trend. For example, IDMC figures show 14,348 events between 2020–2024 compared to 5,664 during the previous five years (2015–2019). However, this is partly due to recent investment in methods to record and validate events, particularly smaller-scale disasters.
- Where are disasters happening? The geographical distribution of recorded disasters is quite concentrated, with 84% of events occurring in only 37 countries. A number of often-overlapping factors explain this: these countries are prone to disasters due to the high number of people and assets exposed to natural hazards; they are capable of recording the impacts of hazards when disasters do occur. It also suggests that there is significant under-reporting across most of sub-Saharan Africa, where millions of people are exposed to flooding, drought and other hazards, and relatively few events recorded.
- During this period, disasters affected some 697 million people,⁷ caused 105 million displacements and killed 271,000 people.⁸ With hydrometeorological hazards driving most of this risk, these numbers are expected to rise in the coming years.
- How do communities respond, or take anticipatory action, with or without international actors? Around 94% of disasters are managed by national governments and communities without international assistance, and this is inherently due to their resilience.

- Which of these questions can be answered confidently and with what data? It is a bitter irony that the most disaster-prone countries often lack the institutional capacity to systematically capture data on their impacts. As a result, the many disasters which are handled with local response actors and through community resilience go under-reported. The growing risks to the funding of humanitarian data and information management organizations and services threaten to exacerbate this gap.



Asks, aims and recommendations

Asks

Sustain investment in disaster data systems at local, national and international levels, with particular support to National Societies and local actors to strengthen their capacities for data collection, analysis and sharing.

Aims

Enable more evidence-based humanitarian action and decision-making through improved data collection, analysis and use.

Recommendations

States and policy-makers

- Invest in comprehensive and reliable disaster data collection systems.
- Commit to greater transparency and the use of evidence-based funding allocation processes.

Humanitarian actors

- Commit to generating and sharing the strongest possible evidence collectively, rather than advancing isolated evidence.
- Work transparently to acknowledge data gaps, limitations and biases to ensure data and evidence are used appropriately.

Communities and local leaders

- Insist on being involved in data collection and analysis, and assert your essential role in validating data and evidence generated by others.
- Collaborate with humanitarian actors and donors to ensure community insights inform ongoing responses and interventions.

Endnotes

- 1 The data analysed for the following insights comes from the Montandon, the world's largest open-source, open-access disaster database, which integrates data on natural hazards and their impacts from several sources.
- 2 We use the term 'displacements' rather than 'displaced people' because a fraction of the 160 million displacements reflects individuals who were displaced multiple times.
- 3 Given the way data on affected populations is collected, this figure 'double counts' anyone who was affected by more than one disaster during this period.
- 4 From Montandon, EM-DAT and IDMC-GIDD.
- 5 European Commission. INFORM Severity Index. <https://drmkc.jrc.ec.europa.eu/inform-index/INFORM-Severity/Severity-Facts-Figures>
- 6 IPCC. Summary for Policymakers. In: *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press, 2022) www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_SummaryForPolicymakers.pdf
- 7 Given the way data on affected populations is collected, this figure 'double counts' anyone who was affected by more than one disaster during this period.
- 8 From Montandon, EM-DAT and IDMC-GIDD.



The fundamental principles

Humanity

The Movement, born of a desire to bring assistance without discrimination to the wounded on the battlefield, endeavours – in its international and national capacities – to prevent and alleviate human suffering wherever it may be found. Our purpose is to protect life and health, and to ensure respect for the human being. We promote mutual understanding, friendship, cooperation and lasting peace among all peoples.

Impartiality

The Movement makes no discrimination as to nationality, race, religious beliefs, class or political opinions. We endeavour to relieve the suffering of individuals, being guided solely by their needs, and to give priority to the most urgent cases of distress.

Neutrality

In order to continue to enjoy the confidence of all, the Movement may not take sides in hostilities or engage at any time in controversies of a political, racial, religious or ideological nature.

Independence

The Movement is independent. National Societies, while auxiliaries in the humanitarian services of their governments and subject to the laws of their respective countries, must always maintain their autonomy so that they may be able at all times to act in accordance with the principles of the Movement.

Voluntary service

The Movement is a voluntary relief movement not prompted in any manner by desire for gain.

Unity

There can be only one Red Cross or one Red Crescent Society in any one country. It must be open to all. It must carry on its humanitarian work throughout its territory.

Universality

The Movement, in which all National Societies have equal status and share equal responsibilities and duties in helping each other, is worldwide.