

Policy Brief

Misinformation, Disinformation and the Crisis of Trust in Public Institutions

A Multilevel Agenda for Information
Integrity in an Era of Fragmented
Data Governance

Simon Chesterman



UNU
CPR



GGI | Global
Governance
Innovation

About the Author

Simon Chesterman is David Marshall Professor of Law and Vice Provost (Educational Innovation) at the National University of Singapore, where he is also the founding Dean of NUS College. He is the editor, with Araz Taeihagh and Audrey Yue, of the forthcoming *Oxford Handbook of Misinformation and Disinformation*.

Acknowledgements

This brief was prepared for the United Nations University Centre for Policy Research (UNU-CPR) and draws on a presentation delivered at UNU-CPR in April 2026, including a discussion led by UNU-CPR's David Passarelli. Many thanks to Fakhar Abbas, Jack Durrell, David Passarelli, Eka Nugraha Putra, Araz Taeihagh and Jun Yu for comments on earlier drafts.

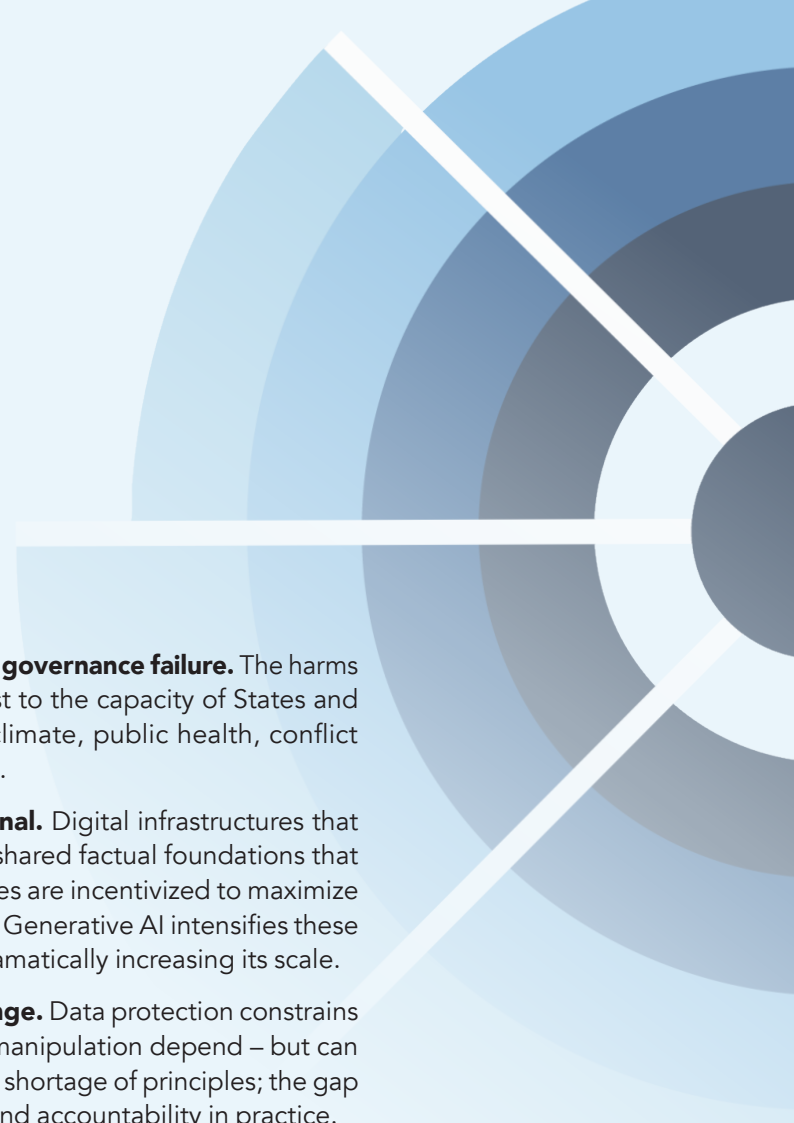
Disclaimer

This brief is an output of the Global Governance Innovation Platform initiative. The views and opinions expressed do not necessarily reflect the official policy or position of the United Nations University.

ISBN: 978-92-808-6675-9 © United Nations University 2026.

All content (text, visualizations, graphics), except where otherwise specified or attributed, is published under a Creative Commons Attribution-NonCommercial-ShareAlike IGO license (CC BY-NC-SA 3.0 IGO). Using, reposting and citing this content is allowed without prior permission.

Citation: Simon Chesterman, "Misinformation, Disinformation and the Crisis of Trust in Public Institutions: A Multilevel Agenda for Information Integrity in an Era of Fragmented Data Governance", GGI Platform Policy Brief (New York, United Nations University, 2026).



Key Insights and Recommendations

Main argument and findings:

- **Misinformation and disinformation are a multilevel governance failure.** The harms extend beyond electoral politics or institutional trust to the capacity of States and international organizations to act collectively on climate, public health, conflict prevention and the Sustainable Development Goals.
- **The problem is structural, not merely informational.** Digital infrastructures that in principle democratize expression also erode the shared factual foundations that make collective action possible; private sector entities are incentivized to maximize engagement rather than promote reasoned debate. Generative AI intensifies these dynamics by lowering the cost of fabrication and dramatically increasing its scale.
- **Data governance is one layer of this wider challenge.** Data protection constrains the personal data pipelines on which profiling and manipulation depend – but can also make manipulation harder to detect. There is no shortage of principles; the gap is mechanisms that reconcile privacy, transparency and accountability in practice.
- **The United Nations (UN) has built the normative case but not the operational architecture.** The adoption of the *2024 Global Principles*, the *Global Digital Compact* and the work of the Department of Global Communications are genuine achievements. What remains underdeveloped is how those principles travel across the different layers at which the problem operates.
- **The institutional landscape is becoming more crowded.** The launch of the World Data Organization in March 2026 signals that data governance norms are being shaped in multiple venues outside the UN system. Unless information integrity is built into those processes now, the two agendas risk permanent divergence.
- **The costs of delay are cumulative.** When institutions cannot see enough to respond intelligently, blunt responses follow: vague “fake news” laws, emergency takedowns, broad liability rules or shutdowns. Platform opacity, unequal research capacity and AI-enabled amplification, meanwhile, increase the risk of epistemic fragmentation (the breakdown of a shared basis for knowing and judging what is true) and weaken the basis for multilateral cooperation itself.

Correcting common misperceptions:

- **This is more than a content-moderation problem.** The deeper issue is the governance of data, incentives and infrastructure.
- **A broad UN regime for policing “truth” is neither desirable nor realistic.** The UN’s advantage is architectural: building standards, convening coalitions and embedding information integrity into live policy domains.
- **National regulation alone will not suffice.** Platforms, data flows and influence operations are transnational. The response must be layered: local and national where possible, regional and global where necessary.

Recommendations:

- **Develop a UN model protocol for privacy-preserving researcher access to platform data** (United Nations Educational, Scientific and Cultural Organization (UNESCO), with the UN's Department of Global Communications (DGC) and the Office of the United Nations High Commissioner for Human Rights (OHCHR)). Define categories of systemic data – ad libraries, recommender outputs, AI provenance metadata, coordinated inauthentic behaviour indicators – that vetted researchers may access, with clear safeguards and independent adjudication of refusals.

Benchmark: Three pilot arrangements in elections, health and climate within five years.

- **Establish a light coordination mechanism on data governance and information integrity** (Executive Office of the Secretary-General umbrella, DGC as operational lead). A small cross-agency steering group to align guidance, support UN Country Teams, maintain a multilingual resource hub and transfer lessons across climate, health, elections and conflict. Stress the need for State regulation rather than relying on the goodwill of private actors.

Benchmark: A common UN guidance note and four sector-specific action plans within five years.

- **Issue OHCHR guidance on the privacy–integrity balance in national regulation** (with UNESCO, DGC and the Special Rapporteurs on freedom of expression and privacy). Clarify how transparency obligations, researcher safeguards and limits on emergency powers and shutdowns can be designed consistently with international human rights law.

Benchmark: Citation in at least two intergovernmental processes and inclusion in at least one regional instrument within five years.

Scope and ambition:

- **Scale the coalition model across policy domains.** Climate shows that a focused, cross-agency coalition tied to a live intergovernmental process produces more durable outcomes than abstract reform agendas. The model should be extended to public health, electoral integrity and conflict prevention.
- **The UN cannot solve misinformation and disinformation.** But it can offer a framework in which to build a practical settlement that enables privacy-preserving scrutiny of high-risk systems, reduces the temptation toward blunt State responses and supports sector-specific efforts to strengthen resilience.

Introduction

Misinformation and disinformation are often treated as pathologies of public debate: distortions of electoral politics, symptoms of social polarization or threats to trust in public institutions. They are all of those things. But they are also something more consequential. In an international system that depends on shared facts, minimum trust and the possibility of reasoned disagreement, the erosion of information integrity is increasingly a governance problem. It affects not merely what citizens believe, but whether societies, States and international organizations can mobilize collective action on climate, public health, conflict prevention, humanitarian response and the Sustainable Development Goals.¹

The problem is structural as much as informational. A handful of private platforms hold extraordinary leverage over the design of information environments, while their business models often reward engagement rather than integrity. Generative AI intensifies these dynamics by lowering the cost of fabrication, increasing the scale of synthetic content and complicating verification.² The result is even more misleading material, but also a broader weakening of the conditions under which judgment can be exercised at all.³ A purely State-centred response is inadequate: many of the most important levers lie in the hands of private actors whose operations are transnational and whose incentives are structurally misaligned with information integrity. A multilevel response is therefore not optional. It is the only plausible response.

Data governance enters this picture not as a separate debate, but as one layer of the wider challenge. Data protection can help constrain manipulation by limiting

the personal data pipelines on which profiling and microtargeting depend. At the same time, privacy protections can make manipulation more difficult to detect when researchers and regulators cannot see enough of the underlying system. That does not mean that one must choose between privacy and information integrity. On the contrary, the challenge is to build trusted, rights-based arrangements that allow scrutiny without normalizing surveillance. That question has become more urgent as the data governance landscape grows more crowded: the launch of the World Data Organization (WDO) in March 2026 underscores that rules on data access, governance and digital infrastructure may increasingly be shaped in multiple venues, not all of them closely connected to the human rights-based information integrity agenda of the United Nations (UN).⁴

The UN has already assembled much of the normative case. The Secretary-General's 2023 policy brief,⁵ the 2024 *Global Principles for Information Integrity*,⁶ the *Global Digital Compact*⁷ and the practical work led by the Department of Global Communications (DGC) together represent a serious effort to frame the problem in human rights terms and to move from principle to practice. The question is no longer whether information integrity belongs on the multilateral agenda. It plainly does. The question is what kind of agenda is now needed: not a sweeping new regime for policing truth, but a more modest and more consequential architecture for digital resilience that connects data governance, platform accountability, human rights safeguards and sector-specific practice across the multiple layers at which the problem now operates.

¹ The first United Nations Global Risk Report identified mis- and disinformation as a top global vulnerability – a significant institutional acknowledgment that the international community remains insufficiently prepared. See *2024 Global Risk Report* (United Nations publication, 2025).

² Shaleen Khanal, Hongzhou Zhang and Araz Taeihagh, "Why and How Is the Power of Big Tech Increasing in the Policy Process? The Case of Generative AI", *Policy and Society*, vol. 44, No. 1 (2025), p. 52; A/HRC/61/68.

³ Simon Chesterman, "Lawful but Awful: Evolving Legislative Responses to Address Online Misinformation, Disinformation, and Mal-Information in the Age of Generative AI", *American Journal of Comparative Law*, vol. 72, No. 4 (2024).

⁴ Yichao Du, "From Initiative to Institution: China, the World Data Organization, and the Politics of Global Data Governance", *SSRN*, 8 April 2026. Available at <http://dx.doi.org/10.2139/ssrn.65401>.

⁵ *Information Integrity on Digital Platforms, Our Common Agenda Policy Brief 8* (United Nations publication, 2023). Available at <https://www.un.org/en/civil-society/information-integrity-digital-platforms>. See also A/75/982, para. 36.

⁶ United Nations, "Information Integrity", n.d. Available at <https://www.un.org/en/information-integrity/global-principles>.

⁷ A/79/L.2.

Disinformation as a Challenge Across the UN System

- 1. Peace and security:** Information manipulation is now a standard instrument of armed conflict and political destabilization. The UN Security Council⁸ has debated disinformation as a threat to peacekeeping mandates.
- 2. Climate change:** The UN *Global Risk Report* put mis- and disinformation just behind climate change and large-scale pollution as the top global concerns,⁹ though in fact these issues are connected.
- 3. Public health:** The COVID-19 “infodemic”¹⁰ demonstrated the life-or-death stakes. The Pandemic Treaty recognizes the importance of timely information-sharing to prevent misinformation and disinformation, and requires parties to strengthen pandemic literacy and access to transparent, timely, accurate, science- and evidence-based information – suggesting one way such commitments may be embedded in an international legal instrument.¹¹
- 4. Election integrity:** The *Global Principles*¹² include specific recommendations on protecting elections from disinformation and on the responsibilities of political actors.
- 5. Press freedom and journalist safety:** The work of the United Nations Educational, Scientific and Cultural Organization (UNESCO) on the safety of journalists,¹³ including environmental journalists and women journalists, addresses the systematic targeting of information defenders.
- 6. Education:** UNICEF has flagged the importance of media literacy for children to ameliorate the consequences of disinformation.¹⁴
- 7. AI governance:** The UN’s *Governing AI for Humanity* report¹⁵ (2024) directly links AI development to information integrity risks, including at scale; the *Global Principles* include a dedicated section on AI actors.

⁸ United Nations, “Secretary-General’s remarks to the Security Council Open Debate on Artificial Intelligence and International Peace and Security”, 24 September 2025.

⁹ 2024 *Global Risk Report*, p. 5.

¹⁰ World Health Organization, “Infodemic”, n.d.

¹¹ WHO Pandemic Agreement, 20 May 2025, WHA78.1, Annex, art. 16. This is, however, less than was initially proposed. The 2022 Conceptual Zero Draft contained explicit operative language on tackling misinformation and disinformation, managing infodemics and conducting social media analysis.

¹² *United Nations Global Principles for Information Integrity: Recommendations for Multi-Stakeholder Action* (United Nations publication, 2024).

¹³ UNESCO, “Safety of Journalists”, n.d.

¹⁴ Philip N. Howard, Lisa-Maria Neudert, Nayana Prakash and Steven Vosloo, *Digital misinformation/disinformation and children* (UNICEF, 2021).

¹⁵ *Governing AI for Humanity* (United Nations publication, 2024).

Section 1 ●●●●

What the UN Has Built – and Why It Is Not Enough

Over the past three years the UN has moved misinformation and disinformation from the margins of digital-policy debate toward the centre of discussions about governance, rights and resilience. The progress is real. The 2024 *Global Principles for Information Integrity* established a five-part framework – societal trust and resilience, healthy incentives, public empowerment, independent and pluralistic media, and transparency and research – grounded in human rights law and addressed to every major stakeholder group.¹⁶ The *Global Digital Compact* deepened the normative shift, committing Member States to supporting independent journalism, media literacy curricula and data access for researchers.¹⁷ And the Global Initiative for Information Integrity on Climate Change¹⁸ demonstrated that information integrity can be integrated into a live policy process, producing the Belém Declaration at COP30 and embedding the issue in international climate governance for the first time.¹⁹

Yet the framework remains incomplete in three important respects:

1. Principles without mechanisms. The *Global Principles* call for privacy-preserving data access, greater transparency and stronger support for researchers. But they do not resolve the practical questions on which implementation depends: who may access

which platform data, under what safeguards, for what public purpose, and with what oversight. This is the central governance gap.

2. Norms do not by themselves alter incentive structures. Large technology companies control the architecture of attention and shape the conditions under which information is produced, distributed and monetized. Their business models often reward precisely the content most likely to inflame or manipulate.²⁰ The UN can set standards and convene coalitions. It cannot, on its own, compel platforms to redesign recommender systems, open data to scrutiny or abandon profitable forms of opacity. Any agenda that ignores this structural reality will overpromise and underdeliver.

3. The UN system itself remains fragmented. DGC has developed an information integrity function and a practical 3R model (Research, Risk Assessment, Response). UNESCO has advanced platform governance, AI ethics, and journalist safety. The Office of the United Nations High Commissioner for Human Rights (OHCHR) provides essential human rights guardrails. The United Nations Development Programme (UNDP) and operational agencies encounter the downstream effects in democratic governance and peacebuilding. But those pieces

¹⁶ United Nations *Global Principles for Information Integrity*.

¹⁷ *Global Digital Compact* (United Nations, September 2024). The Department of Global Communications' 2025 issue brief, *From Principles to Practice*, pushed this further by setting out a theory of change and an operational model centred on research, risk assessment and response. *Issue Brief 1: From Principles to Practice: Strengthening Information Integrity* (United Nations publication, 2025). These commitments and exhortations exist in a complementary normative context, including: the *UNESCO Guidelines for the Governance of Digital Platforms* (2023), the *UNESCO Recommendation on the Ethics of AI* (2021), the UN *Strategy and Plan of Action on Hate Speech* (2019) and the UN *Guiding Principles on Business and Human Rights* (2011).

¹⁸ UNESCO, "Global Initiative for Information Integrity on Climate Change", n.d.

¹⁹ United Nations Climate Change, "Countries Seal Landmark Declaration at COP30 – Marking First Time Information Integrity Is Prioritized at UN Climate Conferences", 12 November 2025.

²⁰ Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Public Affairs, 2019); Simon Chesterman, "Silicon Sovereigns: Artificial Intelligence, International Law, and the Tech-Industrial Complex", *American Journal of International Law*, vol. 120, No. 1 (2026).

have not been joined into a coherent operational architecture.²¹ Data governance and information integrity, meanwhile, continue to be developed

through adjacent but insufficiently connected processes – while the wider institutional landscape, exemplified by the WDO, grows more crowded.²²

Section 2 ●●●●

The Governance Gap

The governance gap is more than a lack of means to deal with false content. Misinformation, disinformation and the wider erosion of information integrity are being addressed through disconnected institutions, norms and policy tools, even though the problem now operates across several interlocking layers. At one level lie the data practices and platform architectures that make manipulation possible. At another lie the institutional arrangements through which States, international organizations, platforms, researchers and civil society attempt to respond. At a third lie the concrete policy domains – climate, health, elections, conflict – in which failures of information integrity translate into failures of collective action. Each layer is difficult in itself, but treating them separately misses possible synergies and amplifies the risk of confounding responses.

The data governance dimension matters because the effectiveness of modern information manipulation depends heavily on data practices. The same systems that allow platforms and advertisers to profile users, target messages and amplify emotionally resonant material also enable coordinated manipulation to spread quickly and

at scale.²³ Yet those systems are opaque to independent scrutiny. Data protection law performs a dual function: properly designed, it constrains the personal data pipelines on which profiling and microtargeting depend; but it can also make manipulation harder to detect when researchers and public institutions cannot gain trusted access to systemic data. The paradox sits at the heart of the governance problem.²⁴

When governments cannot see enough to act intelligently, they tend to act crudely: through vague “fake news” laws, emergency takedowns, broad intermediary liability rules or outright shutdowns. The human rights critique of such measures is well established.²⁵ The institutional point is equally important: in the absence of credible mechanisms for privacy-preserving scrutiny, pressure builds for action directed at downstream symptoms rather than upstream conditions. A rights-consistent response requires more than restraint. It requires better ways of seeing.

The wider data governance landscape sharpens the urgency. New institutions – including the WDO – are emerging to shape standards on data access, governance

²¹ The Global Principles themselves call for a central unit in the Secretariat, a multilingual resource hub, and support for multi-stakeholder action plans – though it is unclear whether there is appetite or budget for any significant new institutional response. *United Nations Global Principles for Information Integrity*, p. 38.

²² A/79/L.2, paras. 37–49.

²³ Lynnette H.X. Ng and Araz Taeihagh, “How Does Fake News Spread? Understanding Pathways of Disinformation Spread Through APIs”, *Policy and Internet*, vol. 13, No. 4 (2021).

²⁴ J. W. F. Small, “Data Protection Regulations as Disinformation Resilience”, in *Digital Resilience*, Matthew Stubbs, Samuel White and Dale Stephens, eds. (Springer, 2025), p. 51; Alexander Romanishyn, Olena Malyska and Vitaliy Goncharuk, “AI-Driven Disinformation: Policy Recommendations for Democratic Resilience”, *Frontiers in Artificial Intelligence*, vol. 8 (2025).

²⁵ Gregory T. Nojeim, Evelyn Mary Aswad and Asaf Lubin, “Internet Shutdowns and International Law”, *Proceedings of the ASIL Annual Meeting*, vol. 119 (2025), p. 247.

and digital infrastructure outside the UN system. If upstream rules on data access and governance are shaped without adequate regard for the conditions of trustworthy information, the multilateral response to manipulation will remain structurally incomplete. The task for the UN is to ensure that the data governance and information integrity agendas are joined up before institutional fragmentation hardens into policy incoherence.

Seen in this light, asking whether the UN can solve misinformation and disinformation is the wrong question. It plainly cannot. The more useful question is whether it can help build a practical settlement across these layers: one that enables privacy-preserving scrutiny of high-risk systems, reduces the temptation toward blunt State responses and supports sector-specific efforts to strengthen resilience. That is a narrower ambition than grand talk of “governing truth.” But it is the more plausible one.

Section 3 ●●●●

Three Priorities for the Next Five Years

If the governance gap is layered, the response must be layered as well. The three priorities below are calibrated to the structural limits already identified: the UN cannot compel platforms to redesign their systems, cannot eliminate geopolitical incentives for manipulation, and is unlikely to secure support for a large new bureaucracy in the current political and budgetary environment. What it can do is help build a more credible architecture for digital resilience.

Priority 1: A model protocol for privacy-preserving researcher access

The first priority sits at the infrastructure layer. The *Global Principles* call for privacy-preserving data access for researchers;²⁶ the *Global Digital Compact* urges platforms to provide researchers with relevant data while protecting user privacy.²⁷ What remains missing is a practical mechanism specifying who may access which data, under what safeguards, for what purpose and with what oversight.

The UN should sponsor a **model protocol for privacy-preserving researcher access to platform data**: a voluntary but authoritative template that States, regulators, platforms and civil society could adopt or reference in national and regional frameworks. The protocol should define categories of systemic data that vetted researchers may request – ad libraries and targeting criteria, recommender-system outputs, trust-and-safety resourcing by language and geography, AI provenance metadata and indicators of coordinated inauthentic behaviour – with vetting criteria, safeguards and independent review where access is refused. The aim is controlled visibility into high-risk systems, not a general erosion of privacy.

UNESCO, with DGC and OHCHR, should lead. The principal obstacle will be resistance from platforms invoking privacy, security and trade secrecy, as well as unease from some States that fear scrutiny of their own information practices.²⁸ That is precisely why the initial ambition should be modest and demonstrable:

²⁶ *United Nations Global Principles for Information Integrity*, pp. 16–17, 32–33.

²⁷ A/79/L.2, para. 36(b).

²⁸ See, for example, Raphael Satter and Alexandra Alper, “US Orders Diplomats to Fight Data Sovereignty Initiatives”, *Reuters*, 25 February 2026; Edith M. Lederer, “UN Approves 40-Member Scientific Panel on the Impact of Artificial Intelligence over US Objections”, *Associated Press*, 13 February 2026.

the European Union's Digital Services Act Article 40 architecture – vetted researcher access supervised through independent coordinators – shows what is possible, even if its implementation has faced persistent platform resistance.²⁹

Benchmark: Three pilot arrangements – elections, public health and climate – within five years, supported by published methodology and periodic synthesis reporting.

Priority 2: A light coordination mechanism

The second priority sits at the institutional layer. Fragmentation remains a central obstacle: the DGC, UNESCO, OHCHR, UNDP and sectoral bodies all hold relevant parts of the problem, but they do not yet amount to a coherent operational architecture. The answer is not a large new office – that would be politically difficult, fiscally implausible and likely to provoke suspicion about UN oversight of speech. The climate initiative shows that a focused, cross-agency coalition tied to a live policy process can produce tangible results without large institutional overhead.

A light coordination mechanism on data governance and information integrity, operating under the umbrella of the Executive Office of the Secretary-General with DGC as operational lead, should bring together a small cross-agency steering group (UNESCO, OHCHR, UNDP and one rotating sectoral partner). Its mandate should be tightly defined: aligning guidance across the system; supporting UN Country Teams with shared tools and risk frameworks; maintaining a multilingual information integrity resource hub; and ensuring that lessons learned in one domain are transferred to others. Crucially, its reach should extend outward – to emerging data governance processes such as the WDO – not only inward across UN agencies.

Benchmark: A common UN guidance note on the data governance and information integrity nexus, a functioning multilingual resource hub and four sector-specific action plans within five years.

Priority 3: OHCHR guidance on the privacy–integrity balance

The third priority sits at the normative and guardrail layer. Two familiar pathologies produce similar effects: governments that invoke “disinformation” to justify censorship, and governments that invoke “privacy” to shield systems from scrutiny while allowing manipulation to flourish. In both cases, the language of governance becomes a means of evading accountability.

OHCHR should issue focused guidance on the privacy–integrity balance in national regulation, in consultation with UNESCO, DGC and the Special Rapporteurs on freedom of expression and privacy. The guidance should address the design of transparency obligations on platforms, the permissible scope of restrictions on political advertising, disclosure requirements for State requests, safeguards for independent researchers and fact-checkers and the strict limits international law places on emergency powers, intermediary liability rules and Internet shutdowns. The human rights vocabulary of legality, necessity, proportionality and non-discrimination can narrow the space for opportunistic responses even where States disagree sharply on the underlying terminology of “disinformation.”

Benchmark: Citation in at least two intergovernmental UN processes, inclusion in at least one regional instrument or declaration, and demonstrable use by UN Country Teams in legislative or regulatory dialogue within five years.

²⁹ 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); Catalina Goanta and others, “The Great Data Standoff: Researchers vs. Platforms Under the Digital Services Act”, arXiv 2505.01122v3 [cs.CY] (2025).

Section 4 ●●●●

What Is at Stake

The stakes are larger than the regulation of false content. What is at issue is the erosion of the shared informational conditions under which collective action remains possible. The UN's normative framework has emphasized that this erosion disproportionately harms those in vulnerable or marginalized positions and countries with weaker institutional capacity. In many parts of the world, data voids and information voids are structural features of the environment. Legitimate institutions are often slower, less resourced and less adaptive than actors seeking to manipulate public understanding. Where reliable information is absent or difficult to access, deceptive narratives do not merely compete with the truth – they arrive first.

One of the more insidious effects of contemporary manipulation is that it goes beyond simply persuading individuals to believe false claims; it also distorts perceptions of what others believe. When public support for collective action is systematically underestimated – whether on climate policy, public health measures or democratic norms – political will can be paralysed even where the underlying support exists.³⁰ This feedback loop of manufactured dissent has profound implications for multilateralism. Generative AI intensifies every dimension: it lowers the cost of fabrication, accelerates production of synthetic content, and increases scale across languages and jurisdictions. The core risk is more than deception. It is saturation – and saturation, over time, is corrosive of trust in any source at all.

The legitimacy of the UN itself is therefore implicated. If the organization cannot help shape a credible response, it risks being seen as peripheral to one of the defining governance challenges of the digital era. Yet the inverse danger is equally real: if the UN is seen as a vehicle for censorship or politically selective enforcement, its authority in this area will be quickly undermined. That is why the human rights foundation of the UN's information integrity agenda is more than normative positioning. It is a condition of institutional credibility – and of the UN's capacity to act in this area at all.

The UN cannot compel platforms to redesign their business models, eliminate geopolitical incentives for manipulation or erase the deeper conditions in which distrust flourishes. But it does occupy a position few other actors can match: the capacity to connect data governance, platform accountability, human rights safeguards and sector-specific practice within a single multilateral frame. The alternative – leaving the information environment to market forces, fragmented national regulation and reactive crisis management – has already been tried. A more plausible ambition is therefore the one advanced here: a multilevel agenda for digital resilience, modest in its claims but serious in its architecture, and grounded in the recognition that the defence of information integrity is, increasingly, part of the defence of multilateralism itself.

³⁰ Kristina Lerman, Xiaoran Yan and Xin-Zeng Wu, "The 'Majority Illusion' in Social Networks", *PLOS One* e0147617, vol. 11, No. 2 (2016); Peter Andre and others, "Globally Representative Evidence on the Actual and Perceived Support for Climate Action", *Nature Climate Change*, vol. 14 (2024), p. 253.

About GGI

The Global Governance Innovation Platform helps UN Member States design new multilateral institutions and reform existing ones. Through interactive mapping and data visualization, we identify and showcase innovative governance models and mechanisms that can be adapted to address urgent global challenges and collective action problems.

About UNU-CPR

The United Nations University Centre for Policy Research is a think tank within the United Nations that carries out policy-focused research and capacity-building on issues of strategic interest and importance to the UN and its Member States. The Centre prioritizes urgent policy needs requiring innovative, practical solutions oriented toward immediate implementation and sustainability over the long term.