IDENTIFYING SOLUTIONS TO PROTECT INFORMATION INTEGRITY ON PRIVATE MESSAGING PLATFORMS

FINAL REPORT OF THE WORKSTREAM
OF THE PARTNERSHIP FOR
INFORMATION AND DEMOCRACY
CO-CHAIRED BY **LUXEMBOURG** AND **UKRAINE**

FID FORUM INFORMATION DEMOCRACY

Knowledge Partner
NYU STERN
Center for Business and Human Rights

March 2026

FOREWORD

“

By Luxembourg & Ukraine

As co-Chairs of the Workstream on «Identifying Solutions to Protect Information Integrity on Private Messaging Platforms,» we are proud to present this Final Report, which marks a significant milestone in the Partnership for Information and Democracy’s efforts.

The rapid evolution and widespread adoption of private messaging services presents a governance challenge that touches every one of our 57 partner states. These platforms, characterized by their scale, immediacy, and use of encryption, serve as indispensable tools for communication and civic coordination. Yet, they are also increasingly exploited as vectors for foreign manipulation, disinformation, and coordinated inauthentic behavior.

The regional aspect of this challenge cannot be overstated. For Ukraine, this discussion is not theoretical; it is a matter of national security and resilience. During a full-scale war, private messaging platforms become both a critical lifeline for civil defense and emergency alerts, and a primary battleground for Russian disinformation campaigns aimed at sowing panic and eroding trust. This urgent context underscores the absolute necessity of implementing the pragmatic, rights-respecting recommendations contained herein.

Even during peacetime, the abuse of private messaging platforms to spread disinformation, for instance to influence elections or undermine citizens’ trust in their democratic institutions, is an immediate concern to all of us. Similarly, civil society organizations and human rights defenders rely on the integrity of private messaging platforms, notably their encryption, for secure communications. Attempts to weaken the cryptographic standards of private messaging platforms are rejected by human rights organizations, security researchers, and privacy advocates alike, given that they risk weakening security for everyone.

This report underscores that no single solution will suffice. While regulatory frameworks must evolve to address the hybrid public-private nature of these platforms, the long-term protection of our information ecosystems hinges on media literacy and societal resilience. We must collectively invest in empowering citizens to critically navigate their information environment.

We urge all 57 countries of the Partnership for Information and Democracy to carefully study these findings. The continued international cooperation and shared commitment shown by all participating states are vital. We look forward to transforming these recommendations into concrete policy actions to safeguard democratic discourse globally and to strengthen our collective ability to respond to future threats.

”

EXECUTIVE SUMMARY

Private messaging platforms have become central to the global information ecosystem. Their reach, networked design, and (variable) use of encryption make them both vital channels for private communication and significant vectors for disinformation. This report is developed by the Forum on Information and Democracy (FID) with the support of the NYU Stern Center for Business and Human Rights and builds upon practices, research and insights shared by states of the Partnership and members of the Forum's civil society coalition during a series of meetings of the Workstream. It examines current governance approaches and provides recommendations for safeguarding information integrity while protecting privacy and freedom of expression. It builds upon previous work of the FID and in particular its recommendations on mixed private and public spaces on closed messaging services (How to End Infodemics, 2020).

KEY FINDINGS

- **Disinformation on messaging platforms poses distinct challenges.** Research shows that closed groups, broadcast lists, and hybrid "semi-public" messaging channels amplify misinformation while eluding oversight. Case studies from Brazil to Lebanon illustrate how manipulated content circulates virally through encrypted groups during electoral periods and political crises, often via coordinated influence campaigns that mimic grassroots communication.
- **Private messaging spaces occupy a regulatory gray zone.** Across jurisdictions, definitions of "private" and "public" communications remain ambiguous. Many regulatory and self-regulatory frameworks — such as the EU Digital Services Act (DSA) and Australia's voluntary Code of Practice on Disinformation and Misinformation — either explicitly exclude or only partially cover private messaging functions. The UK's Online Safety Act (OSA) is a notable exception, treating messaging as a user-to-user service subject to transparency and risk mitigation requirements.
- **Media literacy is the most consistent policy response.** Most jurisdictions emphasize education over enforcement. Ukraine's "Filter" initiative, Ireland's "Be Media Smart" campaign, and Australia's national media literacy investments exemplify this trend. These programs aim to foster user awareness and resilience but cannot substitute for structural measures addressing virality and amplification features.
- **The balance between encryption and accountability remains unresolved.** Authorities widely recognize end-to-end encryption (E2EE) as essential to the right to privacy and freedom of expression. However, tensions persist between encryption and content moderation. Many governments seem to be grappling with this dilemma, underscoring the need for proportionate and transparent measures.
- **The introduction of artificial intelligence and advertising further blurs the categorization of private messaging platforms and poses challenges to encryption.** The evolving nature of private messaging platforms modifies the "private" aspect of these platforms and demonstrates that regulation is not yet fit for purpose.
- **International coordination is limited but growing.** Several countries express a demand for cooperative frameworks to share best practices and harmonize standards around the regulation of private messaging platforms.

RECOMMENDATIONS IN BRIEF

To Governments:

1. Regulate platform features rather than platform categories. Legal obligations should be tied to specific functionalities (e.g., group forwarding, broadcast lists, AI and advertising integration) rather than broad platform categories, reflecting the hybrid nature of modern communication services.

2. Clarify definitions of "public" and "private." Distinguish obligations based on audience size, discoverability, encryption, access controls, and shareability to ensure proportionate oversight.

3. Protect encryption. Avoid mandates that render E2EE technically unworkable; ensure that any content-related obligations exclude end-to-end encrypted spaces or apply only to unencrypted functionalities.

4. Impose transparency and user empowerment. Including transparency about the implementation of platform's terms and services, notice-and-action procedures and information to users about new features and their implications.

5. Support media literacy. Develop national strategies to strengthen societal resilience to disinformation, measure their effectiveness, and share lessons with other governments.

6. Encourage international responses. Enlarge disinformation coordination mechanisms to private messaging platforms and develop international standards.

RECOMMENDATIONS IN BRIEF

To Platforms:

1. Separate messaging, broadcasting and AI functions. Differentiate encrypted private communication from large-scale or public-facing features; consider distinct products or user interfaces for each. Ensure that AI functions depend on opt-in consensus and are clearly distinguished from encrypted features.

2. Apply encryption judiciously. Limit E2EE to contexts where users have a reasonable expectation of privacy; introduce friction to mass forwarding to reduce virality.

3. Integrate human rights into the design by default. Respecting the United Nations Guiding Principles on Business and Human Rights, establishing clear terms of service and empowering users through opt-in features, notice-and-action procedures and review mechanisms.

4. Strengthen authentication and anti-abuse systems. Restrict multi-account creation and mitigate "phone farm" manipulation through device-level limits, rate controls and user empowerment.

5. Expand and promote verified "tiplines". This enables fact-checking and user empowerment.

6. Empower users through in-app verification tools. Support and monitor research into in-app affordances that enable users to fact-check information.

7. Engage in cross-platform cooperation. Participate in information-sharing initiatives to detect coordinated inauthentic behavior across messaging ecosystems.

8. Promote transparency and independent research. Proactively share information with independent civil society, researchers and regulatory authorities on features, functionalities and content moderation decisions while respecting privacy to enable independent evaluation.

CONTENT

1. INTRODUCTION

Navigating Information Warfare in the Age of Encrypted Messaging

Since Russia's full-scale invasion in 2022, Ukraine has faced an unprecedented onslaught of disinformation aimed at undermining public trust, sowing panic, and eroding confidence in national institutions. Telegram, now one of the country's most widely used platforms, has played a central role in this struggle. Originally valued for its accessibility, the app has become both a lifeline and a liability: a key channel for emergency communication and community resilience, but also a vector for false narratives and manipulation campaigns orchestrated from abroad.

Telegram's mixed architecture—combining public channels, semi-public group chats, and encrypted "secret chats"—creates an information ecosystem that straddles the line between social media and private messaging[1]. This hybrid structure has enabled covert influence operations that spread rapidly through "news" channels impersonating trusted outlets and amplifying fabricated stories about the war.

The Ukrainian government's response has been multifaceted. Recognizing that heavy-handed regulation could inadvertently stifle free expression or push users toward even less transparent channels, Ukraine has prioritized strategic communication, transparency, and media literacy over direct platform control. Ukraine's approach underscores both the urgency and complexity of governing information flows on private messaging platforms during conflict. It highlights the need to balance security, openness, and rights protection in an environment where encryption, anonymity, and virality intersect. The country's leadership in this Workstream reflects a genuine governmental interest in finding nuanced, context-sensitive solutions that other states may adapt to their own democratic and security challenges.

The Ukrainian case thus illustrates the dual nature of messaging platforms in the modern information environment: indispensable for civic coordination and crisis response, yet vulnerable to exploitation through design features that blur the boundary between private communication and public broadcasting. Understanding these features, and how they interact with regulation and user behavior, is essential for any credible effort to ensure information integrity on private messaging platforms.

[1] Unlike fully encrypted services such as Signal or WhatsApp, Telegram's end-to-end encryption applies only to "secret chats," while the platform's default mode leaves most content accessible to the company and, effectively, to external actors who can exploit its openness.

Background on Messaging Platforms

Private messaging platforms are services that enable two-way or multi-way digital communication between a finite number of users, often protected by end-to-end encryption. They include global applications such as WhatsApp, Telegram, Signal, Viber, and Messenger, each of which now combines traditional messaging functions with features typical of social media: group chats that can reach thousands, broadcast "channels," in-app commerce, and public "stories." The introduction of artificial intelligence chatbots marks the latest evolution of these platforms.

These platforms have evolved from simple peer-to-peer tools into hybrid communication ecosystems that blur the line between private and public spaces[2]. For example, Telegram hosts millions of public channels where administrators can broadcast to unlimited audiences, while simultaneously offering encrypted one-to-one chats. WhatsApp's introduction of "Communities"—umbrella spaces linking up to 100 group chats—exemplifies how messaging platforms have scaled communication while retaining the veneer of privacy.

Such semi-public uses—large, encrypted groups, viral forwarding, and broadcast lists—now represent one of the most dynamic and least regulated fronts of the information environment. While messaging platforms differ in ownership and business model—WhatsApp and Messenger under Meta, Telegram privately held, Signal non-profit, Viber part of Japan's Rakuten Group—they share global reach. WhatsApp alone counts more than two billion users across over 180 countries; Telegram exceeds 900 million, and Signal has tens of millions. In many regions, especially the Global South, these applications substitute for the open web as primary news channels and coordination tools.

Major messaging platforms, main features, and business models

This convergence has complicated both governance and user expectations. Their encryption protocols serve crucial privacy and security purposes, particularly in repressive or high-risk contexts. Yet the same protections hinder independent auditing or real-time moderation.

Selected messaging platforms: Encryption & Features

	LEVEL OF ENCRYPTION	RELEVANT FEATURES
WhatsApp	***Partial*** End-to-end encryption (E2EE) applied by default to all features except channels; meta data is also not E2EE.	• Individual chats • Group chats (max 1,024 members) • Communities (i.e., a "supergroup" of max 100 groups) • Intra-app forwarding • Status updates • Broadcast lists (up to 256 individual contacts) • Channels (unlimited audience; not E2EE) • WhatsApp Business Platform • Ads on Channels & Status Updates • Generative AI-powered chatbot and image creation (Meta AI), availability varies according to country

[2] Olaizola Rosenblat et al (2024), Covert Campaigns: Safeguarding Encrypted Messaging Platforms from Voter Manipulation, p. 4-6.

Telegram	***Low*** E2EE only for individual "secret chats," voice and video calls; all other app functions are not E2EE.	• Group chats (max 200,000 members) • Intra-app forwarding • Channels (unlimited audience) • Secret chats (one-to-one chats under E2EE) • Global search function allowing users to search groups • Newsfeed • Stories • Telegram for Business • Ads in Broadcasting channels (official Telegram Ads) and push-style ads, banners, videos etc in Mini Apps (other Ad providers) • AI chatbots can be integrated
Signal	***Full*** E2EE applied to all content and metadata.	• Individual chats • Group chats (max 1,000 members) • Intra-app forwarding • Stories • Stickers • No advertisement • No official AI chatbot
Viber	***Partial*** E2EE applies only to individual chats, group chats, and 1-on-1 calls; all other features have encryption-in-transit.	• Individual chats • Group chats (max 250 members) • Intra-app forwarding • Communities (groups with unlimited membership) • Private channels (require invite links) • Public channels (searchable and open to anyone) • Chatbots • Viber Business Messages • Stickers • Ads on chat lists, home screen, business chats, calls, etc.

Updated from Covert Campaigns: Safeguarding Encrypted Messaging Platforms from Voter Manipulation.

Other services that offer private messaging features are Facebook Messenger, which is characterized by end-to-end encryption and includes advertising, as well as LinkedIn messaging which does not offer E2EE and enables sponsored messages.

Rationale and Scope

The purpose of this report is to identify practical, rights-respecting solutions for strengthening information integrity on private messaging platforms. It draws on the FID's policy recommendations on mixed public and private messaging services; existing research on the use of covert influence operations across open and encrypted environments; questionnaire responses from participating governments summarizing their regulatory and policy approaches; and Workstream discussions reflecting multistakeholder perspectives from governments, regulators, and civil society.

Taken together, these inputs demonstrate that governments and platforms must balance three objectives: (1) Preserving privacy and encryption as enablers of human rights; (2) mitigating the systemic risks of virality, coordinated inauthentic behavior, and covert manipulation; and (3) promoting societal resilience through transparency, literacy, and cross-sector cooperation.

2. HOW MESSAGING PLATFORMS IMPACT THE INFORMATION ENVIRONMENT

Mechanics of Disinformation Dissemination

Messaging platforms have transformed interpersonal communication, enabling billions of people to share information instantly and often securely. Yet the same affordances that make them indispensable for private communication also make them fertile ground for manipulation. Covert influence operations employ at least three strategies to achieve mass communication under the guise of private messaging[3].

The "broadcasting toolkit"

The first strategy involves the systematic use of platform features that make private communication function like mass media. Forwarding, group links, broadcast lists, channels, and "community" functions create a chain of amplification that is largely invisible to outsiders. Through these features, messages crafted by a few operators can reach thousands of recipients within minutes, often via a sequence of trusted intermediaries. Because users experience each message as a personal or group exchange, the information feels authentic and intimate even when it originates from a coordinated campaign.

Forwarding limits and other friction-reducing measures help but do not eliminate the problem. The architecture itself favors velocity: messages move horizontally across overlapping networks rather than vertically through public feeds. The result is an information space that mirrors social media in reach but differs in accountability, since its circulation is largely unobservable to regulators, researchers, or the public.

Exploitation of business-messaging infrastructure

A second strategy capitalizes on the privileges granted to verified or commercial accounts. Many messaging platforms provide business platforms and automation tools that allow high-volume communication with customers or subscribers. These systems are intended for legitimate outreach, yet they also furnish covert actors with the means to deliver political or propagandistic material at scale.

By registering as small businesses, news outlets, or civic initiatives, operators obtain accounts that can send bulk messages, include multimedia attachments, and avoid spam filters. Some integrate these channels with online advertising to draw users into chat groups subject to high-volume messaging. In practice, this converts a commercial marketing ecosystem into a pipeline for political manipulation.

[3] Olaizola Rosenblat et al (2024), Covert Campaigns: Safeguarding Encrypted Messaging Platforms from Voter Manipulation

Sock-puppet coordination and identity laundering

A third strategy involves the creation of networks of fabricated identities—so-called "sock puppets"—that impersonate ordinary users. These accounts join numerous groups, cross-post identical content, and reinforce one another's messages to create the appearance of consensus. Because participation in group chats or channels often rests on social trust or invitation, recipients seldom question whether a sender is real.

Weak identity controls further enable these operations. Where messaging platforms allow multiple accounts per device or registration through virtual phone numbers, a single operator can manage dozens of personas simultaneously. Coordinated timing and repetition give the impression of widespread grassroots engagement, when in fact the discourse originates from a small set of organized manipulators.

Together, these strategies convert tools meant for connection and commerce into instruments of influence. Indeed, empirical evidence from the NYU Stern Center's 2024 comparative study on voter manipulation indicates that these strategies can be effective. In a survey of messaging app users in nine countries, 52% of respondents said that the political content they had received from strangers on those apps had "significantly" or "somewhat" influenced their opinions[4]. The line between private conversation and public broadcasting thus becomes blurred, allowing large-scale information manipulation to flourish in spaces that remain beyond conventional oversight.

In addition, the latest evolutions of these platforms, notably the introduction of advertising and the integration of artificial intelligence features, bear threats to information integrity.

The monetization of disinformation

As platforms' business models rely mainly on advertising revenue, their objective is to increase user engagement. This attention economy has led to the monetization of disinformation, as disinformation drives engagement shown by the meta-analysis of the Observatory on Information and Democracy[5]. The lack of transparency of the advertising market enables inadvertent funding of disinformation and makes it nearly impossible for even well-intentioned brands to avoid contributing to the problem. The introduction of advertising on private messaging platforms bears the risks of repeating similar patterns and leads to addictive design choices that compromise content quality in a search to encourage user engagement with features where advertising revenue can be earned.

[4] Olaizola Rosenblat et al (2024), Covert Campaigns: Safeguarding Encrypted Messaging Platforms from Voter Manipulation, p. 17. The nine countries surveyed were: South Africa, Mexico, the Philippines, Turkey, the U.S., Indonesia, India, Brazil, and Hungary.

[5] Forum on Information and Democracy (2025), Observatory on Information and Democracy: Information Ecosystems and Troubled Democracy

Artificial intelligence

Private messaging platforms are introducing AI functionalities in their services with the promise to facilitate conversations and provide the benefit of AI capabilities such as summaries or content creation. Yet, as AI models generally do not run on individual devices, the integration of AI opens a communication channel with shared AI models. This creates cybersecurity risks and might break encryption, especially if input data is used for inference or training of the AI model. Beyond encryption risks, the integration of AI could potentially render disinformation campaigns more effective, cheaper, widescale and credible.

Country case studies

Brazil

Brazil's experience illustrates how messaging platforms can become engines of political mobilization and disinformation. During Brazil's 2024 municipal elections, the circulation of manipulated political content on messaging platforms followed clear patterns of coordination and amplification. A study by the Alafia Lab[6] found that while 17% of links shared in WhatsApp groups contained disinformation, disinformation was much higher when looking at specific topics, with the highest amount circulating about the Federal Supreme Court (40%). Many links led to opinion pieces presented as news, blurring genre boundaries and eroding users' ability to distinguish commentary from reporting. In numerous instances, headlines contained the deceptive element while the article body included accurate information—a deliberate strategy to attract attention and mislead readers. In addition, opinion pieces were presented as factual reporting. While WhatsApp groups tended to favor rapid dissemination of disinformation on (international) politics or economics, Telegram groups tended to circulate fabricated information on the Supreme Court.

Lebanon

In Lebanon, the interplay between economic crisis, political polarization, and declining trust in media created conditions for large-scale manipulation and hate speech through WhatsApp.

WhatsApp has become Lebanon's dominant news channel after television: roughly 84% of Lebanese adults use the app, and more than half trust information shared through it more than they do traditional media. This shift accelerated after the 2019 protest movement, when distrust of sectarian and privately-owned television and newspapers pushed many toward hyper-local WhatsApp "news" groups.

[6] Alafia Lab (2025), Abaixo do radar: Desinformação em grupos de extrema direita no WhatsApp e no Telegram nas eleições de 2024. The study monitored 35 WhatsApp and 22 Telegram groups aligned with the far right during the official campaign period (2 September–27 October 2024), analyzing thousands of messages containing 480 unique links on WhatsApp and 192 on Telegram.

In 2023, Lebanon slipped into an acute social, economic, and institutional crisis, which provided an opportunity for political propagandists to amplify sensationalistic and conspiratorial content. According to research by the Samir Kassir Foundation, some topics such as currency rates were discussed more often in news-oriented groups than in traditional media, and often stemmed from questionable sources, promoting speculation and fear. With respect to security, one of the most discussed and debated topics, it included news on events not covered by traditional media, suggesting potential links between security forces and WhatsApp group administrators. Unethical reporting, creating a sense of looming threats and news items without or with questionable sources characterize these groups. The groups also replicated each other's information. They are not registered as official news agencies, but some impersonate reputable media (e.g., Annahar, Addiyar, BBC Arabic).[7] The study also highlights that WhatsApp provides a low-cost advertising opportunity to reach a large audience, that groups can be infiltrated by false accounts, and people can bypass local regulation through foreign sim cards.

These cases underscore a shared challenge: how to preserve the integrity of information flows on messaging platforms and the trusted character these spaces represent, without undermining the privacy and security that make them valuable. The next section examines how different jurisdictions have attempted to govern messaging platforms—through regulation, national strategies, and media-literacy initiatives—and what lessons can be drawn for proportionate, effective oversight.

[7] Samir Kassir Foundation (2023), WHATSAPP 360: A Look into the WhatsApp News Ecosystem in Lebanon Focusing on Misinformation and Hate Speech. The study mapped 37 public news-oriented groups across all seven governorates, reaching 59,653 users directly and an estimated three million through extended networks of mirrored channels.

3. GOVERNANCE OF MESSAGING PLATFORMS

Regulation

Where "private" meets "public"

Across jurisdictions, regulators struggle to delineate what counts as a "private" online space that can be distinguished from "public" fora. The distinction is not merely semantic: it determines which obligations apply, whether under general online safety laws, platform regulation or specialized disinformation frameworks.

Questionnaire responses from governments[8] show that this definitional question remains unsettled almost everywhere.

The EU's Digital Services Act (DSA) adopts a differentiated approach to regulating "intermediary services," but the classification of private messaging platforms under that scheme has emerged through gradual interpretation rather than through clear ex ante rules. Recital 14 explicitly excludes interpersonal communication services, defined as "communications between a finite, that is to say not potentially unlimited, number of natural persons," from the definition of online platforms. However, the European Commission has taken a more flexible, feature-based approach when clarifying regulatory obligations. This approach became explicit in January 2026, when the Commission designated WhatsApp as a Very Large Online Platform (VLOP) based on its monthly active user base exceeding 45 million, treating WhatsApp Channels, rather than private messaging, as an online platform. Meta has until May 2026 to comply with the resulting obligations, including systemic risk assessments of its Channels feature. In parallel, the Commission is assessing whether Telegram meets the VLOP threshold, a designation Telegram contests, and whether comparable obligations should apply to certain of its features.

In the interim, messaging services have taken divergent positions: WhatsApp states that its DSA transparency reporting covers its messaging service, Telegram limits reporting to optional or ancillary features, and Viber does not specify which services are covered in its reports. In addition, the EU Code of Conduct on Disinformation as foreseen by Article 45 of the DSA, specifically includes messaging apps, notably in its commitment 25 focusing on empowering users. As adherence is voluntary, among the messaging platforms examined in this paper, only Meta, with WhatsApp and & Messenger, subscribed to that commitment[9]. The other messaging platforms (or their companies) are not signatories of the Code. As a result, the EU's regulatory approach reflects a growing but incremental move toward feature-based classification of private messaging platforms, one that is being developed through staged designation decisions and enforcement practice.

[8] The following countries provided replies to the questionnaire, either via their government or independent regulatory institutions: Armenia, Australia, Croatia, France, Greece, Ireland, Lithuania, Luxembourg, North Macedonia, Portugal, United Kingdom & Ukraine

[9] EU CODE OF PRACTICE ON DISINFORMATION 2025: Subscription Document for Meta, https://disinfocode.eu/storage/subscription-forms-files/meta-1.pdf

By contrast, the UK under the Online Safety Act (UK OSA) has taken a clear and an expansive approach from the beginning, explicitly categorizing private messaging as a form of user-to-user service, thus subject to many of the same transparency and risk-assessment duties as social-media platforms.

Other jurisdictions—France and Croatia, for example—regulate platforms only to the extent that they host some public content, leaving purely private spaces outside formal oversight. This partial coverage approach attempts to respect privacy but risks allowing disinformation operations to flourish in semi-public environments such as large groups or channels. The privacy of communications is guaranteed in many countries, as in the Constitution of the Republic of North Macedonia.

From a governance standpoint, the lack of consistent definitions undermines both enforcement and interoperability. Several respondents emphasized the need for international clarification of what constitutes a private, semi-public, or public space online—a recurring theme that informs this report's recommendations on proportionality and feature-based regulation.

Illegal content versus disinformation

Legislators and regulators generally distinguish between illegal content and disinformation on online platforms, but they diverge significantly in how these categories are defined and governed. Illegal content is typically grounded in criminal or administrative law, linked to judicial procedures, and subject to clear enforcement mechanisms. Disinformation, by contrast, occupies a more contested regulatory space: some policymakers treat it as a subset of illegal content, while others frame it as a distinct category better addressed through transparency obligations, education, and forms of co- or self-regulation. This distinction is particularly consequential for private messaging platforms, where enforcement tools are constrained by encryption and expectations of privacy.

Australia illustrates a relatively clear separation between these two approaches. Illegal content is addressed through binding legal instruments, while disinformation is largely governed through voluntary measures. The Criminal Code Act 1995 creates offences targeting the dissemination of certain types of content, such as child abuse material or violent extremist material, using a carriage service, which includes private messaging platforms. The Australian Online Safety Act (Australian OSA) complements this framework by regulating illegal and seriously harmful online content across a wide range of services, including those using end-to-end encryption. Under the Act, the eSafety Commissioner can compel providers—including relevant electronic services such as iMessage and WhatsApp—to report on their handling of illegal and non-illegal material and request the removal of content, through both formal removal notice powers and informal engagement with platforms. The eSafety Commissioner can also request information from platforms on how they comply with Basic Online Safety Expectations, a key element of the Australian OSA.

Disinformation, however, is treated differently. Australia's voluntary Code of Practice on Disinformation and Misinformation, developed by the industry association Digital Industry Group Inc (DIGI), excludes most messaging services, even semi-public group messaging platforms playing a significant role in dissemination of harmful material. The Australian Communications and Media Authority (ACMA) has called for inclusion of these services in the Code to establish minimum industry standards.

The UK adopts a more integrated approach under the UK OSA, which regulates both illegal content and certain types of content designated as harmful to children. Crucially, the Act explicitly applies to private messaging as "user-to-user services." In practice, this means messaging platforms must conduct risk assessments on features that enable private communications; mitigate exposure to illegal material, including terrorism and child-exploitation content; and manage complaint mechanisms--even within encrypted environments. While the UK framework still draws a line between illegal and harmful-but-legal content, it brings private messaging more clearly within the scope of statutory oversight.

The EU's approach under the DSA differs again. The DSA imposes baseline obligations on all covered services with respect to illegal content, transparency reporting, designated points of contact, notice-and-action mechanisms and complaint systems. Obligations related to disinformation, however, are primarily tied to the mitigation of "systemic risks" and apply only to services designated as Very Large Online Platforms (VLOPs) or Very Large Online Search Engines (VLOSE). As a result, disinformation is not treated as illegal per se under EU law but as a risk that may warrant additional duties for certain services under specific conditions.

In practice, this distinction means that most private messaging platforms operating in the EU are accountable primarily for illegal content, not for the spread of disinformation. Telegram[10], Viber[11] and WhatsApp[12] publish transparency reports, have designated points of contacts and legal representatives, and put in place notice-and-action mechanisms as required. Yet these measures largely focus on illegal content such as violence, fraud, and child sexual exploitation. Only WhatsApp explicitly reports "misinformation" as a type of illegality. The terms of service of private messaging platforms focus mainly on illegal content, violence and fraud. As a result, regulators have limited grounds under the DSA to hold messaging platforms accountable for failures to address disinformation or large-scale manipulation, unless specific features of those platforms are brought within the VLOP regime.

[10] Telegram: User guidance for the EU Digital Services Act, https://telegram.org/tos/eu-dsa
[11] Viber: EU Digital Services Act (DSA) Transparency Report, www.viber.com/en/terms/eu-digital-services-act-dsa-transparency-report/
[12] WhatsApp: Regulatory and Other Transparency Reports, www.whatsapp.com/legal/transparencyreports

Finally, some countries address disinformation indirectly through criminal or sector-specific laws. Criminal liability may attach to false information where it qualifies as defamation, as in Luxembourg, or where laws prohibit the dissemination of certain types of content. The Lithuanian Law on the Provision of Information to the Public, for example, forbids spreading disinformation, war propaganda, incitement to war, the encouragement of violence, or promoting or inciting terrorist crimes. Ukraine's Law on Media includes content restrictions on calls for violent change or overthrow of the constitutional order, incitement of hatred, discrimination or terrorism. However, it often remains unclear whether and how these provisions apply to messaging platforms, particularly in encrypted or close communication contexts.

The encryption challenge

One of the challenges regulators face is end-to-end encryption, which serves to protect privacy and human rights. How to treat and detect information that is shared in encrypted spaces is a regular point of contention and debate, often with regulatory proposals that are a threat to end-to-end encryption or would weaken it effectively.

The EU is faced with an ongoing debate to implement regulation regarding Child Sexual Abuse Material (CSAM) (draft Combat Child Sexual Abuse Regulation), which would cover private messaging platforms and pose risks to encryption. In its latest version adopted on 26 November 2025 by the European Council the draft proposes voluntary client-side scanning[13], a technique not compatible with cybersecurity and encryption, as automatic reporting can leak content to third parties. Moreover, even without automatic reporting to third parties, client-side scanning is problematic because of the unreliability of the technology that underlies it—in particular, the high number of false positives and the sensitivity of the data needed to train the models.

In the UK, under the Online Safety Act, encryption remains a point of contention. Under Chapter 5, the UK's regulator for communications services, Ofcom, may give a notice requiring a service to deploy "accredited technologies" to identify and remove terrorism and child-exploitation materials, although Ofcom must be satisfied it is necessary and proportionate to issue such notices, taking account of a range of factors. This power stops short of mandating client-side scanning but leaves the possibility open, giving rise to privacy concerns.[14]

While in Australia, the industry standard "Relevant Electronic Services Standard" requires services to detect known CSAM and pro-terror material, it only applies if it is technically feasible and reasonably practicable for the service. If not, risk mitigation measures or alternative actions are needed.

[13] European Council (2025), Child sexual abuse: Council reaches position on law protecting children from online abuse, www.consilium.europa.eu/en/press/press-releases/2025/11/26/child-sexual-abuse-council-reaches-position-on-law-protecting-children-from-online-abuse/

[14] Kubi, G. (2025), Encryption Under Threat: The UK's Backdoor Mandate and Its Impact on Online Safety www.internetsociety.org/blog/2025/05/encryption-under-threat-the-uks-backdoor-mandate-and-its-impact-on-online-safety/

In Brazil, the regulator tried to bypass the encryption issue in mandating meta data retention in its Draft Bill 2630/2020, which was eventually withdrawn following civil society advocacy. Data Privacy Brazil argues that "data transmission such as time, date, and participants—is as revealing as the content itself."[15] Moreover, mass data collection also poses a danger when the rule of law is attacked after regime changes.

Oversight is dispersed

The lack of comprehensive strategies or regulation addressing private messaging platforms is also reflected in scattered oversight responsibilities of such platforms.

Most countries that replied to the questionnaire mentioned some responsibilities of the Data Protection Authority, notably with regard to their role in privacy protection. In addition, regulatory institutions responsible for platform regulation might have some responsibilities, such as the Office of the eSafety Commissioner responsible for Australia's Online Safety Act, Ofcom, and the European Digital Services Coordinators.[16] Relevant ministries and media regulators are also involved in information integrity policies. Ukraine plans to establish a Unified Digital regulator.

This dispersal of oversight, despite some attempts for coordination among different agencies and institutions, hinders effective monitoring of these platforms and the challenges they pose to information integrity.

Countries shared several challenges with regard to oversight of these platforms and promoting information integrity. These range from low public digital literacy, cross-border disinformation flows, insufficient resources for monitoring and response, the lack of public access to data or public transparency reports and the use of anonymous channels. Inadequate or lack of official cooperation channels between authorities and platforms further hinder their effective oversight.

Approaches to Regulating Private Messaging

• Illegal content but not disinformation • Messaging included (e.g., Australia)	• Illegal + disinformation • Messaging included (e.g., UK)
• Illegal + disinformation • Messaging excluded (EU includes messaging for illegal content, disinformation only once designated as VLOP)	• Soft-law / media literacy (e.g., Armenia, North Macedonia)

The scope of regulation remains inconsistent and sometimes ambiguous, especially regarding whether private messaging qualifies as a public-facing service, and disinformation is treated alternately as illegal, harmful, or merely undesirable content. Against this backdrop, media literacy has become the default policy tool where direct regulation is constrained by privacy or technical limitations.

[15] Data Privacy Brazil submission to the workstream (2025)

[16] In the European Union, the Belgian DSC is responsible for Telegram and the Irish for WhatsApp.

National Action Plans and Media-literacy Strategies

In lieu of direct regulation, many governments have invested in media-literacy initiatives and national strategies to foster public resilience against disinformation, sometimes these also include media self-regulatory strategies, fact-checking initiatives and user reporting. These approaches vary from broad national frameworks to targeted campaigns. Some examples include:

- **Ukraine:** The Action Plan for the Implementation of the Information Security Strategy and the Filter National Media Literacy Project integrate education, civic engagement, and cross-sector coordination. Ukrainian users can also report disinformation through a dedicated portal.
- **Ireland:** The Be Media Smart campaign and the Media Literacy Ireland Network combine state and NGO leadership, emphasizing critical-thinking skills and cross-platform awareness. The country also has a National Counter Disinformation Strategy.
- **Armenia:** The country has co-developed its Strategy against Disinformation with civil society which includes, among others, strategic communication of state institutions.
- **Lithuania:** Media and Information Literacy Competence Development Program implemented through the network of public libraries, which organize educational activities for various target groups across the country.
- **North Macedonia:** Its Action Plan Against Disinformation (2019) includes awareness campaigns and capacity-building for journalists. It complements strategies on cybersecurity, resilience and hybrid threats.
- **Australia:** Supports media-literacy programs as part of its Online Safety Strategy, with significant financial investment and accessible eLearning tools for schools.
- **Luxembourg:** Combines a guide for media education in schools («Media Compass») with public awareness campaigns focusing in particular on groups considered as vulnerable: young people and seniors and a center for political education.

These strategies demonstrate a growing recognition that education and societal resilience are indispensable complements to regulation, especially in encrypted or semi-private environments where direct moderation is infeasible. Yet, none of these strategies and action plans focuses specifically on the challenges of private messaging platforms and further indicators and measurements of their success would need to be investigated.

4. RECOMMENDATIONS

Based on discussions during the Workstream, research and approaches shared by its participants and previous works of the FID and NYU Stern Center for Business and Human Rights, the Forum on Information and Democracy recommends the following policy interventions to both governments and platforms. It should be noted that the following recommendations, while co-developed with the inputs from signatory states of the Partnership for Information and Democracy, do not represent their official views and commitments.

To governments:

1. Regulate platform features rather than platforms as a whole.

Lawmakers should design rules that attach obligations to specific platform features and affordances rather than categorizing entire platforms as one type of service. While internet applications once tended to be single-purpose, today's platforms combine many functions—ranging from public broadcasting to private, one-to-one messaging—each carrying distinct risks and justifying different levels of government oversight.

Recognizing this diversity, regulation should differentiate obligations based on the features in question. Enforcement bodies in jurisdictions with existing frameworks should also clarify how obligations such as risk assessments or notice-and-take down requirements apply to particular features. For instance, under the EU Digital Services Act (DSA), the European Commission and national Digital Services Coordinators should specify which features of the private messaging platforms fall within scope to reduce ambiguity and ensure proportionate implementation.

In the same spirit, and as stated in the 2020 report of the Forum on How to End Infodemics, states should impose obligations on platforms to "inform users on any new features that have an impact on the design of an app, in all languages and dialects where the service operates"[17] and enable opt-in and out of these features.

2. Clearly distinguish between "public" and "private" venues.

Regulation should differentiate obligations based on whether communications take place in "public" or "private" venues, but these terms must be carefully defined. Too often, laws and implementation guidance invoke this distinction without explanation, creating uncertainty for platforms and enforcers.

[17] Forum on Information and Democracy (2020), How to End Infodemics

Regulators should articulate clear criteria for what counts as "private" versus "public" in the online environment. Relevant factors may include:

- Audience size (how many users can access the content)
- Discoverability (whether groups/channels are searchable)
- Encryption (use of end-to-end encryption)
- Access controls (invite-only, admin approval, or other restrictions)
- Shareability (ease of forwarding or redistributing content)

Regulators should also provide clear criteria on what kind of AI features and ads can be introduced and under which circumstances with the objective of curbing disinformation and guaranteeing encryption.

Such clarity will help ensure that obligations are applied proportionately and consistently across different types of online spaces.

3. Do not create obligations that would render all end-to-end-encrypted messaging untenable.

End-to-end encryption is essential to the right to privacy and the secure exercise of human rights, including freedom of expression and association, particularly in repressive contexts. Regulation should not create obligations that make E2EE technically or legally impossible. Sweeping mandates—such as requiring all platforms, without differentiation, to proactively scan and report harmful or illegal material—are incompatible with encrypted services, where providers cannot access user content. Put simply, automatic scanning and reporting cannot coexist with genuine E2EE. To protect both safety and privacy, lawmakers should ensure that obligations requiring proactive detection combined with third-party reporting apply only to non-encrypted features.

Relatedly, lawmakers and those in charge of implementation should avoid vague or open-ended language when imposing obligations that might apply to end-to-end encrypted services. Terms such as "reasonable methods" create significant uncertainty about what platforms are expected to do, and with respect to which features. Clear, precise drafting is essential to prevent overbroad interpretations that could undermine the viability of end-to-end encryption.

Similarly, states should not "encourage the collection, storage and usage of the metadata of all users' communications."[18] The storage of data bears privacy concerns and a presumption of guilt.

[18] Forum on Information and Democracy (2020), How to End Infodemics

4. Impose transparency and user empowerment.

States should design and impose transparency and user empowerment measures that enable them to better understand and monitor these platforms and their different features.

This should include transparency about platforms' terms of service and their implementation, how technical protocols and algorithms operate and with what objectives, how users' metadata is used. There should also be transparency about what activities were conducted to identify risks of harm, the negative effects of their activities on users and other people, and take-down requests and outcomes of complaint handling systems.

Platforms should also be required to provide granular data on content in public spaces enabling researchers, regulators and the public to understand the reach and impact of specific messages and the tactics of disinformation.

In addition to the reports published by platforms, "states should publish detailed transparency reports on all content-related requests issued to service providers."[19]

5. Support media-literacy initiatives.

When direct moderation is infeasible, as in encrypted or small-group environments, societal resilience becomes the first line of defense. Every jurisdiction surveyed recognized media literacy as an essential complement to regulation. Governments should embed media literacy within national information-security strategies, create measurable benchmarks, and ensure sustained funding and evaluation rather than ad-hoc campaigns.

Models to emulate and iterate upon include Ukraine's "Filter" Project, which integrates formal education with fact-checking partnerships and public-sector transparency, and Ireland's Media Literacy Ireland Network, which provides a platform for multistakeholder coordination linking broadcasters, NGOs, and regulators.

6. Encourage international responses.

Given the often international or cross-border nature of disinformation campaigns and the international character of these companies, international responses are needed. This starts with spaces for policy exchange and learning such as the Partnership for Information and Democracy and furthermore requires joint actions and global agreements.

In the European Union, the Rapid Alert System to strengthen coordinated and joint responses to disinformation could also pay attention to private messaging platforms. At the global level, UN instruments as a resolution by the UN Human Rights Council could provide a way forward on private messaging platforms covering their human rights risks.

[19] Forum on Information and Democracy (2020), How to End Infodemics

To platforms:

1. Separate messaging, broadcasting and AI functions.

The tendency of messaging platforms toward feature bloat undercuts the platforms' mission as private and secure channels for communication and leads to confusion among users. Platforms like WhatsApp, Telegram, and Viber, which have expanded into social media territory with their breadth of virality-promoting features, should establish a clear separation between their messaging service, meant for individual and small-group chats, their social media features and their AI integration.

Platforms should consider bifurcating their services into separate apps—one purely for private messaging, in which all content and metadata are protected with end-to-end encryption, and another for broadcasts, channels, stories, and large groups, in which content is left in plaintext and moderated rigorously. AI integration into encrypted spaces should be avoided to guarantee E2EE, but it can be integrated in other functionalities as long as users are asked to opt-in. Similarly, advertising should not be introduced into encrypted spaces as it aids disinformation dissemination.

Establishing a clear separation between encrypted and non-encrypted functionalities would also help users have more accurate mental models of the security and privacy guarantees of each side of that separation, thereby minimizing the risks that come with having a false sense of security and privacy.

2. Apply end-to-end encryption judiciously.

End-to-end encryption is most appropriate in contexts where users have a reasonable expectation of privacy, such as one-to-one or small group communications. Extending E2EE to very large group chats or broadcast channels, however, undermines that rationale. For instance, WhatsApp currently allows up to 1,024 members in a single encrypted group chat, and its "communities" feature can link as many as 100 groups—creating spaces that resemble mass broadcasting rather than private communication.

To strike the right balance, messaging services should limit the audience size of end-to-end encrypted groups and introduce friction to viral forwarding. For example, restricting the number of simultaneous forwards (as WhatsApp does with its cap of five recipients) can help mitigate the risks of rapid, uncontrolled amplification while preserving secure communication where it is most needed.

3. Set human-rights-respecting rules and measures.

As a general rule, private messaging platform providers should apply the United Nations Guiding Principles on Business and Human Rights and integrate human rights into their products and systems by design and by default.[20]

[20] Forum on Information and Democracy (2020), How to End Infodemics

They should establish clear terms of service, in compliance with international human rights law and standards that specify clearly what types of content and activities are prohibited on the provider's services.

They should also conduct evaluations of the risks of their functions and content moderation decisions, investigating their impact on fundamental rights and enabling users to easily notify them of content they consider to be in breach of the terms of service and applicable laws, and inform users about the moderation decisions and provide avenues for review. [21]

4. Curb coordinated inauthentic behavior.

To remain trustworthy channels for communication, messaging services need to crack down on coordinated inauthentic activity. Phone farms and trolls dedicated to spreading manipulative content on messaging platforms rely on the ability to create and manage multiple accounts from a single device. One of the most direct ways for messaging services to curb political manipulation campaigns is by limiting the number of accounts someone can create and manage with a single device. Therefore, from the point of view of abuse mitigation, platforms should place limits of one account per device, as Signal does, or up to two accounts if the platform offers business messaging services, as do WhatsApp and Viber. Telegram's policy of allowing the operation of multiple accounts from a single device seriously undermines the authenticity of communications on its service and facilitates its exploitation. This should also include banning the use of external tools not approved by the service provider to manage accounts.

However, platforms should consult researchers who have extensively studied benevolent uses of the apps and undertake empirical studies to determine an account limit that adequately balances the legitimate use of multiple accounts with the need to curb phone farms. Short of establishing absolute limits on the number of accounts per device, platforms should place limits on the pace at which new accounts can be created.

Empowering users is another gateway to counter inauthentic behavior, including requiring users to opt-in to receiving group messages, allowing them to choose who can add them to groups as well as labeling messages created by bots or business accounts.[22]

5. Support and improve access to accredited tiplines.

Platforms play a critical role in safeguarding their products from manipulation, but users are not powerless either. In fact, on end-to-end encrypted platforms, users are the parties best positioned to control their information diets. Researchers have proposed a number of tools for facilitating user-driven fact-checking on messaging apps, which platforms should consider implementing or improving. A large majority of messaging app users surveyed expressed a demand for such tools, and none of these affordances would violate the privacy and security guarantees of end-to-end encryption.

[21] Forum on Information and Democracy (2020), How to End Infodemics
[22] Forum on Information and Democracy (2020), How to End Infodemics

One approach involves supporting the operation of verified "tiplines." These are dedicated messaging app accounts managed by independent media or civil society organizations to which users can submit "tips" for fact-checking. Platforms do not control tiplines, nor do they have access to the information exchanged. The decision of what to do with the results of the fact-checking process is entirely that of the user who submitted the evidence and of the organization that provides the fact-checking. But platforms can play a constructive role by authenticating tiplines and providing an accessible user interface. Currently, few users know about or use tiplines, and the cumbersome way in which tiplines are accessed is part of the problem. According to a survey carried out by the NYU Stern Center for Business and Human Rights in 2024, only 7% of app users across the nine countries surveyed said they had ever contacted a tipline, although 83.4% said they would find such an option useful.

6. Empower users through in-app verification tools.

There are other forms of user-driven fact-checking that platforms can directly support through dedicated in-app affordances. One idea, suggested by Kiran Garimella, a professor of information science at Rutgers University, is for messaging apps to implement "one-click reverse image search" tools. Google's News Initiative offers such a tool, which empowers users to find information about an image—including its provenance, history of previous use, and any related images—through a series of simple steps. Messaging platforms should consider collaborating with an Internet search function to enable recipients of images to quickly check the image for relevant information. In this vein, WhatsApp has piloted an "Internet search for forwarded messages" tool. However, the tool is currently available in a minority of countries where WhatsApp operates.

Another user-driven fact-checking affordance that has been proposed, but mostly in theoretical discussions among scholars of content moderation in encrypted environments, is a hash matching and flagging system for known disinformation installed on users' devices. Such on-device matching and flagging is consistent with the privacy guarantees of end-to-end encryption *as long as the matching process and result are kept strictly on device*—that is, no one outside of the communication participants learns any new information. This proposal is different from the so-called "client-side scanning" method, which does violate the privacy of users' communications because it is set up to automatically release information about a match to third parties, such as the platform or law enforcement agencies.

The benefit of an on-device matching system is that it empowers users with information about content that they are about to send or receive, alerting them if it contains previously fact-checked information and informing them of the source of the fact-checking. Users retain full agency—they can decide to send or receive the message regardless of the flag. A similar "speedbump" approach has been implemented on social media to caution users before sending images containing nudity—a warning that they can choose to ignore.

However, an on-device matching method for known misinformation is still merely a viable proposal, one that has not been tested or implemented. As such, platforms should consider supporting and monitoring academic research into such techniques but should not rush into rolling out a feature that might prove too intrusive or counterproductive.

A potential intermediate step might be to create optional plug-ins or extensions for on-device flagging that messaging app users can install on their devices. But doing so should not be a prerequisite for using the messaging service.

7. Engage in cross-platform cooperation.

Disinformation campaigns often span several platforms, ranging from various messaging platforms to social media. Platforms should therefore participate in information-sharing initiatives to detect coordinated inauthentic behavior across messaging and social media ecosystems.

8. Promote transparency and independent research.

Independent civil society, researchers and regulatory authorities play an important role in researching coordinated inauthentic behavior and making suggestions to improve private messaging platforms. Platforms should proactively share information on feature functionalities and content moderation decisions while respecting privacy to enable independent evaluation. This should also include granular data on notice-and-takedown procedures.

CONCLUSION • The path forward

Over the past year, the Workstream on Identifying Solutions to Protect Information Integrity on Private Messaging Platforms co-chaired by Luxembourg and Ukraine convened expert exchanges to analyze the challenges and discuss solutions. The deliberations highlighted several core insights.

First of all, these platforms, while originally conceived for private one-on-one communication, have evolved and are constantly evolving to take on more and more features of online platforms. The introduction of artificial intelligence and advertising are the latest changes. These are exploited for widespread coordinated disinformation campaigns, election manipulation and information warfare. The "trusted" nature and the mixed public and private space make them particularly difficult to regulate. Legal obligations should thus be tied to specific features rather than broad platform categories. The diverging interpretations of legislation by both regulators and platforms themselves reflect an urgent need for regulatory clarity regarding different features and their obligations. Providers, with the same objective, should clearly separate private messaging functions from online platform features, and empower users by enabling opt-into features.

Secondly, private communication is differentiated from public communication in regulation, yet the definitions of public and private are not clear, when messaging platforms allow groups of thousands or several hundred-thousands. Regulators and legislators should establish clear definitions of public and private which consider several factors such as audience size, discoverability, encryption, access controls, and shareability.

Thirdly, encryption remains a contested issue. Regularly, legislative proposals attempt to weaken encryption in the name of security or child protection. Yet, end-to-end encryption is essential for privacy and protection and none of the suggested approaches to bypass encryption, such as client-side scanning, are viable without weakening it. Regulators must therefore clearly define private communications and avoid mandates that render E2EE technically unworkable.

Fourth, media literacy remains a key strategy to counter disinformation and encourage a responsible use of private messaging platforms. This needs to include national strategies to build societal resilience and adaptation to emerging trends such as the introduction of artificial intelligence and advertising.

This Workstream demonstrates that international exchange and cooperation can identify actionable pathways to promote information integrity on private messaging platforms. Implementing the recommendations presented in this report will determine if messaging platforms are governed by democratic rules and in the interest of human rights or continue to be vectors of disinformation and manipulation.

The designation by the European Commission of WhatsApp Channels as an online platform service that needs to comply with DSA VLOP obligations demonstrates that regulators are starting to recognizse the changing nature of private messaging platforms and their impact on information integrity.

ACKNOWLEDGMENTS

This report was drafted by the Forum on Information and Democracy and the NYU Stern Center for Business and Human Rights, the knowledge partner of the Partnership for Information and Democracy's Workstream on identifying solutions to protect information integrity on private messaging platforms. The Workstream is co-chaired by the government of Luxembourg, represented by its Ambassador for Cybersecurity and Digitalization, Luc Dockendorf, and the government of Ukraine, represented by Ganna Krasnostup at the Ministry of Culture.

The Workstream was coordinated by Katharina Zuegel, Policy Director at FID. Mariana Olaizola Rosenblat, Policy Advisor on Technology & Law at NYU Stern Center for Business and Human Rights acted as its rapporteur.

The report is based on the discussions held in the framework of the Workstream, which met for four online meetings on 27 March, 3 June, 11 September, and 6 November 2025. It also draws upon the replies provided by 12 countries to a questionnaire, including Armenia, Australia, Croatia, France, Greece, Ireland, Lithuania, Luxembourg, North Macedonia, Portugal, United Kingdom and Ukraine.

The Workstream benefitted from input from representatives of the 57 states of the Partnership for Information and Democracy as well as of members of the Forum's civil society coalition and other partners including Access Now, Alafia Lab, Chaos Computer Club, Centre pour le dialogue humanitaire, Conscious Advertising Network, Data Privacy Brazil, Digital Security Lab Ukraine, Institute for Strategic Dialogue, IT for Change, New York University, Research ICT Africa and Samir Kassir Foundation.

The recommendations were elaborated by FID and do not necessarily reflect the views of all partners involved.

THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of Foreign and European Affairs,
Defence, Development Cooperation
and Foreign Trade

Luminate

FORD FOUNDATION

MINISTÈRE
DE L'EUROPE
ET DES AFFAIRES
ÉTRANGÈRES
Liberté
Égalité
Fraternité