Annex I

Glossary

Algorithmic amplification

The process by which recommendation systems, ranking algorithms or automated curation tools on digital platforms prioritize and boost the visibility of certain content over others. This amplification is typically based on engagement metrics (e.g., clicks, shares, watch time), personalization or platform incentives, rather than the accuracy or trustworthiness of the information.

AI-generated fake content

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

AI-generated 'slop'

Refers to low-quality media, such as text, images or videos, created using generative artificial intelligence tools and characterized by mass production, minimal effort and lack of meaningful substance or artistic integrity. The term is a pejorative implication similar to 'spam,' highlighting superficial or banal AI outputs often produced solely for volume or engagement.

Artificial intelligence (AI)

The capability of machines, computer systems or software to perform tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, natural language understanding and decision-making.

Cheapfake

Manipulated or misleading media created using simple, low-cost or readily available editing tools.

Cognitive warfare

The use of information, psychological and technological tools to influence, disrupt or control the perceptions, beliefs and decision-making of individuals or groups. It targets the human mind as the contested domain, aiming to weaken trust, sow confusion and manipulate behaviour in ways that serve strategic or political objectives.

Community engagement and accountability (CEA)

A way of working that recognizes and values all community members as equal partners, whose diverse needs, priorities and preferences guide the IFRC's work. This includes integrating meaningful community participation, open and honest communication, and mechanisms to listen to and act on feedback.

Confirmation bias

A cognitive bias whereby people tend to seek, interpret and recall information in ways that confirm their pre-existing beliefs, while discounting or overlooking contradictory evidence.

Conspiracy theory

An explanatory belief that attributes the cause of an event, phenomenon or social condition to the secret actions of powerful groups acting in concert for malevolent purposes. Conspiracy theories are typically resistant to falsification (evidence against them does not disprove them, instead, it often gets absorbed into the theory), simplify complex realities into intentional plots, and often portray the world as controlled by hidden actors.

Content moderation

The process by which platforms, organizations and institutions monitor, review and manage user-generated content to ensure it complies with legal requirements, community standards and safety guidelines. Moderation can involve the removal, labelling or restriction of harmful, illegal or policy-violating material, as well as the promotion of accurate and trustworthy information.

Coordinated inauthentic behaviour

A term introduced by Meta (Facebook) to describe organized efforts where groups of accounts or pages work together to mislead people about who they are and what they are doing. It is not defined by the content of the messages but by the deceptive, coordinated and concealed nature of the actors and their activities. It often involves fake accounts, covert networks or hidden sponsorship, designed to manipulate public debate, amplify narratives or influence political outcomes.

Cybercriminals

Individuals or organized groups who use computers, networks, or digital technologies to commit crimes. Their activities include unauthorized access to systems, theft of data, identity fraud, financial scams, ransomware attacks and the distribution of malicious software. Cybercriminals may act independently, as part of loosely connected online networks, or within highly organized transnational criminal enterprises.

Debunking

A reactive strategy that addresses misinformation after it has begun to circulate. Its effectiveness depends on being timely, clear and delivered by trusted messengers. Effective debunking not only identifies a claim as false but also explains why it is false. Communication techniques such as the 'truth sandwich' – which begins and ends with accurate information while addressing the falsehood in the middle – help to minimize the risk of reinforcing the myth.

Decontextualization

The practice of presenting accurate information, images or quotes outside their original context in a way that alters their meaning or implications. By stripping away crucial details such as time, place, source or intent, decontextualization can mislead audiences, distort events and reinforce false narratives, even without fabricating new content.

Deepfake

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

Demystification

The process of unpacking how misinformation operates. It entails explaining its forms and functions, as well as the psychological, social and technological triggers that drive its spread. Unlike fact-checking, which focuses on verifying the truth or falsehood of individual claims, demystification highlights the structural, emotional, algorithmic and contextual factors that influence the information people encounter, thereby fostering a deeper understanding of why misinformation is persuasive and persistent.

Disinformation

False information that is deliberately created or spread with the intention to deceive or cause harm.

Doxing (or 'doxxing')

The intentional gathering and publishing of someone's private or personally identifiable information without their consent, typically carried out to shame, embarrass, harass, intimidate, threaten or cause harm.

Fact-checking

The process of verifying the factual accuracy of information, claims or statements, usually in journalism, public communication or online content. It involves systematically evaluating evidence, consulting credible sources

and providing transparent corrections when false or misleading claims are identified. Fact-checking can be proactive (anticipating claims) or reactive (debunking after claims spread).

False dichotomy

A logical fallacy that presents a situation as having only two opposing options or outcomes, when in reality there are additional possibilities. By reducing complex issues to a simplistic 'either/or' choice, false dichotomies distort reasoning, polarize debate and limit consideration of alternative perspectives. For example: "You're either with us or against us."

Farm

A 'farm' in the misinformation ecosystem refers to an organized operation that systematically produces, amplifies or manipulates content at scale. **Content farms** typically pursue profit by generating click-driven misinformation, while **troll farms** are politically or strategically motivated, aiming to distort public debate and influence opinion.

Filter bubble

A state of intellectual or informational isolation created by personalized algorithms that curate online content based on a user's past behaviour, interests and preferences. Within a filter bubble, individuals are primarily exposed to information and viewpoints that reinforce their existing beliefs, while alternative perspectives are downplayed or excluded. This phenomenon can limit diversity of information, strengthen confirmation bias and contribute to polarization.

Foreign information manipulation and interference

A pattern of behaviour that threatens or has the potential to negatively impact values, procedures and political processes. Such activity is manipulative in character, conducted in an intentional and coordinated manner, by state or non-state actors, including their proxies in and outside their own territory. (Also known as FIMI.)

Generative AI

A class of AI systems that can create new content, such as text, images, audio, video or code, based on patterns learned from large datasets. Unlike traditional AI systems designed primarily for classification or prediction, generative AI models are capable of producing novel outputs that resemble human-created work, often through deep learning architectures such as generative adversarial networks or large language models.

Hacktivists

A hacktivist is an individual or collective that uses hacking techniques as a form of political or social activism. Hacktivism combines 'hacking' and 'activism' to pursue goals such as promoting free information, protesting government or corporate actions, or drawing attention to social causes. Tactics may include website defacement, denial-of-service attacks, data leaks and hijacking online accounts.

Harmful information

Information that has the potential to cause, contribute to or result in harm to an individual or entity.

Hashtag hijacking

A reactive or strategic misuse of a trending or branded social media hashtag. It occurs when users adopt a hashtag intended for a specific topic or campaign and repurpose it to promote a substantially different or conflicting message – often involving spam, trolling, political agendas or conspiracies. This diversion can disrupt the original conversation and obscure or distort the intended content.

Hate speech

Any kind of communication in speech, writing or behaviour that attacks or uses pejorative or discriminatory language with reference to a person or group based on their identity, such as religion, ethnicity, nationality, race, colour, descent, gender or other factor.

Illusory truth effect or familiarity effect

A cognitive bias in which repeated exposure to a statement increases the likelihood that it will be perceived as true, regardless of its factual accuracy. This effect arises because familiarity enhances processing fluency – the ease with which information is mentally processed – which people often mistake for accuracy.

Infodemic

An overabundance of information, including accurate, false and misleading content, circulating in digital and physical environments during a crisis, which makes it difficult for people to find trustworthy sources and reliable guidance.

Infodemic management

The systematic response to these challenges – using tools like social listening, debunking and prebunking to manage the information environment and support public health decision-making.

Information integrity

Refers to the reliability, accuracy and trustworthiness of information within an ecosystem. It encompasses efforts to safeguard information environments from manipulation, misinformation, disinformation and other harmful content, ensuring that people can access factual, contextual and credible information to make informed decisions.

Information operations

Coordinated activities that use information, communication and influence tactics to affect the perceptions, decisions and behaviours of target audiences in support of political, military or organizational objectives. They often combine psychological operations, cyber activities, propaganda and disinformation campaigns to achieve strategic effects, both online and offline.

Inoculation theory

This proposes that people can build resistance to persuasion in a manner similar to medical inoculation: by being exposed to a weakened form of an argument, combined with refutations, they develop 'mental antibodies' that help them resist stronger attacks later. In the context of harmful information, inoculation theory underpins prebunking strategies – educational interventions that expose people to examples of manipulation techniques (such as scapegoating, conspiracy framing or emotional appeals) before they encounter them in real life. Research shows that inoculation can improve people's ability to detect and reject misinformation in public health, humanitarian and political contexts.

Lone wolf

An individual who plans and carries out violent or disruptive actions independently, without direct operational support from an organization, even if they are ideologically inspired by one. In the information domain, the term also extends to individuals who act alone in spreading extremist or conspiratorial content online, amplifying harmful narratives without belonging to a coordinated network.

Malinformation

Genuine information shared with the intent to cause harm – for example, by taking something true out of context or using it to discredit someone.

Misinformation, disinformation and hate speech

An umbrella term that refers to different types of false, misleading or damaging information. It includes misinformation (false information shared without

intent to harm), disinformation (false information deliberately created or spread to deceive or cause harm) and hate speech. (Also known as MDH.)

Misinformation

False or inaccurate information that is shared without intent to deceive.

Open-source intelligence

The collection, analysis and use of information that is publicly available and legally accessible. Sources include traditional media, academic publications, government reports, social media, satellite imagery and other open data. Open-source intelligence is widely used by governments, law enforcement, journalists, researchers and civil society to support decision-making, investigations and situational awareness.

Political manipulation

The deliberate use of deceptive, coercive or manipulative tactics by political actors to shape public perceptions, attitudes and behaviours in ways that serve their own interests, often at the expense of informed democratic decision-making. It can involve controlling narratives, spreading misinformation, exploiting emotional appeals, suppressing dissent or engineering consent through covert influence operations.

Prebunking

A proactive method aimed at building resilience to misinformation by exposing people – in advance – to weakened examples of misleading strategies or false claims, coupled with refutations. By leveraging inoculation theory, it prepares individuals to better recognize and resist manipulative content when they encounter it later. There are two types of prebunking: **narrative inoculation** counters broader storylines or persuasive frames that might be used to mislead, while **tactical inoculation** focuses on techniques of manipulation rather than the storyline itself, e.g., teaching people how tactics like emotional manipulation, decontextualization of data or scapegoating can make them more resistant when they encounter them. Together, these build 'mental immunity' – the cognitive resilience that helps people identify, resist and discount misinformation before it takes root.

Pre-emption

Refers to proactive strategies that anticipate misleading claims, narratives or harmful information before they spread, with the aim of reducing their impact.

Propaganda

The systematic and deliberate use of communication – through symbols, messages or media – to shape perceptions, manipulate cognition and direct behaviour to achieve the objectives of its sponsor. It is typically selective in its presentation of facts and may appeal to emotions rather than rational analysis.

Proxy

An intermediary actor – such as an organization, media outlet or individual – used by a state or group to conduct activities indirectly on its behalf. In information operations, proxies amplify or disseminate narratives while concealing the true origin or sponsor of the content. Proxies may act knowingly (aligned actors) or unknowingly ('useful idiots'), providing plausible deniability for the real orchestrators.

Resilience

The ability of individuals, communities, organizations or countries exposed to disasters, crises and underlying vulnerabilities to anticipate, reduce the impact of, cope with and recover from the effects of shocks and stresses without compromising their long-term prospects.

Rumours

Unverified information that spreads quickly within communities. Rumours may later prove to be true, false or partially true, but they can still cause confusion or mistrust while circulating.

Satire-as-truth

Refers to situations where parody or satirical content is mistaken for genuine information. When satire circulates without its original humorous or ironic context, such as memes, parody news articles or comedy sketches, audiences may interpret it literally and spread it as fact.

Scams and fraud (related to aid)

Deliberate acts of deception aimed at gaining money, resources or sensitive information, often targeting vulnerable populations in crises. These can include fake aid offers, fraudulent fundraising or misrepresentation by individuals or organizations. Disinformation often amplifies these schemes, creating confusion, urgency or mistrust among communities and donors.

Social listening

The systematic process of monitoring, collecting and analysing conversations, trends and mentions across digital and social media platforms to gain insights into public perceptions, behaviours and narratives. Unlike basic social media monitoring, which tracks metrics (likes, shares, mentions), social listening emphasizes qualitative analysis of context, sentiment and emerging issues.

Sock puppet account

A false online identity created and controlled by a person or organization to deceive others.

Synthetic content

Refers to text, images, audio, video or other media that is generated, manipulated or modified using AI or algorithmic systems, rather than being directly created or recorded by humans.

Troll

An individual who deliberately disrupts online conversations, communities or platforms by posting inflammatory, off-topic or deceptive content with the intent to provoke, mislead or manipulate others. In misinformation contexts, trolls may operate independently or as part of organized campaigns ('**troll farms**'), amplifying divisive narratives or spreading false information for political, financial or ideological purposes.

Visual manipulation

Changing or misusing photos and videos – by editing, cropping or adding false captions – to trick people or make them believe something that is not true.

Watermarking

A technique used to embed a marker (visible or invisible) during digital content creation, such as text, images, audio, video or AI-generated media, to indicate its origin, authenticity or ownership.