



SYCHO UNNY

ADIDAS
THE STRIPES BRAND

CIVIL


Chapitre 1



**Crise, chaos
et confusion:
comprendre les
informations
préjudiciables
dans les contextes
humanitaires**



Chapitre 1

A photograph showing a dirt road in a mountainous area. In the foreground, the shoulder of a person wearing a dark jacket is visible. The road is dusty and has tire tracks. To the left, there are utility poles with wires and a transformer. The background shows a steep, rocky hillside.

**Crise, chaos
et confusion:
comprendre les
informations
préjudiciables
dans les contextes
humanitaires**

Sommaire

	Introduction	31
1.1	Définition des informations préjudiciables	35
1.2	Évolution de l'écosystème de l'information	35
1.3	Qui sont les acteurs malveillants ?	39
1.4	Intégrité de l'information dans les situations de crise	40
1.5	De la diffusion à la communication bidirectionnelle	43
1.6	Quelles sont les personnes les plus vulnérables face aux informations préjudiciables et pourquoi ?	46
1.7	Quels sont les effets des informations préjudiciables ?	49
1.8	Intelligence artificielle et informations préjudiciables	53
1.9	Les discours façonnent les perceptions, la confiance et l'action	57
1.10	Contrôle de l'accessibilité et modération des contenus	59
1.11	Analyse et surveillance des informations préjudiciables	61
	Conclusions : les informations préjudiciables ne sont pas un bruit de fond	65
	Notes de fin	69

Introduction

Pour le secteur humanitaire, le tsunami survenu en 2004 dans l'océan Indien a été la première catastrophe d'envergure à faire l'objet d'une large couverture numérique, tandis que les guerres en Afghanistan et en Irak (au début des années 2000) ont constitué les premiers conflits armés analysés en temps réel par des milliers de commentateurs en ligne. Ces événements ont ainsi marqué l'avènement d'une ère de décentralisation de l'information numérique, davantage façonnée par des espaces en ligne rapides et participatifs que par les médias traditionnels. Les premiers contenus numériques reflétaient rarement les réalités locales, en particulier lorsque les plateformes ne permettaient pas de publier en langues locales. Cependant, la communauté émergente de blogueuses et blogueurs a contribué activement à la couverture médiatique, à la vérification des contenus, à la dénonciation des images manipulées¹, des reportages fabriqués de toutes pièces² ou des dommages exagérés. Bien qu'imparfaits³, ces efforts ont mis au jour de nouvelles possibilités en matière de vérification et de responsabilité publique.

Ces communautés de blogueuses et blogueurs en ligne sont alors devenues de puissants acteurs au sein de l'écosystème de l'information. Les innovations technologiques et certains changements sociaux ont donné aux individus la possibilité de partager leurs points de vue sans filtre, souvent en réponse à un mécontentement croissant à l'égard des médias traditionnels. Les blogs permettaient de diffuser des informations immédiates ainsi que des discours parallèles plus personnels, ce qui a favorisé la création de réseaux de discussion informels (des « communautés ») et des premières formes de médias participatifs. En fin de compte, ils ont jeté les bases des modèles de communication décentralisés et axés sur l'utilisateur que nous connaissons aujourd'hui.

Pour les professionnels des opérations et de la communication du secteur humanitaire, cette période a été intense et imprévisible, et laissait présager des changements autrefois inconcevables. Les crises humanitaires, en particulier les guerres, sont toujours allées de pair avec une utilisation de l'information à des fins de tromperie, de manipulation, de discrédit, de perturbation ou de promotion de programmes politiques ou idéologiques⁴. Ce qui a changé, c'est la nature, la vitesse, l'ampleur et l'accessibilité des informations préjudiciables. Aujourd'hui, un large éventail d'acteurs (des professionnels et des amateurs, anonymement ou ouvertement) créent et diffusent des informations préjudiciables sur des canaux en ligne et hors ligne. Internet, les téléphones mobiles, les plateformes de médias sociaux et les applications de messagerie constituent de puissants outils de diffusion. Dans de nombreux espaces, les faits, la vérité et l'exactitude ont laissé la place aux opinions, aux émotions et aux perceptions, la plupart du temps motivées par des considérations financières, politiques ou idéologiques. Souvent, la polarisation et la méfiance ne sont pas seulement des effets secondaires, mais de véritables objectifs. Cette dynamique dépasse largement le cadre de la guerre : les informations préjudiciables influencent désormais la façon dont les communautés comprennent les catastrophes, les situations d'urgence sanitaire, les migrations et d'autres défis humanitaires, mais aussi la manière dont elles s'y préparent et y font face.

Au lieu de faciliter l'accès à des informations fiables et vitales, l'espace informationnel alimente de plus en plus les craintes et les incertitudes des populations. Les informations relatives aux crises sont régulièrement manipulées, politisées ou déformées, notamment en minimisant les causes ou la gravité d'une catastrophe. Par exemple, en 2024, de fausses informations diffusées en ligne affirmaient que l'ouragan Hélène aux États-Unis avait été provoqué par des mesures de géoingénierie⁵. De telles théories ont

largement circulé sur les médias sociaux, ce qui illustre une tendance croissante à la désinformation et à la diffusion de mésinformations concernant les catastrophes ainsi qu'une méfiance accrue à l'égard des informations officielles.

Si le tsunami survenu dans l'océan Indien a été la première catastrophe observée à grande échelle grâce aux technologies numériques, le Covid-19 a été la première à être vécue numériquement. Tous les aspects de la crise (les directives sanitaires, les rumeurs, les craintes et les manifestations de solidarité) se sont propagés en ligne aussi rapidement que le virus lui-même. Le Covid-19, une urgence sanitaire sans précédent, a constitué à la fois une pandémie et une **infodémie**⁶ dont les effets ont touché les populations du monde entier. Si la crise sanitaire a été globalement maîtrisée, l'infodémie a quant à elle entraîné des conséquences durables, à savoir une plus grande méfiance à l'égard des autorités, une réticence et un refus accrus face à la vaccination et l'essor d'une industrie d'influenceurs et de complotistes s'exprimant sur toutes sortes de dangers, qu'ils concernent les traitements médicaux, les professionnels de santé ou encore les antennes-relais de téléphonie mobile⁷.



Infodémie

Le terme « **infodémie** » désigne une surabondance d'informations, qu'elles soient fausses ou trompeuses, circulant dans les environnements numériques et physiques pendant une épidémie.

Ces informations ont des effets immédiats et souvent néfastes, et il s'avère extrêmement difficile de les réfuter. Elles exploitent les divisions sociales, les sensibilités politiques et les préjugés profondément enracinés et incitent ainsi de plus en plus d'individus à s'opposer aux autorités, à refuser l'aide, à déshumaniser les personnes perçues comme « différentes » et à se mobiliser contre elles.



Quelle est ma réaction lorsque je reçois de fausses informations? D'une certaine manière, cela me fait peur. J'ai du mal à déterminer lesquelles sont crédibles et fiables, car elles sont toutes diffusées par des médias. On me dit comment me comporter, mais on me dit aussi que si j'agis de cette manière, il y aura des conséquences. Donc je ne sais pas quoi faire. C'est vraiment très compliqué et déroutant.»

Un membre de la communauté, Ouganda

En 2024, lors des interventions déployées face aux inondations qui ont frappé la région de Valence en Espagne, des mésinformations ont été diffusées selon lesquelles l'aide était détournée au profit des migrants plutôt que d'être acheminée vers les communautés touchées par la catastrophe, ce qui a alimenté les discours xénophobes. La même année, de fausses rumeurs et des contenus incendiaires ont circulé après que plusieurs enfants ont été tués et blessés lors d'un atelier de danse à Southport (Royaume-Uni), suggérant que l'agresseur était un demandeur d'asile musulman. Ce discours a attisé les propos anti-immigrés et islamophobes et a ainsi provoqué des manifestations violentes et des émeutes devant une mosquée, mais aussi des troubles de plus grande ampleur dans de nombreuses villes.

Perspectives des contributeurs 1.1



Diffusion d'informations préjudiciables lors des inondations liées à une dépression isolée à des niveaux élevés à Valence (Espagne) en 2024 (Partie 1 sur 2)

Fin octobre 2024, la région de Valence a été frappée par l'une des inondations les plus meurtrières qu'ait connues l'Espagne au cours des dernières décennies, causée par une dépression isolée à des niveaux élevés, également appelée «goutte froide». Des pluies torrentielles qui ont duré plusieurs heures ont submergé les infrastructures. Au total, 236 personnes ont perdu la vie et des communautés entières ont été dévastées.

La diffusion d'informations préjudiciables sur les médias sociaux a donné lieu à divers incidents, notamment des commentaires négatifs, des insultes et menaces proférés à l'encontre de nos employés et volontaires dans les rues et, dans une moindre mesure, des actes de vandalisme ciblant nos bureaux et véhicules (tels que des graffitis et des pneus crevés).

Des mésinformations circulent depuis plusieurs années sur la Croix-Rouge espagnole, mais au cours de cet événement météorologique, celles-ci ont principalement revêtu la forme de mensonges et de rumeurs affirmant que la Société nationale :

- n'était pas présente sur le terrain pour aider les habitants de Valence ;
- se contentait de prendre des photos et des vidéos mises en scène, plutôt que de fournir une aide réelle («vos vêtements sont propres, vos outils de travail sont neufs et ont été achetés récemment» [voir la figure 1.1]);
- aidait les migrants plutôt que les citoyens espagnols touchés par la catastrophe ;
- détournait les fonds, en mentant sur leur destination et les montants alloués ou en refusant les dons en nature afin de garder l'argent.



Fig 1.1

Messages publiés sur X en novembre et décembre 2024

Un plus petit nombre de messages récurrents critiquaient les actions menées par la Croix-Rouge espagnole sur la scène internationale (par exemple en faveur de l'Ukraine ou

de Gaza) ou faisaient référence à d'anciennes publications sur les migrants en Espagne ou les dons de sang.

Dans ce cas, les informations préjudiciables n'ont ni directement entravé l'action humanitaire ni empêché la fourniture de l'aide. Cependant, elles ont sans aucun doute eu des répercussions sur les personnes contribuant aux interventions à plusieurs égards :

- en créant une charge de travail considérable pour démentir ces informations, fournir des explications ou atténuer leurs conséquences ;
- en suscitant une détresse émotionnelle et même des doutes sur l'organisation, tant au sein de la société que parmi les membres de la Croix-Rouge espagnole ;
- en sapant la confiance de la population (certains donateurs ont remis en question l'intégrité de l'organisation).

Au cours de cet évènement météorologique, nos employés et volontaires (sur le terrain et en dehors), nos partenaires et alliés et, bien entendu, nos bénéficiaires, ont constitué les groupes les plus vulnérables face aux mésinformations. Les migrants ont également été la cible de certaines informations préjudiciables (ces publications ont inévitablement contribué à alimenter la xénophobie), mais l'objectif était plutôt de porter atteinte à la Croix-Rouge espagnole. Ces groupes ont été plus ou moins touchés sur le plan émotionnel et certains ont aussi perdu confiance en l'organisation.

María Trénor Alvargonzález
Directrice de la communication
Croix-Rouge espagnole

Kenan Terzic
Responsable des médias sociaux
Croix-Rouge espagnole

Souvent, les effets néfastes des informations préjudiciables touchent plus particulièrement les communautés, même si, dans certains contextes, ils sont exacerbés par des acteurs externes. Ce phénomène coïncide avec la perte de confiance dans les institutions traditionnelles et l'idée de plus en plus répandue selon laquelle les organisations non gouvernementales (ONG) défendent certaines causes ou certains intérêts politiques. Cette érosion de la confiance est encore accentuée par ce qui a été qualifié de « mort de l'expertise⁸ », dans la mesure où les journalistes, les universitaires et les institutions publiques (autrefois considérés comme des sources fiables d'information) ne bénéficient désormais que d'une confiance limitée dans de nombreux pays. En parallèle, l'intelligence artificielle (IA) est utilisée pour produire des contenus qui imitent les discours d'experts faisant autorité, mais peuvent s'avérer inexacts, trompeurs ou dépourvus de tout contexte.

1.1 Définition des informations préjudiciables

L'écosystème de l'information que nous connaissons aujourd'hui est extrêmement complexe et englobe de nombreuses formes de contenus préjudiciables, notamment la mésinformation, la désinformation, la malinformation, les discours de haine, la propagande et les activités de manipulation de l'information et d'ingérence étrangères. Ces mécanismes se recoupent et se renforcent souvent mutuellement, leur objectif étant le principal facteur qui les distingue.

Le présent rapport utilise le terme « informations préjudiciables » pour mettre en avant leur impact et les réponses qui y sont apportées, plutôt que d'employer des classifications rigides, qui ont souvent des connotations politiques et dépendent du contexte. Bien qu'il n'existe pas de définition universelle de ce terme, aux fins du présent rapport, il désigne les **informations susceptibles de porter atteinte, directement ou indirectement, à une personne ou à une entité, ou d'y contribuer.**

Les critiques légitimes, même lorsqu'elles s'avèrent dérangeantes ou difficiles à accepter, ne sont pas considérées comme préjudiciables. Toutefois, les organisations doivent en avoir connaissance et y répondre, car les critiques ignorées peuvent être exploitées par autrui.

Il est essentiel de bien comprendre la nature du préjudice, non seulement pour évaluer son incidence sur les individus, les communautés, les organisations et les sociétés, mais aussi élaborer des stratégies de réponse appropriées, proportionnées et fondées sur des principes.

1.2 Évolution de l'écosystème de l'information

Les informations préjudiciables sont souvent considérées comme un phénomène Internet, mais les crises humanitaires subissent de longue date l'influence des rumeurs, des mythes et de la propagande hors ligne, propagés par le bouche-à-oreille, des pamphlets, des émissions de radio et de télévision, des réunions communautaires et d'autres voies officielles ou semi-officielles. L'influence de ces modes de diffusion reste immense, en particulier dans les contextes où l'accès au numérique est limité ou où le niveau de maîtrise des médias est faible. Il est important de noter que les informations préjudiciables circulent souvent de manière fluide entre le virtuel et la vie réelle, ce qui amplifie leur portée et a des conséquences concrètes aux niveaux personnel et communautaire.

Perspectives des contributeurs 1.2



Impact des informations préjudiciables sur les interventions humanitaires au Soudan du Sud

Incident: À Djouba et dans les environs, certaines rumeurs circulaient selon lesquelles des ONG internationales distribuaient de la nourriture empoisonnée à des fins politiques.

Conséquences: Les bénéficiaires de l'aide ont hésité à venir récupérer les denrées alimentaires et des menaces verbales ont été proférées contre des membres du personnel local dans au moins deux cas documentés.

Réaction: L'équipe du projet Sentinel a utilisé la plateforme WikiRumours pour vérifier rapidement les informations et les réfuter publiquement, en collaboration avec les stations de radio locales et les dirigeants communautaires afin de rassurer les habitants et de rétablir la confiance dans les activités humanitaires de distribution alimentaire.



Incident: Des membres de la communauté ont déclaré avoir entendu dire que des jeunes armés se dirigeaient vers des villages du comté de Lainya, ce qui a incité des familles à fuir pendant la nuit.

Conséquences: Bien que la rumeur se soit révélée fausse, elle a provoqué des déplacements de population dus à la panique et entraîné la suspension temporaire des activités humanitaires locales.

Réaction: Des représentants de la communauté ont vérifié les faits sur le terrain et des messages rectificatifs ont été diffusés à la radio et par SMS le lendemain matin. Cette clarification rapide a permis de reprendre les activités humanitaires et d'empêcher de nouveaux déplacements ou conflits.

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

Christopher Tuckwood
Projet Sentinel

L'environnement de l'information numérique (qui englobe à la fois les domaines numérique et cybernétique) est de plus en plus complexe, polémique et politisé. Les États l'abordent sous différents angles: certains d'entre eux se concentrent sur les **infrastructures** (c'est-à-dire le **matériel**, l'architecture physique qui permet la circulation des données, comme les centres de données, les serveurs cloud et les câbles sous-marins), tandis que d'autres mettent l'accent sur le **contenu** (les propos, les données et les discours qui circulent sur les différentes plateformes et entre elles). Ces deux perspectives (le matériel et le contenu) ne s'excluent pas mutuellement, mais déterminent la manière dont une certaine influence et un certain contrôle s'exercent sur

l'espace informationnel, en entraînant des répercussions en matière de souveraineté, de surveillance et de résilience.

Contrôler le contenu peut consister à façonner, restreindre ou déformer les flux d'information à des fins politiques, idéologiques ou de sécurité, notamment au moyen de campagnes de désinformation, de la censure, de la réglementation des plateformes et d'activités de manipulation de l'information ou d'influence. Par conséquent, les frontières entre la cybersécurité, les opérations visant à influencer l'information et la concurrence géopolitique sont de plus en plus floues. Les décisions stratégiques relatives au matériel et au contenu ont des conséquences directes sur l'accès à l'information, la liberté d'expression et la capacité des communautés d'utiliser de manière sûre et efficace l'espace numérique.

En parallèle, les tactiques, méthodes et procédures des acteurs malveillants ne visent plus seulement les systèmes techniques. Elles exploitent de plus en plus les informations préjudiciables à des fins financières, politiques ou idéologiques, ce qui fait de la manipulation de l'information un phénomène aussi dangereux que les cyberattaques ciblant les systèmes physiques. Différents facteurs psychologiques sous-tendent cette menace : les informations préjudiciables exploitent les peurs, les émotions, les identités et les griefs, ce qui aggrave les divisions sociales et complique considérablement la riposte fondée uniquement sur des faits.

Perspectives des contributeurs 1.3

Premiers intervenants dans l'écosystème de l'information : renforcer la résilience face aux informations dans les interventions humanitaires

À une époque où les crises sont exacerbées par des informations trompeuses et manipulées, le secteur humanitaire doit considérer l'environnement informationnel comme un terrain opérationnel. Les fausses informations ne circulent plus en marge : elles influencent désormais les perceptions et les comportements, et déterminent la légitimité même de l'aide. Qu'il s'agisse d'itinéraires d'évacuation manipulés ou d'attaques utilisant l'information pour porter atteinte aux acteurs humanitaires, les contenus préjudiciables ne sont plus un effet secondaire des crises mais en font partie intégrante.

Au cours des premières semaines de la pandémie de Covid-19, les faux remèdes, les théories du complot et les discours xénophobes ont fortement concurrencé les directives de santé publique vérifiées, et les ont ainsi souvent dépassées, voire éclipsées, dans la course à l'attention. Dans les crises plus récentes, notamment au Soudan et en Ukraine, la diffusion de mésinformations a entravé l'évacuation en toute sécurité des civils, suscité la méfiance des populations locales à l'égard des travailleurs humanitaires et conduit à de violentes représailles. Ces exemples montrent que les informations préjudiciables ont

des répercussions concrètes qui sont immédiates et potentiellement mortelles. Attendre pour réagir, c'est prendre du retard dans la course.

Nous traversons actuellement une crise dans notre écosystème de l'information, où les discours déformés érodent la confiance, accentuent la polarisation et détériorent les relations entre les intervenants et les communautés. Selon le Baromètre de confiance Edelman 2025, près de 70 % des personnes interrogées dans le monde s'inquiètent d'être délibérément induites en erreur par des sources faisant autorité, telles que les gouvernements et les médias. Il est alarmant de constater que 25 % des personnes interrogées se disent prêtes à diffuser des informations erronées pour encourager le changement, un chiffre qui atteint 34 % chez les jeunes adultes.

Par ailleurs, des données récentes collectées par l'Institut Reuters indiquent que 58 % des personnes interrogées dans le monde craignent de ne pas pouvoir distinguer le vrai du faux en ligne. Dans de nombreuses régions, les influenceurs en ligne et les politiciens nationaux sont désormais considérés comme les principales sources d'information inexactes, devant les médias traditionnels, les puissances étrangères ou les acteurs malveillants.

Pour le secteur humanitaire, le message est clair : la lutte pour gagner la confiance des populations ne doit pas commencer après l'apparition d'une crise. Tout comme les chaînes d'approvisionnement sont prépositionnées, les personnes chargées de la communication et des protocoles de surveillance numérique ainsi que les messagers de confiance à l'échelle locale doivent être identifiés et déployés en amont, prêts à être mobilisés en cas de crise, mais surtout être présents en permanence sur le terrain.

Ces «premiers intervenants dans l'écosystème de l'information» jouent un rôle essentiel pour rivaliser dans les espaces numériques polémiques et protéger les populations contre les mensonges qui se propagent rapidement. Il convient donc de mobiliser des personnes fiables qui sont des porte-paroles crédibles au sein de communautés clés, d'instaurer des canaux multilingues et de former des équipes sur le terrain afin de repérer et contrer les discours préjudiciables. Il faut en outre mettre en place des mécanismes de surveillance des médias sociaux, des systèmes d'alerte en temps réel ainsi que des outils permettant de réfuter rapidement les mensonges. Si l'IA peut causer d'immenses préjudices en favorisant la production et la diffusion de fausses informations, elle offre également la possibilité de créer des contenus précis et attrayants.

Renforcer la résilience face à la manipulation de l'information dans un écosystème pollué ne se limite pas à établir la vérité. Cela implique de concevoir des systèmes de communication souples, dirigés par les communautés et fondés sur la confiance. Si le secteur humanitaire entend rester une force neutre et stabilisatrice, la confiance doit être considérée comme un outil essentiel à la réalisation de notre mission, qui nécessite des ressources et doit être défendu.

Oliver Hayes OBE

Responsable, Lutte contre la désinformation
en Europe, au Moyen-Orient et en Afrique

Agence Edelman

Dave Fleet

Responsable, Crises numériques et lutte
contre la désinformation au niveau mondial

Agence Edelman

D'autres facteurs qui contribuent à façonner l'écosystème de l'information sont abordés au [Chapitre 2, page 75](#) et au [Chapitre 3, page 147](#).

1.4 Intégrité de l'information dans les situations de crise

Les Nations Unies considèrent le renforcement de l'intégrité de l'information comme «l'une des tâches les plus urgentes de notre époque». Il est essentiel pour garantir le respect des droits fondamentaux, la paix et le développement durable. L'intégrité de l'information renvoie à l'exactitude, à la cohérence et à la fiabilité de l'information. Face aux menaces croissantes liées à la désinformation, à la désinformation, aux discours de haine et à l'utilisation abusive des technologies numériques, y compris de l'IA, les Nations Unies ont défini les Principes mondiaux pour l'intégrité de l'information⁹ suivants :

- Confiance et résilience de la société;
- Incitations saines;
- Responsabilisation du public;
- Médias indépendants, libres et pluralistes;
- Transparence et recherche.

Avant de définir ces principes, les Nations Unies ont publié une note d'orientation traitant de l'intégrité de l'information sur les plateformes numériques en 2023 afin d'appeler à la mise en œuvre de mesures de suivi, notamment des consultations et la mise au point d'un éventuel code de conduite, en vue de contribuer à «guider les États membres, les plateformes numériques et les autres parties prenantes dans les efforts qu'ils déploient pour rendre l'espace numérique plus inclusif et plus sûr pour tous et toutes, tout en défendant vigoureusement le droit à la liberté d'opinion et d'expression ainsi que le droit d'accès à l'information¹⁰». Les consultations ont abouti à l'élaboration et à l'adoption de ces principes et, en septembre 2025, les Nations Unies ont publié le premier numéro d'une nouvelle série de notes d'information intitulée *From Principles to Practice: Strengthening Information Integrity*¹¹ (Des principes à la pratique : renforcer l'intégrité de l'information).

Perspectives des contributeurs 1.4

S'appuyer sur les normes des Nations Unies pour renforcer l'intégrité de l'information

Avant d'élaborer un plan de lutte contre les informations préjudiciables, nous devons comprendre les cadres mondiaux qui guident nos actions. Toutes les stratégies doivent se fonder sur les trois principaux cadres élaborés par les Nations Unies, à savoir :

- 1 Les Principes mondiaux des Nations Unies pour l'intégrité de l'information,** qui contiennent des recommandations axées sur cinq principes : la confiance de la société, des incitations saines, la responsabilisation du public, un espace médiatique libre, la transparence et la recherche. Ces principes constituent un plan d'action favorisant la collaboration multipartite.
- 2 Les Principes pour la gouvernance des plateformes numériques de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO)** qui définissent cinq responsabilités, notamment la réalisation d'évaluations d'impact et l'accès à des recours efficaces.
- 3 Le Pacte numérique mondial,** adopté en septembre 2024, dans lequel les gouvernements se sont engagés à œuvrer en faveur d'un avenir numérique ouvert, sûr et fondé sur les droits, et qui contient des mesures spécifiques visant à protéger l'intégrité de l'information.

Ces cadres ne sont pas des contraintes, mais les fondements d'une action innovante et multidimensionnelle.

Principes directeurs et nécessité d'adopter une approche multipartite :

- Les Principes mondiaux des Nations Unies appellent l'ensemble des acteurs (gouvernements, entreprises technologiques, publicitaires, médias, etc.) à renforcer l'intégrité de l'information.
- Ils exhortent les Nations Unies à accroître les efforts déployés, ce que nous faisons en favorisant la coordination à l'échelle du système, le partage de connaissances et la collaboration dans le cadre d'un groupe de travail sur l'intégrité de l'information.
- Aucun acteur ne peut réussir seul ; il est essentiel de nouer des partenariats pour renforcer la résilience mondiale.

Quatre pistes d'action concrètes fondées sur les enseignements tirés :

- 1 Coalitions et actions multipartites :** mettre en place une instance permanente réunissant des acteurs de confiance, notamment des gouvernements et des groupes de jeunes, afin de partager les informations, de signaler les risques et de coordonner des interventions rapides, en particulier pendant les élections ou les crises de santé publique.
- 2 Renforcement des capacités et de la résilience :** doter les journalistes, les éducateurs, les personnes chargées de vérifier l'information et les dirigeants communautaires des compétences, des ressources et des fonds nécessaires pour repérer et contrer les discours préjudiciables, en accordant une attention particulière aux zones défavorisées.
- 3 Recherche et évaluation des risques :** créer un centre de recherche régional pour surveiller les discours, publier des notes d'information accessibles à tous et diffuser rapidement des alertes au-delà des frontières.
- 4 Protocoles de prévention, d'atténuation et d'intervention :** élaborer des procédures claires et fondées sur les droits fondamentaux permettant aux

gouvernements, aux médias, aux plateformes et aux acteurs de la société civile d'agir rapidement et de manière transparente lorsque des contenus préjudiciables font leur apparition.

Hiroyuki Saito

Directeur

Centre d'information des Nations Unies, Dakar

Perspectives des contributeurs 1.5



La qualité des données, un élément fondamental dans la gestion des catastrophes

Même lorsqu'elles ne sont pas créées et diffusées délibérément, les fausses informations peuvent compromettre l'efficacité des services humanitaires et de la gestion des catastrophes. Lors du séisme qui a frappé la ville de Bam (Iran) en 2003, faisant des milliers de victimes et touchant plus de 250 000 personnes, une opération de secours d'envergure a été déployée aux niveaux national et international. Cette opération a bénéficié d'une attention et d'un soutien considérables à l'échelle nationale et internationale. Si elle a été couronnée de succès, certaines difficultés ont tout de même été rencontrées, notamment en matière de gestion de l'information.

La diffusion d'informations inexactes et exagérées sur l'ampleur ainsi que sur le montant des contributions financières et de l'aide humanitaire internationale a engendré trois principaux problèmes. Premièrement, les informations inexactes ont perturbé la planification des interventions, leur mise en œuvre et l'établissement de rapports. Deuxièmement, les déclarations exagérées concernant le montant de l'aide internationale ont fait douter les donateurs potentiels quant à la nécessité de fournir une aide et des contributions supplémentaires, ce qui a compromis la mobilisation des ressources et les efforts de collecte de fonds à l'échelle mondiale. Troisièmement, les chiffres gonflés relatifs au montant et à la valeur de l'aide internationale ont accru les attentes des personnes touchées et des bénéficiaires, entraînant ainsi des difficultés de gestion. Ces expériences soulignent la nécessité absolue pour toutes les parties prenantes aux interventions en cas de catastrophe et aux services humanitaires de respecter des normes minimales de qualité en matière de données et d'informations (à savoir **l'exactitude, la rapidité, la pertinence et la fiabilité**) avant de les partager et de les diffuser.

Outre l'exactitude et la qualité, les informations doivent être **exhaustives et intégrées** afin de pouvoir gérer les catastrophes et réduire les risques de manière efficace. Des données et des informations incomplètes peuvent entraver la réalisation des résultats escomptés. Par le passé, les acteurs chargés de la gestion des catastrophes considéraient les **aléas** naturels comme le principal facteur à l'origine des catastrophes. Plus tard, nous avons compris que ces derniers n'étaient pas les seuls responsables des catastrophes. Ce sont plutôt les vulnérabilités humaines et sociétales, notamment **physiques, sociales, économiques et environnementales**, ainsi que l'exposition des personnes, des systèmes

et des biens aux différents risques qui en sont les principaux facteurs. Ces vulnérabilités découlent souvent de comportements humains et de politiques et pratiques inadéquates. Il est donc impératif d'intégrer les données sur les vulnérabilités physiques, socioéconomiques et environnementales aux informations sur les risques afin de mieux comprendre les **risques de catastrophe**. Cette approche est en adéquation avec la première priorité définie dans le Cadre de Sendai pour la réduction des risques de catastrophe (2015-2030), à savoir comprendre les risques de catastrophe.

Mostafa Mohaghegh

Coordonnateur principal du Centre Asie-Pacifique pour le développement de la gestion de l'information sur les catastrophes (CAPGIC)

Commission économique et sociale des Nations Unies pour l'Asie et le Pacifique (CESAP)

1.5 De la diffusion à la communication bidirectionnelle

La promesse de disposer de nouveaux outils favorisant la mobilisation des communautés devait marquer un changement profond vers une communication plus inclusive. Les plateformes numériques ont permis de passer de la diffusion unidirectionnelle de l'information (communiqués de presse, sites Internet, déclarations officielles) à une communication bidirectionnelle. Toutefois, dans la pratique, ce changement très appréciable apporte son lot de difficultés. N'importe qui peut désormais répondre, contester, réinterpréter ou amplifier un message instantanément et publiquement.

Si la communication bidirectionnelle encourage le dialogue, elle ouvre également la porte à la critique, aux contenus préjudiciables et à la perturbation du discours, notamment de la part de publics internationaux et d'acteurs malveillants qui n'ont aucune connaissance des réalités du terrain. Les médias sociaux favorisent rarement les échanges constructifs, en particulier lorsque les réponses s'éloignent considérablement du message initial. Pour les organisations humanitaires, qui s'efforcent de rester neutres et impartiales, il est difficile de surveiller et de gérer cet environnement imprévisible et alimenté par les émotions, mais leur incapacité à le maîtriser peut avoir des conséquences directes sur les opérations.

En juin 2025, les plateformes Google (26 ans), Facebook (21 ans), YouTube (20 ans), Twitter/X (19 ans), VK (18 ans), WhatsApp (16 ans), Instagram (14 ans), Signal (12 ans), Telegram (11 ans), TikTok (9 ans en Chine, 8 ans dans le monde) étaient utilisées par des milliards de personnes dans le monde entier. Ces plateformes ont transformé la manière dont les individus accèdent aux informations et les partagent, interagissent et communiquent. Elles sont en outre devenues de puissants vecteurs de contenus préjudiciables. Leurs algorithmes privilégient les contenus qui suscitent l'intérêt du public, souvent les plus choquants ou les plus clivants, car ils génèrent davantage de clics, d'attention et, par conséquent, de profits. Il ne s'agit pas d'un dysfonctionnement, mais d'une caractéristique inhérente à ces systèmes qui renforcent les préjugés, créent des chambres d'écho, faussent les perceptions et accentuent les divisions.

La diffusion délibérée d'informations préjudiciables sape la confiance et remet en cause les intentions humanitaires, l'action fondée sur des principes et la légitimité.

Dans certains cas, elle a directement conduit à des menaces et à des actes de violence : des réfugiés et des migrants ont été menacés, des travailleurs humanitaires ont été attaqués et des volontaires se sont heurtés à des comportements hostiles.



Lors de l'épidémie d'Ebola qui a frappé la Guinée en juillet 2014, le président de la Croix-Rouge guinéenne a signalé que les violences contre les volontaires entravaient l'accès aux communautés touchées et perturbaient les inhumations sans risque. Ces attaques résultaient d'idées fausses et de craintes liées à la maladie, malgré les efforts déployés pour contrer les rumeurs à la radio, à la télévision et au moyen d'activités de mobilisation communautaire. En collaboration avec les chefs religieux, les familles et les parties prenantes locales, la Société nationale a mis en place des pratiques funéraires respectueuses et adaptées aux normes culturelles afin de permettre aux populations d'honorer leurs morts¹².

Les utilisateurs des médias sociaux attendent de plus en plus des réponses rapides et authentiques. Les organisations sont donc contraintes de communiquer des informations en temps réel, ce qui est souvent incompatible avec les processus d'approbation centralisés, les besoins de coordination avec les équipes sur le terrain et les exigences en matière d'exactitude. Par exemple, lors d'incidents relatifs à la sécurité ou à l'intégrité, la nécessité d'une communication rapide et transparente peut être en contradiction avec l'obligation de vérifier les faits, de faire preuve de diligence raisonnable et de garantir le respect des procédures. Les messages tardifs ou trop prudents peuvent alors être perçus comme une manière de protéger en priorité la réputation de l'organisation ou les relations financières.

Dans un environnement où les informations préjudiciables circulent plus rapidement que les faits et où la crédibilité des institutions est constamment examinée, même les communications bien intentionnées peuvent être perçues comme réactives, défensives, hypocrites ou manquant d'empathie. Pour que la communication soit véritablement bidirectionnelle, il ne suffit pas d'être réactif, il faut également afficher son intention, faire preuve de transparence et se montrer présent. Cependant, lorsque des acteurs malveillants exploitent les informations communiquées par une organisation, par exemple en utilisant les sections de commentaires pour semer le trouble, détourner les messages ou la discréditer, le dialogue bidirectionnel peut ne pas suffire à rétablir la vérité. De nombreuses organisations évitent de répondre par crainte d'une escalade, mais le silence peut laisser les discours préjudiciables se répandre et saper leur crédibilité. Les sections de commentaires, autrefois considérées comme des espaces d'échange, sont devenues de plus en plus difficiles à modérer. Le volume considérable de contenus préjudiciables qu'elles contiennent, conjugué à des moyens limités pour les gérer, a conduit certaines agences de presse et certains médias¹³ à les désactiver afin de réduire les risques d'atteinte à leur réputation et de freiner la propagation de ces informations.

Les mesures mises en œuvre suite au séisme qui a frappé Haïti en 2010 sont un bon exemple. À l'époque, la Fédération internationale avait mis en place un programme de communication avec les bénéficiaires dans le cadre duquel dix points d'accès avaient été créés afin de permettre aux personnes et aux communautés de faire part de leurs commentaires concernant l'aide humanitaire dont elles bénéficiaient et d'exprimer leurs besoins. Ces points d'accès comprenaient des émissions de radio en direct, des

SMS, des centres d'information, des échanges directs avec les communautés et des centres d'appels. Cette initiative a permis à la Fédération internationale d'adapter en permanence ses programmes et de prendre des décisions plus éclairées en s'appuyant sur les contributions des communautés.

Il est de plus en plus difficile de savoir quand et comment communiquer. Les retards ou le silence créent souvent un vide que d'autres personnes, qu'elles soient mal intentionnées ou mal informées, s'empressent de combler.

Perspectives des contributeurs 1.6



Un échec de communication et un moment décisif

Les troubles civils qui ont frappé le Bangladesh en juillet et août 2024 resteront dans les mémoires tant pour leurs conséquences politiques (à savoir la chute du gouvernement en place et son remplacement par un gouvernement intérimaire) que pour la manière dont ils ont mis à l'épreuve les institutions. Le Croissant-Rouge du Bangladesh n'a pas été épargné : cette organisation forte d'une longue tradition de service a soudainement été prise entre deux feux en se heurtant à la méfiance du public et à des défis internes.

Bien que la Société nationale soit intervenue (dans plusieurs districts, avec discrétion et courage), le discours dominant relayé par les médias et certains groupes de population était celui de l'absence, du silence, voire de la complicité. Les accusations d'inaction se sont rapidement multipliées et les rumeurs préjudiciables concernant un certain alignement politique ont éclipsé le travail accompli sur le terrain pour sauver des vies. À mesure que la situation se stabilisait et que le brouillard créé par les troubles se dissipait, nous avons compris qu'il ne s'agissait pas seulement d'un échec de communication, mais d'un véritable moment décisif. Alors que les volontaires du Croissant-Rouge du Bangladesh dispensaient les premiers secours aux manifestants, aux forces de l'ordre et aux personnes qui se trouvaient sur place, et que les ambulances transportaient des centaines de blessés vers des lieux sûrs, le discours général présentait une version différente des faits. Beaucoup ignoraient tout simplement les actions menées par la Société nationale et, à défaut d'une communication claire, rapide et fiable, cette méfiance s'est accentuée.

En réalité, le Croissant-Rouge du Bangladesh a mené une intervention humanitaire d'envergure. Vingt et une équipes ont été déployées dans neuf sections, des centaines de personnes ont reçu les premiers soins et plus de 2 000 familles ont bénéficié d'une aide alimentaire. Toutefois, la Société nationale a rapidement reconnu «[qu'elle n'avait] pas communiqué ces informations à temps, ni de la bonne manière. Et dans une crise où les perceptions influencent la réalité, ce silence a eu un coût».

Compte tenu du contexte instable et politiquement tendu, les dirigeants du Croissant-Rouge du Bangladesh ont pris la décision stratégique de faire profil bas (une discrétion intentionnelle) afin d'atténuer les risques de politisation, de protéger les volontaires et d'éviter toute implication dans des discours partisans. Si cette approche était compréhensible du point de vue de la gestion des risques, ce silence et ce manque de visibilité

publique ont créé un vide informationnel, rapidement comblé par des spéculations, des critiques et des discours politiques échappant au contrôle de l'organisation. Les conséquences ont dépassé la simple atteinte à la réputation. Dans certains districts, les volontaires ont été harcelés, les partenariats établis avec les ONG locales ont été mis à rude épreuve et les questions des donateurs et des partenaires du Mouvement ont soulevé des inquiétudes quant à la capacité de la Société nationale à intervenir dans des contextes politiquement sensibles tout en préservant la confiance de l'ensemble des parties prenantes. Pour se remettre d'une telle atteinte, il ne suffit pas de soigner son image. Il faut déployer des efforts à long terme afin de renforcer l'indépendance institutionnelle, diversifier les structures de gouvernance et regagner la confiance des populations.

Alberto Bocanegra

Chef de délégation pour le Bangladesh

Fédération internationale

1.6 Quelles sont les personnes les plus vulnérables face aux informations préjudiciables et pourquoi ?

Bien que la quantité d'informations dont nous disposons aujourd'hui n'ait jamais été aussi importante, un grand nombre de personnes continuent de vivre dans un vide informationnel où les informations vitales et susceptibles de sauver des vies sont inaccessibles ou peu fiables. Les informations préjudiciables peuvent avoir des répercussions négatives sur les individus, les communautés, les institutions et des sociétés entières, mais certains groupes sont exposés à des risques disproportionnés en raison de vulnérabilités structurelles, contextuelles ou situationnelles. Comme le soulignent les analystes spécialisés dans les politiques et la sécurité Singer et Brooking, «[c]omme toute infection virale, les offensives informationnelles ciblent les membres les plus vulnérables d'une population, en l'occurrence les moins informés¹⁴.»

Dans les contextes humanitaires, les personnes les plus vulnérables peuvent être :

- **Les populations touchées par une crise :** les personnes qui sont confrontées à des conflits armés, des catastrophes, des déplacements, des migrations ou des urgences sanitaires sont souvent exposées à des perturbations des canaux de communication, à une anxiété accrue et à des informations rares ou manipulées ;
- **Les groupes marginalisés ou socialement exclus :** les communautés victimes de discrimination fondée sur la race, l'origine ethnique, le genre, la religion, le handicap ou le statut juridique sont régulièrement la cible de discours préjudiciables et ont rarement accès à des sources d'information fiables ;
- **Les jeunes et les personnes âgées :** les jeunes peuvent être davantage exposés aux informations préjudiciables en raison du temps qu'ils passent

en ligne, de l'influence de leurs pairs ou des tendances sur Internet. Quant aux personnes âgées, elles peuvent avoir des difficultés à utiliser les technologies numériques et à naviguer dans les nouveaux environnements informationnels ;

- **Le personnel humanitaire, les volontaires et les organisations :** ces acteurs sont de plus en plus souvent la cible de discours remettant en question leur neutralité, leur indépendance, leur impartialité ou leur intégrité, ce qui sape la confiance des populations, compromet l'accès à ces dernières et nuit à leur sécurité ;
- **Les journalistes, les défenseurs des droits fondamentaux et les dirigeants de la société civile :** les personnes occupant ces fonctions sont souvent délibérément discréditées, réduites au silence ou menacées dans des environnements polarisés ;
- **Les personnes ayant un accès limité aux informations ou maîtrisant mal les outils médiatiques :** les personnes qui n'ont pas accès à des sources d'information fiables ou n'ont pas développé leur esprit critique sont plus susceptibles de croire et de diffuser de faux contenus.



Je ne pense pas que les personnes à faible revenu ou les personnes âgées sont forcément plus vulnérables. Ce sont plutôt celles qui sont plus facilement influencées par leurs émotions et les informations, ce qui n'a rien à voir avec l'âge. Certaines personnes ont des idées particulièrement arrêtées et refusent de changer d'avis. Certains dirigeants considèrent parfois que leur avis est meilleur que celui des autres et affirment qu'ils ont raison. Leur volonté, ou leurs descriptions subjectives, peut influencer la propagation de l'information, et ils y croient eux-mêmes. Par exemple, en tant que secouriste, même si je dis à une personne de 50 ans ce qu'il faut faire ou ne pas faire du haut de mes 20 ans, elle peut mettre mes propos en doute et me répondre : « J'ai mangé plus de sel que tu n'as mangé de riz, j'en sais plus que toi ». Ce n'est donc pas l'âge qui détermine la vulnérabilité aux fausses informations, mais les préjugés subjectifs. »

Membre de la communauté, Chine

Il est primordial d'identifier et de comprendre ces vulnérabilités ainsi que les risques spécifiques liés aux informations préjudiciables dans différents contextes afin d'atténuer leur impact sur les interventions humanitaires. Lors d'une enquête réalisée par la Fédération internationale auprès de 400 membres du personnel, **50 % ont indiqué avoir été confrontés à des situations dans lesquelles la diffusion de fausses informations ou de discours de haine a nuí à l'action humanitaire**, les principaux cas étant liés à la vaccination, à la santé et aux migrations.

Perspectives des contributeurs 1.7

Enquête auprès du personnel du Haut-Commissariat des Nations Unies pour les réfugiés

Les menaces qui portent atteinte à l'intégrité de l'information sur les plateformes numériques, telles que la désinformation, la désinformation et les discours de haine, sont de plus en plus graves et entraînent des préjudices concrets, en particulier dans les contextes humanitaires les plus difficiles comme les conflits armés, les catastrophes et autres situations d'urgence. Ces préjudices peuvent être physiques (lorsqu'il est question d'actes de violence ou de meurtres par exemple), économiques, sociaux et psychologiques, et érodent ainsi la confiance et les capacités de sécurité. Les acteurs humanitaires sont également exposés à des risques considérables: la sécurité du personnel, la neutralité opérationnelle, le soutien du public et les capacités de collecte de fonds peuvent être menacés. En fin de compte, les inégalités en matière d'information risquent d'empêcher les réfugiés d'accéder à des données vitales sur la protection via les plateformes numériques.



Afin de mieux comprendre l'impact de la désinformation, de la désinformation et des discours de haine, le Haut-Commissariat des Nations Unies pour les réfugiés (HCR) a réalisé deux enquêtes mondiales fin 2023, l'une auprès de communautés de personnes déplacées de force et apatrides, l'autre auprès du personnel de l'organisation. 85% des personnes déplacées de force et apatrides interrogées ont indiqué avoir été confrontées à des discours de haine ou de fausses informations ciblant leurs communautés, tandis que 72% d'entre elles ont déclaré avoir été personnellement visées. La plupart des membres du personnel du HCR ont affirmé que les discours de haine et les fausses informations avaient des répercussions sur les activités de l'organisation: 37% d'entre eux ont vu circuler ce type de contenu chaque mois, 60% ont constaté que les discours de haine avaient des effets directs, 20% ont été personnellement visés et un tiers ont déclaré que des collègues avaient été touchés. La quasi-totalité des membres du personnel (92%) ont exprimé leur inquiétude quant aux effets des discours de haine et des fausses informations sur les personnes bénéficiant de la protection du HCR. Sur une note plus positive, un membre du personnel sur trois a souligné que des groupes dirigés par des réfugiés ou des groupes communautaires s'efforçaient de lutter contre ces phénomènes. La collaboration avec ces acteurs s'est avérée essentielle pour apporter des réponses efficaces.

Gisella Lomax

Conseillère principale, Intégrité de l'information

HCR

1.7 Quels sont les effets des informations préjudiciables ?

Les informations préjudiciables menacent la vie, la sécurité et la dignité des personnes touchées par des crises humanitaires. Elles peuvent exacerber la violence pendant les conflits armés et les catastrophes, déformer la réalité sur le terrain, induire les populations en erreur en ce qui concerne la disponibilité de l'aide et influencer des décisions vitales, telles que celle de rester ou de fuir, d'accepter ou de refuser des soins médicaux.



La mésinformation est dangereuse, extrêmement dangereuse, et pas seulement à grande échelle. Elle peut avoir des répercussions au sein des familles et les détruire, semer la confusion et compromettre l'avenir des personnes. Elle peut bouleverser des vies et engendrer de la violence. Les conséquences peuvent être multiples. Selon moi, c'est un véritable fléau.»

Membre de la communauté, Ouganda

Le HCR a recensé un large éventail de préjudices hors ligne liés à la diffusion d'informations préjudiciables dans les situations humanitaires, notamment la xénophobie, le racisme, la persécution et la violence, les meurtres, les déplacements forcés, la traite et l'exploitation, les obstacles entravant l'accès aux droits et aux services, les atteintes à la réputation, l'érosion de la confiance et de la légitimité, l'affaiblissement de la capacité de protéger et soutenir les réfugiés, les menaces pour la sécurité physique des travailleurs humanitaires et la diminution du soutien apporté par les donateurs¹⁵.

L'Organisation mondiale de la Santé (OMS) a mis en lumière les dommages psychologiques et sociaux causés par les informations préjudiciables, tant au sein des communautés touchées par les conflits armés que parmi les intervenants humanitaires. Comme l'indique une étude de l'OMS, «[d]'après les données disponibles, les personnes ressentent une détresse mentale, sociale, politique et/ou économique en raison des contenus faux et trompeurs liés à la santé qui circulent sur les médias sociaux pendant les pandémies, les situations d'urgence sanitaire et les crises humanitaires¹⁶». Ce phénomène se produit lorsque des informations préjudiciables dissuadent les personnes de solliciter des services humanitaires ou d'y recourir, et compromettent la capacité des organisations de déployer des interventions efficaces.

Le Comité international de la Croix-Rouge (CICR) a attiré l'attention sur le fait que la diffusion de certaines formes d'informations préjudiciables peut constituer une violation du droit international. En effet, le droit international humanitaire impose «des limites importantes à la publication ou à la diffusion de certaines formes [d'informations préjudiciables]¹⁷», telles que celles qui incitent à violer le droit international humanitaire, à recourir à la violence pour semer la terreur parmi les civils, à nuire aux activités humanitaires ou médicales, et à diffuser des images de prisonniers de guerre.

Conformément au droit international des droits humains et à d'autres dispositions du droit international, tout discours de haine qui incite à l'hostilité, à la discrimination, à









la violence ou au génocide peut constituer une violation¹⁸. La Rapporteuse spéciale des Nations Unies pour la liberté d'expression a souligné que la désinformation orchestrée par l'État «a de profondes répercussions sur les droits humains, l'état de droit, les processus démocratiques, la souveraineté nationale et la stabilité géopolitique, ce qui s'explique par les ressources et les moyens des États ainsi que par leur capacité à réduire simultanément au silence toutes les voix indépendantes et critiques de sorte que rien ne puisse plus contester le discours officiel¹⁹».

La Fédération internationale regrette que certains donateurs se soient laissés influencer par des rumeurs au lieu de s'engager directement et officiellement auprès des Sociétés nationales de la Croix-Rouge et du Croissant-Rouge concernées. Des opinions se sont ainsi forgées en se fondant sur de fausses informations plutôt que sur des faits vérifiés. Si ces rumeurs ont probablement servi les intérêts de certaines personnes ou institutions, elles ont eu une incidence négative sur les communautés et ont empêché des Sociétés nationales de se développer et d'assurer des services humanitaires essentiels.

Les lois, les politiques et les plans constituent le fondement de toutes les activités de gestion des risques de catastrophe, et contribuent à la protection et à la préparation des communautés du monde entier. Il est donc primordial d'établir des cadres juridiques et politiques solides afin de favoriser l'intégrité de l'information et de relever les défis résultant de la diffusion d'informations préjudiciables dans ce domaine. Pour réagir efficacement, les acteurs humanitaires doivent comprendre comment ces informations perturbent les interventions. Il convient d'établir une typologie des différents préjudices en vue de constituer une base de données probantes qui appuiera les efforts visant à identifier, mesurer et atténuer ces effets. Chacun des types de préjudice ci-après peut compromettre considérablement l'action humanitaire et l'ensemble d'entre eux doivent être mieux appréhendés, surveillés et traités.



Table 1.1 Typologie des préjudices

Type de préjudice	Exemples de préjudice
 Physique	Blessures physiques, pertes en vies humaines, incitation à la violence, à la panique ou à des comportements dangereux, refus de soins médicaux. Attaques physiques contre des véhicules, des installations ou des bureaux humanitaires.
 Psychologique	Traumatisme émotionnel ou psychologique, peur, anxiété, désorientation, discrimination, honte, méfiance, manipulation, intimidation, harcèlement.
 Social	Perturbation de la cohésion sociale et de la confiance au sein des communautés, érosion des relations, stigmatisation, fragmentation sociale, divisions, polarisation. Déplacement forcé ou induit de populations entraînant la séparation des familles, la perte des liens communautaires et l'affaiblissement des systèmes de soutien collectifs.
 Sociétal	Réduction de l'espace dédié à l'action humanitaire, affaiblissement de la confiance dans les institutions, atteinte à la réputation, perturbation des opérations, détérioration des services, musellement de certains secteurs de la société, érosion des droits, restriction de l'accès à l'information, exclusion, limitation de la liberté d'expression. Utilisation d'informations préjudiciables pour justifier des poursuites judiciaires, la pénalisation de certains propos ou des abus des systèmes judiciaires.
 Informationnel	Déformation, suppression ou manipulation de l'information; perte de l'accès à des informations exactes, opportunes ou fiables; profusion de mensonges (surabondance d'informations); détérioration des principes communs nécessaires à la prise de décisions.
 Financier/ économique/lié à la privation	Perturbation des moyens de subsistance, pertes économiques, accès plus limité à des ressources ou services essentiels, pertes financières, vols, pillages, fraudes, escroqueries, détournement de ressources, incapacité de se procurer des produits de première nécessité, restrictions en matière de financement, extorsion, atteinte à la réputation, perte du soutien des donateurs.
 Numérique/ technologique	Attaques ciblant l'identité numérique, divulgation malveillante d'informations personnelles, amplification algorithmique de contenus préjudiciables, deepfakes, abus causés par des robots, manipulation des plateformes (voir Annexe I, page 397).
 Longitudinal/ intergénération- nel (dimension transversale)	Effets durables sur les enfants exposés à des discours préjudiciables, rupture de la confiance intergénérationnelle, perpétuation des stéréotypes ou des traumatismes, perte d'espoir.



Racisme en ligne et santé mentale après une catastrophe

Au lendemain d'une catastrophe, les communautés ont besoin de réconfort et de solidarité. Pourtant, pour de nombreux groupes racialisés, le racisme en ligne vient aggraver la tragédie. À la suite du crash d'un avion de la compagnie aérienne Air India en juin 2025, qui a coûté la vie à au moins 290 personnes²⁰, les communautés indiennes du monde entier ont subi un torrent d'insultes racistes sur Internet, subissant moqueries, propos injurieux, stéréotypes et blagues déshumanisantes²¹. De même, après les tremblements de terre de 2023 en Turquie (qui ont causé la mort de quelque 50 000 personnes, dont 7 200 demandeurs d'asile syriens), les Syriens survivants, déjà marginalisés, ont été victimes de stigmatisation et d'attaques xénophobes, tant en ligne que hors ligne. Ils ont été accusés de piller ou de «prendre les ressources» destinées aux citoyens locaux²².

Cette forme de racisme n'est pas seulement offensante, elle entraîne également des conséquences psychologiques néfastes. Plusieurs études établissent systématiquement un lien entre la discrimination raciale et l'augmentation des taux d'anxiété, de dépression, de troubles de stress post-traumatique et d'idées suicidaires²³. Lorsque le racisme s'ajoute à une catastrophe, il perturbe le deuil collectif, isole les survivants et compromet le relèvement à long terme²⁴. Pour les communautés de la diaspora et les demandeurs d'asile, il renforce le sentiment de ne pas appartenir aux pays où ils ont émigré ou cherché refuge²⁵.

Ces tendances révèlent des défaillances systémiques : les plateformes de médias sociaux permettent à la haine de se répandre, les gouvernements sous-financent la lutte contre le racisme et les systèmes humanitaires négligent souvent les liens qui existent entre race, santé mentale et vulnérabilité. Pour faire face aux conséquences du racisme sur la santé mentale après une catastrophe, il convient d'adopter une approche coordonnée et proactive, car «le bien-être des humains dépend de la capacité personnelle des individus et de la capacité collective des peuples de relever les défis et de surmonter les difficultés qu'ils rencontrent dans leur environnement en utilisant de manière intelligente, habile et éthique les atouts et ressources dont ils disposent. Les individus et les communautés coexistent de manière dynamique, interdépendante et intégrale²⁶».

Les plateformes de médias sociaux doivent être tenues responsables de la diffusion des discours de haine. Il convient donc de mettre en place des systèmes de détection avancés (éventuellement fondés sur l'intelligence artificielle) assortis de mécanismes de signalement transparents et de délais de suppression des contenus clairement définis. En parallèle, les besoins en matière de santé mentale des modérateurs humains chargés de visionner et de supprimer les contenus racistes et traumatisants doivent également être véritablement pris en compte²⁷. Tous les programmes d'intervention mis en œuvre suite à une catastrophe, y compris ceux liés à la santé mentale et au soutien psychosocial, devraient intégrer des soins et un soutien antiracistes et adaptés à la culture. Les professionnels du secteur devraient recevoir une formation leur permettant de reconnaître les conséquences psychologiques propres au racisme, en particulier au lendemain d'une crise, et d'y faire face. Quant aux gouvernements, ils devraient mener des campagnes d'éducation fondées sur des données probantes et la mobilisation communautaire afin de lutter contre la stigmatisation, la xénophobie et la mésinformation après une catastrophe. Ces campagnes devraient être déployées dans les écoles, les lieux de travail

et les institutions publiques afin d'avoir un impact durable. En fin de compte, il s'agit de défendre la dignité humaine et notre humanité.

Yasin Duman

Spécialiste de la recherche

Centre du Mouvement international de la Croix-Rouge et du Croissant-Rouge pour la santé mentale et le soutien psychosocial

Shona Whitton

Conseillère technique

Centre du Mouvement international de la Croix-Rouge et du Croissant-Rouge pour la santé mentale et le soutien psychosocial

1.8 Intelligence artificielle et informations préjudiciables

Nos opinions sont de plus en plus influencées par des systèmes que nous ne comprenons pas pleinement et que nous ne pouvons pas véritablement remettre en question. Un grand nombre de personnes pensent avoir accès aux mêmes contenus que les autres sur leurs fils d'actualité, mais tel est rarement le cas. Les algorithmes de recommandation fondés sur l'IA sélectionnent les contenus proposés, ce qui contribue souvent à accentuer les réactions émotionnelles et les chambres d'écho. Cette dynamique érode de manière subtile mais efficace notre capacité de choisir comment nous accédons, interprétons et utilisons l'information.

Les contenus préjudiciables se sont rapidement multipliés avec la prolifération des outils d'IA. Selon le premier *Rapport scientifique international sur la sûreté de l'IA avancée* (communément appelé «rapport Bengio») publié en janvier 2025, s'il était déjà difficile de distinguer le vrai du faux, **l'IA a fondamentalement modifié le rapport de force en faveur de ceux qui contrôlent la production et la diffusion de l'information.** Le rapport souligne que l'IA redéfinit la manière dont l'information est produite, qui en tire profit ou en pâtit, et comment elle est manipulée. Cette technologie permet de générer rapidement et à grande échelle des contenus persuasifs et semblables à ceux créés par l'homme, malgré son manque de compréhension approfondie des concepts. Ces contenus étant souvent impossibles à distinguer de ceux produits par des humains, les individus ont tendance à surestimer leur capacité de les repérer, ce qui augmente leur vulnérabilité à la manipulation. Les contenus chargés d'émotion ou personnalisés, en particulier lorsqu'ils sont associés à des données provenant des médias sociaux, peuvent fortement influencer les perceptions et les croyances²⁸.

L'impact global des fausses informations générées par l'IA fait encore débat. Certaines études indiquent que leur diffusion et leurs répercussions pourraient être limitées, tandis que d'autres soulignent qu'elles pourraient être plus concentrées, plus difficiles à tracer et susceptibles d'entraîner des effets indésirables. Ce qui est certain, c'est que l'espace informationnel devient de plus en plus complexe et polémique. Bien qu'aucun consensus scientifique n'ait été établi sur l'impact global des fausses informations sur les sociétés, leur potentiel de propagation virale est bien documenté. Diverses mesures telles que l'ajout de filigranes ou le filtrage des contenus semblent prometteuses pour repérer et atténuer les manipulations générées par l'IA, mais celles-ci restent insuffisantes. De plus, il convient de trouver un équilibre entre la lutte contre la manipulation de l'information et la protection de la liberté d'expression. Plusieurs rapports font état d'une prévalence accrue des contenus deepfake générés par l'IA (voir

l'encadré), mais les preuves scientifiques restent globalement limitées. De plus en plus de rapports isolés sont publiés sur les préjudices causés par les faux contenus générés par l'IA, mais sans preuves plus solides, il demeure difficile d'évaluer leur ampleur et leurs répercussions réelles²⁹.



Faux contenu généré par l'IA

Contenu audio, textuel ou visuel, produit par l'IA générative, qui dépeint des personnes ou des événements d'une manière malveillante ou trompeuse qui diffère de la réalité. Il peut par exemple montrer des personnes faisant des choses qu'elles n'ont pas faites ou disant des choses qu'elles n'ont pas dites, modifier le lieu d'événements réels ou montrer des événements qui n'ont pas eu lieu³⁰.



Deepfake

Médias synthétiques – consistant souvent en contenu vidéo ou audio ou en images – créés à l'aide de techniques d'IA, en particulier l'apprentissage profond, pour manipuler ou générer de manière réaliste des contenus qui représentent des événements ou des personnes d'une manière qui n'est pas conforme à la réalité.

Ils constituent un sous-ensemble de faux contenus générés par l'IA. Peuvent être cités par exemple les échanges de visages dans des vidéos et les faux enregistrements vocaux.

Les contenus générés par l'IA devenant de plus en plus difficiles à distinguer de ceux créés par l'homme, leur détection demeure un défi permanent. Les techniques d'authentification des médias, telles que les filigranes numériques, offrent une certaine protection, mais elles sont faciles à contourner ou à supprimer, en particulier dans les environnements antagonistes ou à haut risque³¹.

Perspectives des contributeurs 1.9

Détection des contenus synthétiques dans les contextes complexes

L'IA générative offre de nouvelles possibilités de créer et diffuser des informations préjudiciables. Qu'il s'agisse de «conversations divulguées» entre des candidats aux élections fabriquées par l'IA ou de vidéos manipulées montrant d'importantes personnalités militaires communiquant de fausses informations, ce type de contenus peut être utilisé pour induire en erreur et manipuler. Les contenus synthétiques trompeurs perturbent l'écosystème de l'information de diverses manières. Ils peuvent être utilisés pour produire des représentations très réalistes d'événements qui ne se sont pas produits, mais aussi pour discréditer des preuves et des vidéos authentiques en permettant à des personnalités politiques d'affirmer que des images réelles sont fausses et ont été générées par l'IA. Ces contenus peuvent en outre fausser la perception du public vis-à-vis des crises humanitaires en cours en fabriquant et en diffusant des vidéos générées par l'IA qui dépeignent des événements réels.

Diverses solutions techniques permettent de lutter contre les menaces engendrées par l'IA générative, notamment les outils de détection de l'IA qui identifient les schémas non naturels ou les incohérences dans le contenu en vue de détecter une éventuelle manipulation par l'IA. Par exemple, la «Deepfakes Rapid Response Force», une initiative de l'ONG WITNESS, contribue aux efforts de détection afin d'atténuer les crises liées à l'information en temps réel. Cette initiative met en relation les acteurs du secteur de l'information qui se trouvent en première ligne avec d'éminents spécialistes de la détection de l'IA afin de les aider à analyser les contenus suspectés d'avoir été générés par cette technologie.

Le travail réalisé par la Deepfakes Rapid Response Force revêt une dimension sociotechnique, l'objectif étant de rendre la détection de l'IA plus efficace dans les situations réelles, tout en soulignant la nécessité de disposer d'outils capables de traiter des contenus de mauvaise qualité et fortement compressés, de s'adapter à divers contextes linguistiques et culturels, et de fournir des analyses claires et exploitables. Pour que les technologies de détection de l'IA puissent réellement aider les acteurs humanitaires dans les situations de crise, les solutions proposées doivent être techniquement avancées et adaptées aux besoins, aux contraintes et aux expériences des utilisateurs opérant dans des contextes complexes et à haut risque.

Zuzanna Wojciak

Associée de programme, Menaces et possibilités technologiques

WITNESS

Les avis divergent sur la question de savoir si le réalisme accru des faux contenus rend la manipulation plus efficace ou si leur diffusion constitue la principale difficulté. Selon certains spécialistes, le plus grand défi n'est pas la création de faux contenus mais leur diffusion à grande échelle³². Plusieurs études indiquent en outre que les «cheapfakes» (c'est-à-dire les manipulations simples et moins sophistiquées de contenus audiovisuels) peuvent être tout aussi préjudiciables que les deepfakes sophistiqués, ce qui renforce l'idée selon laquelle la diffusion et la portée peuvent avoir plus d'importance que la qualité³³. Si les plateformes de médias sociaux ont recours à la modération, à

l'étiquetage et à la vérification de la crédibilité des sources en vue de limiter la propagation de contenus manipulés, ces mesures suscitent certaines préoccupations quant à la liberté d'expression. Par ailleurs, des études montrent que les algorithmes «privilégient souvent les interactions et la viralité plutôt que l'exactitude ou l'authenticité des contenus», ce qui peut favoriser la propagation rapide de contenus générés par l'IA dans le but de manipuler l'opinion publique³⁴.

Le rapport des Nations Unies intitulé *Gouverner l'IA au bénéfice de l'humanité* souligne qu'il est «plus utile d'examiner les risques du point de vue des populations vulnérables et des biens communs³⁵». Il se fonde sur l'Enquête sur les risques liés à l'IA, dans laquelle 348 spécialistes du domaine ont donné leur avis sur les tendances et les risques émergents liés à cette technologie. Le rapport précise toutefois que la gestion des risques ne doit pas se limiter à dresser une liste des risques ou à les classer par ordre de priorité. Il préconise de les définir sur la base des vulnérabilités, en axant les interventions non plus sur le **type** de risque (par exemple «risque pour la sécurité»), mais sur **la personne** exposée, le **lieu** et le **responsable**. Cette approche met en lumière les vulnérabilités des individus, des systèmes politiques, de la société, de l'économie et de l'environnement. S'agissant de la sécurité, le rapport insiste sur l'importance de garantir la fiabilité et l'interprétabilité des systèmes d'IA, mais aussi d'évaluer et d'atténuer les risques pour les droits individuels et collectifs, la sécurité nationale et internationale et la sûreté publique dans différents contextes³⁶.

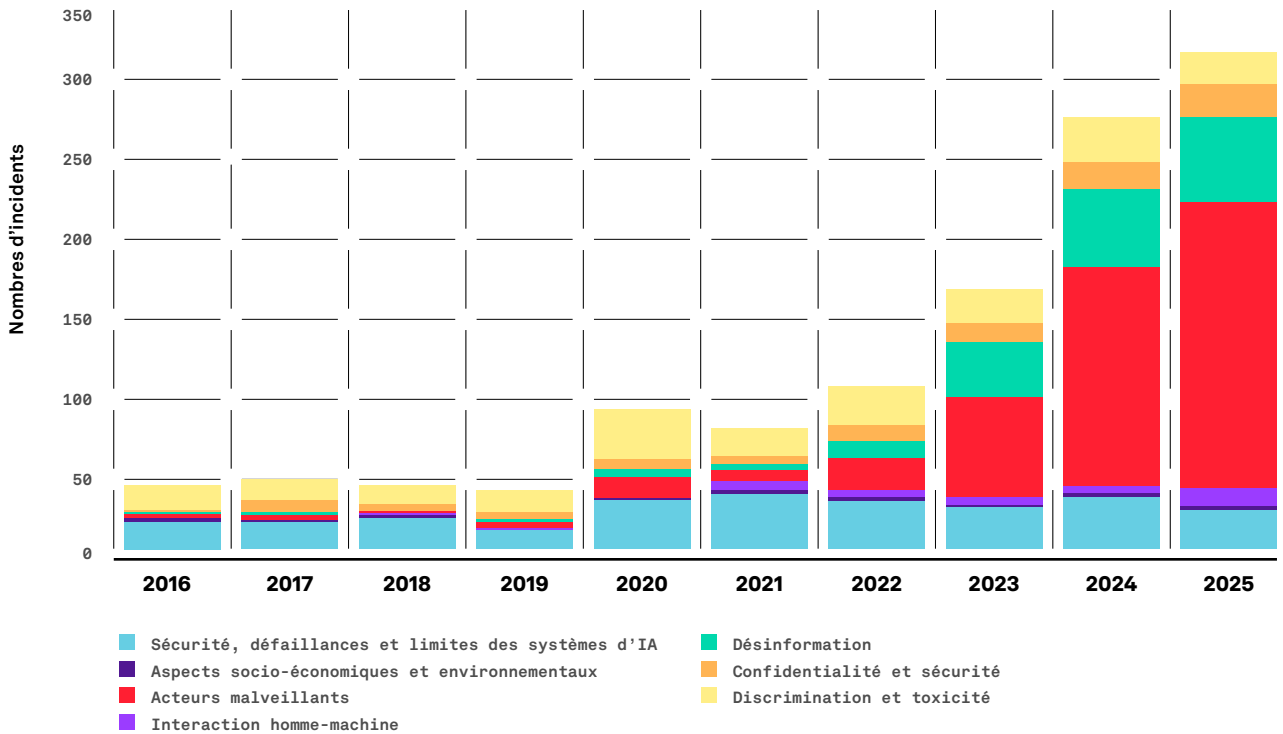
L'accès aux outils d'IA reste inégal, ce qui creuse les fossés numériques existants. Les personnes qui ne disposent pas d'une connexion fiable, ne maîtrisent pas les outils numériques ou n'ont pas accès à des contenus dans leur langue continuent d'être exclues ou mal informées.

Il est clair qu'un suivi systématique est nécessaire.

Fig 1.3

Système de suivi des incidents d'IA

Ce système de suivi montre l'augmentation des incidents d'IA au fil du temps



Sources: MIT AI Risk Initiative – Tableaux de bord des outils «AI Incident Tracker» et «Scalable AI Incident Classification»

De plus amples informations sur la réglementation, les droits et les politiques sont présentées au [Chapitre 5, page 185](#).

1.9 Les discours façonnent les perceptions, la confiance et l'action

Les discours sont les cadres à travers lesquels les personnes interprètent leurs expériences et le monde qui les entoure. Ils donnent aux réalités personnelles une dimension politique, aux réalités locales une dimension mondiale. Les récits percutants simplifient les situations complexes et offrent une certaine cohérence, souvent au détriment des nuances. La manière dont les faits sont présentés à toute son importance. Dans les campagnes d'informations préjudiciables, les discours revêtent généralement un caractère historique ou culturel, et de nombreux acteurs adoptent une stratégie à long terme.

Un événement n'a de poids que si les populations **croient** qu'il s'est produit. Dans l'écosystème de l'information actuel, les perceptions peuvent l'emporter sur la vérité: des incidents inventés peuvent avoir des conséquences réelles, alors que ceux qui sont avérés peuvent être ignorés. Les résultats sont de moins en moins déterminés par les *faits*, mais de plus en plus par des facteurs psychologiques, politiques et algorithmiques.

La confiance est facilement érodée par la déformation et la distraction. Les discours politiques s'inspirent souvent de tactiques d'utilisation de l'information en temps de guerre, tandis que les conflits armés sont influencés par des batailles en ligne³⁷. En cliquant, en aimant ou en partageant, nous prenons tous le risque d'y participer et d'alimenter une lutte plus large pour attirer l'attention, influencer les croyances et renforcer les préjugés.

Dans les conflits armés, les civils sont souvent entraînés malgré eux dans le tourbillon de l'information, ce qui les expose à de sérieux risques en matière de sécurité, par exemple en publiant sur les médias sociaux des photos ou des vidéos géolocalisées de bombardements, en relayant des informations sur le contrôle de certains territoires ou en documentant des attaques pour obtenir de l'aide ou demander des comptes.

Ces combats qui se chevauchent, souvent dans l'ombre, déterminent notre perception du monde et notre comportement. Dans ce «réseau de réseaux», il n'existe aucun terrain neutre³⁸ : l'objectif est d'exercer une influence, un contrôle et un pouvoir. Comme le soulignent Singer et Brooking, l'influence ne s'obtient pas seulement par la force, mais également en déterminant la manière dont les populations interprètent le monde, en provoquant des réactions émotionnelles qui incitent à l'action et en cultivant un sentiment d'identité et d'appartenance communes. Les acteurs qui y parviennent sont ceux qui élaborent et diffusent des récits convaincants à l'échelle mondiale, avec cohérence et régularité³⁹.

Dans le paysage numérique, les discours efficaces reposent sur la simplicité, car ils permettent d'assimiler rapidement les messages. C'est pourquoi les images ont un tel impact. La résonance et la nouveauté stimulent la participation. Or, la simplicité n'est pas le point fort du secteur humanitaire ; certains affirment même qu'il a compliqué la compréhension des efforts déployés⁴⁰.

Les tensions géopolitiques intensifient la polarisation au sein des sociétés et au-delà des frontières. Ce phénomène alimente les divisions, la méfiance et les discours contradictoires qui compliquent l'action humanitaire et nuisent à la cohésion sociale. Pour revenir sur le matériel et les contenus, le contrôle de ces deux éléments est de plus en plus considéré comme une priorité stratégique, ce qui a des conséquences directes sur l'action humanitaire.

Perspectives des contributeurs 1.10



L'impact des informations préjudiciables sur les civils : réflexions fondées sur deux exemples

Partout dans le monde, les informations préjudiciables accentuent la méfiance à l'égard des institutions, divisent les communautés et compromettent la capacité des populations de prendre des décisions éclairées. En 2025, le réseau Communiquer avec les communautés touchées par une catastrophe (CDAC) a réalisé une étude dans deux situations de crise

très différentes – au Soudan et au Liban – afin de déterminer comment les informations préjudiciables faussent la prise de décisions des civils, érodent la confiance et perturbent la vie quotidienne. Bien que les deux contextes diffèrent sensiblement, les schémas sont les mêmes et permettent de mieux comprendre le processus global de diffusion d'informations préjudiciables dans les crises humanitaires.

Dans les deux pays, les informations préjudiciables ont influencé des décisions cruciales, souvent irréversibles. Au Liban, des familles de réfugiés ont retardé ou précipité leur retour en Syrie en se fiant aux rumeurs circulant sur les médias sociaux concernant la fermeture des frontières ou de nouvelles mesures d'indemnisation mises en place par les Nations Unies, alors qu'aucune de ces informations ne pouvait être vérifiée. Au Soudan, des discours manipulés au sujet du contrôle de certains territoires ont incité les civils à retourner dans des zones contestées ou à reporter leur évacuation, les exposant ainsi à de réels dangers.

Les informations préjudiciables fracturent également les communautés. Au Liban, l'incertitude économique et la désillusion politique ont créé un terrain propice à la stigmatisation. Les réfugiés ont été présentés comme favorisés, les communautés d'accueil comme négligées, tandis que les fausses informations circulant dans les groupes WhatsApp ont attisé le ressentiment. Au Soudan, le personnel des ONG, les groupes d'entraide et les minorités ethniques ont été accusés de servir des intérêts politiques, ce qui a sapé la confiance et entraîné une vague de violence contre des membres de la société civile.

Les répercussions sur les moyens de subsistance ont été tout aussi importantes. Au Soudan, les rumeurs visant les groupes déplacés ont conduit à leur exclusion de différents marchés, notamment celui du logement et des emplois informels. Dans certaines régions, des secteurs professionnels tout entiers, tels que la boucherie ou la vente de thé, ont été déstabilisés, la peur et la méfiance ayant engendré une baisse de la fréquentation et des revenus. Ces dynamiques ont exacerbé les inégalités existantes et attisé les tensions intercommunautaires.

Enfin, plusieurs études montrent à quel point les informations préjudiciables sapent la confiance tant dans les institutions que dans les structures informelles. Dans les deux pays, les enseignants, les militants et les travailleurs humanitaires qui assuraient autrefois la transmission de l'information hésitent désormais à s'exprimer par crainte des représailles. Dans les deux contextes, le constat général est simple : les informations préjudiciables ne se contentent pas de semer la confusion, elles divisent, isolent et déstabilisent, ce qui a de lourdes conséquences pour les populations touchées par les crises.

Ila Schoop Rutten

Responsable de l'intégrité de l'information

Réseau CDAC

1.10 **Contrôle de l'accessibilité et modération des contenus**

Les plateformes modèrent les contenus (en associant une supervision humaine à des outils automatisés, notamment pour la vérification des faits et la suppression

de contenus) selon leurs propres conditions d'utilisation. Elles jouent ainsi un rôle de gardiennes, en déterminant ce qui est autorisé et en s'exposant à la critique tant pour leur incapacité de lutter efficacement contre les contenus préjudiciables que pour leur censure excessive. Récemment, les principales plateformes ont réduit leurs efforts de modération, sous l'effet de pressions économiques, de préoccupations liées à leur réputation et d'influences politiques ou idéologiques. Elles s'appuient de plus en plus sur les signalements des utilisateurs ou le déclassement des contenus pour se conformer aux exigences juridiques.

Bien qu'Internet ait été fondé sur les principes d'ouverture et de confiance, il reste vulnérable aux fermetures, aux coupures de services et aux restrictions ciblées.

Perspectives des contributeurs 1.11



L'action locale dans la lutte pour la connectivité humanitaire

Depuis 2021, le gouvernement militaire a coupé la connexion Internet de millions de personnes dans tout le Myanmar, en particulier dans les régions touchées par des conflits. L'ONG Access Now et la coalition #KeepItOn (qui regroupe plus de 334 organisations nationales, internationales, régionales et locales) ont enregistré 85 coupures rien qu'en 2024⁴¹.

Il ne s'agit pas d'un cas isolé. En 2024, la coalition #KeepItOn a recensé 296 coupures dans 54 pays, ce qui constitue un record historique et une augmentation de 35% du nombre de pays touchés par ce type d'incident par rapport au précédent pic enregistré en 2022.

Les données de 2024 confirment que les conflits ont été la principale cause des coupures Internet pour la deuxième année consécutive et que les mesures déployées pour couper l'accès (utilisation de dispositifs de brouillage, coupure de câbles, destruction d'infrastructures et sabotage des fournisseurs de services) ne cessent de se multiplier. Les crises humanitaires prennent de plus en plus la forme d'urgences complexes, certains pays étant touchés par une succession de catastrophes, d'événements liés aux changements climatiques et de conflits.

Le Myanmar illustre parfaitement ce phénomène : le pays a été frappé par le cyclone Mocha en 2022 et un séisme dévastateur en 2025, tout en continuant de faire face à des violences internes. Lorsque le cyclone Mocha a touché l'ouest du pays, l'absence de connectivité Internet a exacerbé les effets de la tempête. Les habitants n'ont pas pu accéder aux informations vitales concernant les opérations d'évacuation, les alertes et même les interventions de secours, car les services étaient indisponibles. Trois ans plus tard, le séisme qui a dévasté la région de Sagaing a fait plus de 3 000 victimes. La région était déjà soumise à un arrêt prolongé des services depuis 2021. Cette coupure, qui a touché plus d'un tiers du pays, a une fois de plus compliqué les opérations de sauvetage et de secours.

Le rôle essentiel que joue la connectivité pour la population civile au cours d'une catastrophe montre combien il est important d'assurer une gestion de l'information résiliente et locale. En imposant ou en facilitant les coupures, en particulier dans les crises complexes, les acteurs qui nuisent délibérément à la mise en œuvre d'une intervention humanitaire rapide, sûre et efficace sont en fin de compte responsables de décès qui auraient pu être évités. Cette situation peut en outre empêcher les opérateurs de télécommunications de déployer des infrastructures d'intervention suffisantes, car l'obtention des autorisations nécessaires est retardée ou bloquée pour des raisons politiques ou tactiques.

Alors que les rapports humanitaires établis par les acteurs internationaux se concentraient sur les conséquences visibles des catastrophes et des conflits, des groupes de défense locaux ont systématiquement attiré l'attention sur les perturbations de l'accès à Internet, leurs conséquences et les dommages qui en découlent.

Wai Phy

Analyste des politiques
pour l'Asie-Pacifique

[Access Now](#)

Faiz Naeem

Associé de programme
pour l'Asie-Pacifique

[Access Now](#)

Giulio Coppi

Spécialiste des affaires
humanitaires

[Access Now](#)

1.11 Analyse et surveillance des informations préjudiciables

Il est primordial d'analyser les informations préjudiciables et les risques qu'elles entraînent pour comprendre les tactiques, les techniques et les processus utilisés par ceux qui les créent et les diffusent. Cette analyse doit tenir compte de facteurs contextuels, historiques, politiques, sociaux, culturels et économiques. Il existe plusieurs cadres analytiques, notamment les 5 W⁴² (une méthode couramment employée par les journalistes) et les cadres ABCDE⁴³, DISARM⁴⁴ et MITRE ATT&CK⁴⁵. Ces derniers requièrent des capacités d'analyse et des niveaux de formation différents, et sont étudiés plus en détail au [Chapitre 3, page 147](#).

Pour réaliser une analyse efficace, il convient en outre de surveiller les environnements opérationnels et informationnels afin de repérer les contenus préjudiciables, tout en intégrant des approches communautaires permettant de mieux comprendre leur fonctionnement et leurs effets. De nombreuses organisations humanitaires utilisent toute une gamme d'outils, disponibles sur le marché ou mis au point en interne, qui combinent des méthodes automatisées et manuelles pour extraire des données publiques. Trop souvent, elles se concentrent uniquement sur leur réputation, plutôt que d'évaluer l'incidence de ces informations sur leur sécurité, leurs opérations et l'action humanitaire au sens large.

Le secteur humanitaire doit de toute urgence mettre en place une surveillance et une écoute sociale plus systématiques, y compris un dispositif d'analyse des sentiments, afin de repérer les informations préjudiciables, de les comprendre et de définir des mesures appropriées. À cet égard, la surveillance reste fragmentée, sous-financée et souvent ponctuelle. En intégrant des approches communautaires et l'analyse de sentiments aux mécanismes d'écoute sociale, les organisations pourront repérer et suivre les informations qui circulent, mais aussi les sentiments des populations (peur,

colère, méfiance ou soutien) et ainsi mieux comprendre l'évolution des perceptions et des risques.

Sans une détection et une analyse structurées en temps réel des discours en ligne et hors ligne, les acteurs humanitaires risquent d'être réactifs plutôt que d'anticiper, ce qui permettrait à des informations préjudiciables de se propager de manière incontrôlée avant qu'elles puissent être réfutées et contribuerait donc à éroder la confiance. Une surveillance systématique, une écoute sociale et une analyse de sentiments, fondées sur des garanties éthiques et la participation des communautés, permettraient aux organisations de repérer rapidement les discours émergents, d'anticiper les risques, d'adapter leurs messages et leurs programmes en temps réel et de renforcer la résilience tant au niveau communautaire qu'institutionnel.

Cependant, force est de constater que les dispositifs de suivi à disposition ne sont pas suffisamment performants, non pas en raison d'un manque d'outils, mais principalement de contraintes budgétaires. Certaines préoccupations demeurent également quant aux répercussions juridiques et éthiques de l'extraction de données, qui peut être perçue comme intrusive ou assimilée à de la surveillance.

Perspectives des contributeurs 1.12

Détection et suivi des informations préjudiciables dans les contextes humanitaires et de crise

Dans les contextes humanitaires et de crise, les informations préjudiciables ou trompeuses relatives aux interventions humanitaires peuvent mettre en danger le personnel humanitaire, nuire à la sécurité opérationnelle et compromettre l'efficacité de l'aide. Chez LinkAlong, nous avons travaillé activement sur ce problème et mis au point une plateforme de surveillance open source fondée sur l'IA, spécialement conçue pour détecter et suivre ces menaces.

Grâce aux progrès récents en matière de traitement du langage naturel fondé sur l'IA, il est désormais possible de surveiller de manière systématique et automatisée de grands volumes de contenu open source. En s'appuyant sur de grands modèles de langage, on peut repérer, classer, regrouper et caractériser les informations préjudiciables avec une grande précision et en temps quasi réel. Le processus s'articule généralement autour de trois étapes :

- 1 **La collecte de données** à partir de diverses sources, notamment les médias sociaux ouverts, les forums publics et les organes d'information pertinents ;
- 2 **L'analyse approfondie des textes** à l'aide de l'IA en vue d'identifier les différents discours, de classer les types de contenus préjudiciables et d'évaluer les risques potentiels ;

- 3 La communication des résultats** aux parties prenantes au moyen de rapports structurés et opportuns afin qu'elles puissent intervenir rapidement et de manière éclairée.

Cependant, plusieurs obstacles entravent son adoption à grande échelle. Premièrement, les grands modèles de langage peinent encore à traiter les langues et dialectes peu documentés mais couramment utilisés dans les régions touchées par les crises. Deuxièmement, l'accès aux données pertinentes des médias sociaux est de plus en plus limité, tandis que les discours préjudiciables migrent souvent vers des plateformes cryptées ou fermées telles que Telegram ou WhatsApp. Troisièmement, les différences régionales concernant l'utilisation des médias sociaux exigent des solutions adaptées, ce qui peut compromettre leur déploiement à plus grande échelle. Quatrièmement, la plupart des organisations ne disposent pas du budget nécessaire pour assurer un suivi continu, et les donateurs sous-estiment souvent ces activités malgré leur importance stratégique.

Les efforts sont également fragmentés : un grand nombre d'organisations conçoivent des outils isolés en interne, des outils dont la maintenance est coûteuse et qui ne permettent pas de partager les informations, alors que les menaces et les discours malveillants se recourent souvent.

Les organisations doivent de toute urgence **renforcer leur coopération** afin de mettre en commun leurs ressources, de partager les informations sur les menaces et de tirer conjointement parti de solutions durables. Les bases techniques existent déjà ; il reste à établir un argumentaire et à convaincre les parties prenantes concernées que la surveillance collaborative est une nécessité opérationnelle et non un luxe.

Karl Aberer

Professeur, École polytechnique fédérale de Lausanne

LinkAlong⁴⁸

Perspectives des contributeurs 1.13

Renforcer la résilience : outils permettant de vérifier les faits, de repérer les fausses informations et d'identifier la provenance

Parmi les outils de suivi mis au point pour fournir des données structurées et lisibles par la machine concernant les informations fausses et trompeuses qui circulent en ligne, la base de données «False Claim Fingerprints» créée par l'entreprise NewsGuard recense les discours dont la fausseté a été prouvée dans 102 pays, et ce en langues locales. Elle couvre des sujets tels que les crises humanitaires et les guerres, les urgences sanitaires, les séismes, les catastrophes climatiques et les migrations. Elle permet en outre de surveiller les opérations d'information menées par des acteurs malveillants dans le monde entier. La base de données sert de système d'alerte précoce et de dispositif de sécurité fondé sur les faits pour lutter contre les fausses informations à haut risque. Elle crée ainsi une «empreinte» pour chacune d'entre elles (en indiquant les données probantes

qui ont permis de déterminer que l'information était fausse) avec des URL vérifiés par des humains⁴⁷, des mots-clés, des publications sur les médias sociaux, des ressources multimédias et des extraits dans différentes langues.

Les ONG et autres institutions peuvent utiliser la base de données pour repérer les fausses informations qui les visent (ou qui visent les communautés vulnérables auxquelles elles viennent en aide) avant que celles-ci ne circulent à grande échelle. L'entreprise NewsGuard parvient à identifier et à signaler les informations à haut risque avant qu'elles soient rendues publiques ou vérifiées dans deux tiers des cas, et est la première à réfuter les fausses allégations des gouvernements autoritaires dans 83 % des cas. Les équipes chargées des relations publiques peuvent intégrer les évaluations journalistiques réalisées par NewsGuard dans leurs communications, en utilisant des informations précises, transparentes et fondées sur des données probantes pour défendre leur réputation lorsqu'elles sont attaquées. Alors que de plus en plus d'internautes se tournent vers les robots conversationnels fondés sur l'IA générative pour s'informer, ce type de base de données peut aider à vérifier si de fausses informations ont contaminé les réponses générées par l'IA et ainsi permettre aux organisations de réagir rapidement et de manière stratégique.

Chine Labbe

Vice-présidente principale chargée des partenariats et
rédactrice en chef pour l'Europe et le Canada

*NewsGuard Technologies*⁴⁸

Perspectives des contributeurs 1.14

Garantir la sécurité des communications dans les contextes humanitaires

À une époque où les informations préjudiciables peuvent se propager rapidement et affaiblir la confiance dans les interventions humanitaires, il est primordial de veiller à ce que les canaux de communication soient sécurisés, fiables et dignes de confiance. Twilio est une plateforme de communication avec les clients qui collabore avec des gouvernements et des organismes de services du monde entier pour fournir des informations essentielles aux personnes dans le besoin. Lorsque les plateformes technologiques ont un impact direct sur la vie des populations, les discussions sur la réglementation en matière de sûreté et de sécurité deviennent essentielles.

Les entreprises peuvent fournir de précieuses informations sur le fonctionnement des plateformes de communication et les effets potentiels de la réglementation sur l'innovation et la sécurité des utilisateurs. L'approche de Twilio consiste à collaborer de manière proactive avec les acteurs du secteur afin de protéger les utilisateurs tout en garantissant des communications ouvertes et fiables, en particulier lorsque la sécurité et le bien-être sont en jeu. Twilio dispose de politiques et de mécanismes de mise en œuvre solides afin de prévenir toute utilisation abusive de sa plateforme, y compris les violations de sa politique d'utilisation acceptable. Des systèmes automatisés surveillent les incidents de sécurité ainsi que les utilisations abusives des services.

La technologie et la transition numérique ont une véritable incidence sur les interventions d'assistance humanitaire. Selon un [rapport de l'Union européenne](#), le passage au numérique peut simplifier les démarches pour les réfugiés et les migrants, notamment en facilitant leur intégration dans les communautés d'accueil.

La plateforme Twilio a conclu un [partenariat pluriannuel avec la Fédération internationale](#), dont l'objectif est de fournir une aide sous forme de transferts monétaires et de bons aux personnes touchées par une catastrophe ou un conflit. Le dispositif, qui permet à la Fédération internationale de confirmer l'identité des personnes déplacées à grande échelle, a été amélioré grâce à [l'intégration d'un système d'enregistrement numérique sécurisé](#) par SMS et WhatsApp. Cette collaboration montre comment les partenariats intersectoriels d'aide humanitaire peuvent s'appuyer sur des plateformes de communication sécurisées.

Chiara Kunnie

Responsable des nouvelles activités,
Ventes à impact social

Twilio

Alexandros Koronakis

Directeur principal, responsable des affaires
gouvernementales, politiques et réglementaires
pour l'Europe, le Moyen-Orient et l'Afrique

Twilio

Conclusions : les informations préjudiciables ne sont pas un bruit de fond

Les informations préjudiciables ne sont pas un bruit de fond. En plus d'influencer directement la façon dont les personnes comprennent les crises, la confiance qu'elles accordent et leur accès à l'aide humanitaire et aux services de protection, elles influent aussi, directement et indirectement, sur la sûreté et la sécurité. La lutte contre les informations préjudiciables concerne autant le matériel que le contenu, c'est-à-dire les infrastructures et les propos qui ont un effet sur l'accès, la confiance et le pouvoir.

L'écosystème de l'information devenant de plus en plus complexe, la capacité de le comprendre, d'y répondre et de protéger les populations, les personnes et les organisations concernées de ses atteintes doit être adaptée en conséquence. La gestion de cet écosystème fait désormais partie intégrante des interventions en cas de crise humanitaire. Cette nécessité doit guider la conception et la mise en œuvre des interventions, et sous-tendre un plaidoyer en faveur de changements systémiques plus vastes.

La responsabilité ne peut incomber uniquement aux organisations humanitaires. La lutte contre les informations préjudiciables appelle un renforcement de la résilience, de la confiance et de la mobilisation des communautés, et exige des gouvernements, des entreprises technologiques, des médias, des communautés et des organisations de la société civile qu'ils mènent une action coordonnée et multipartite.

La lutte contre les informations préjudiciables n'est pas facultative. Elle ne requiert rien de moins qu'un changement systémique. Ce changement est essentiel pour protéger les vies, garantir l'application des principes humanitaires et faire de sorte que l'humanité

elle-même reste la force la plus puissante face à la manipulation, à la méfiance et aux divisions.

Il importe maintenant de veiller à ce que les garanties, la mobilisation, les activités de proximité, la confiance et les capacités de résilience nécessaires ne soient pas laissées au hasard ou à la chance, mais soient délibérément utilisées pour protéger les personnes et permettre une action humanitaire fondée sur des principes.

Exigences, objectifs et recommandations

Exigences

Affronter les informations préjudiciables comme une crise humanitaire systémique qui compromet la sécurité, la dignité et l'accès. Intégrer les droits universellement reconnus, la préparation, la redevabilité et la transparence dans les interventions en cas de crise, et agir avec la même urgence que pour les autres menaces humanitaires. Renforcer la collaboration pour garantir le suivi dans l'ensemble du secteur, notamment pour la surveillance des rumeurs, l'écoute sociale et l'analyse de sentiments.

Objectifs

Protéger les personnes : Donner la priorité aux informations fiables et précises dans le cadre de la préparation, de l'intervention et du relèvement en cas de crise, afin de préserver la sécurité, la dignité et l'accès.

Garantir la détection rapide et l'adaptation : Surveiller les discours, les perceptions et les sentiments pour permettre des interventions coordonnées, transparentes et opportunes.

Protéger l'action humanitaire : Protéger le personnel, les volontaires, les populations touchées et les opérations contre les informations préjudiciables, les cybermenaces et les perturbations des infrastructures essentielles, tout en préservant la confiance et la légitimité.

Recommandations

États et décideurs

- Intégrer la gestion des informations préjudiciables dans les cadres de préparation et d'intervention en cas de crise.
- Investir dans des systèmes d'alerte précoce et de vérification permettant la fourniture en temps utile d'informations vitales fiables.
- Faire respecter les règles applicables en ligne et protéger les organisations

humanitaires et les infrastructures essentielles contre l'utilisation malveillante des technologies de l'information.

- Établir des cadres juridiques et politiques clairs qui soutiennent et protègent l'action humanitaire face aux ingérences et aux campagnes d'informations préjudiciables.

Plateformes technologiques

- Privilégier la modération rapide et la vérification des faits avant, pendant et après les crises humanitaires, en prévoyant des voies efficaces de remontée des problèmes pour les

organisations humanitaires. Veiller à ce que les outils fonctionnent dans des environnements multilingues, ou la bande passante est faible et où les ressources sont limitées.

- Intégrer des frictions volontaires sur les plateformes numériques (messages guides questionnant l'exactitude des informations, messages d'avertissement, incitations à «lire avant de partager» ou voies de diffusion plus lentes pour les contenus non vérifiés, par exemple), en particulier dans le contexte des crises humanitaires, afin de restreindre la propagation rapide d'informations préjudiciables et d'encourager les utilisateurs à adopter un comportement plus réfléchi et éclairé. Adapter les algorithmes pour réduire l'amplification des discours préjudiciables visant les organisations humanitaires,
- l'action humanitaire fondée sur des principes et les populations touchées.
- Rendre compte de manière transparente des mesures de modération, des ajustements algorithmiques et de leur impact dans les contextes humanitaires.
 - Concevoir des protocoles de réponse aux crises en collaboration avec les acteurs humanitaires afin de garantir que les interventions sont rapides, adaptées au contexte et conformes aux droits et normes internationalement reconnus.

Acteurs humanitaires

- Traiter la gestion des informations préjudiciables comme un risque opérationnel – et non seulement un problème de communication – intégré dans les activités de diplomatie humanitaire et le dialogue sur la protection avec les États, les organisations régionales et les forums multilatéraux, ainsi que dans les programmes, les cadres de gestion des risques et la planification de la préparation.
 - Former le personnel et les volontaires à la vérification des rumeurs, à la culture numérique et aux programmes d'information sûrs, et garantir l'échange de connaissances entre organisations de confiance.
 - Renforcer la surveillance en combinant des outils utilisant l'IA et le savoir-faire humain
- pour faciliter la détection en temps réel, la cartographie des écosystèmes d'informations préjudiciables et l'alerte précoce ; favoriser la collaboration dans l'ensemble du secteur humanitaire à l'appui de la mise en commun des ressources et des capacités.
- Documenter et analyser systématiquement les exemples d'informations préjudiciables afin de constituer une base de données probantes en vue de l'élaboration de politiques, de l'adaptation et de la sensibilisation.
 - Mobiliser les communautés de manière transparente, en rédigeant des messages avec les communautés concernées et en renforçant la neutralité, l'impartialité et l'indépendance.

Communautés et dirigeants locaux

- Concevoir et soutenir des systèmes locaux de suivi et de vérification des rumeurs.
 - Jouer le rôle d'intermédiaires de confiance pour renforcer la solidarité, la résilience et la confiance du public dans l'intervention humanitaire.
 - Fournir des retours d'information pour les acteurs humanitaires et les autorités en ce qui concerne les déficits de confiance et les préoccupations non résolues.
- Garantir l'inclusivité en veillant à ce que les jeunes, les minorités et les personnes en situation de handicap soient représentés dans les systèmes d'information communautaires.
 - Favoriser le dialogue, la collaboration entre pairs et les initiatives dirigées par les communautés, en vue de contrer les discours préjudiciables et de prévenir la polarisation ou la stigmatisation.

Notes de fin

- 1 Par exemple, en 2006, l'agence Reuters a rompu tout lien avec un photographe indépendant après avoir découvert des images manipulées numériquement, notamment des panaches de fumée ajoutés à des photographies illustrant le conflit israélo-libanais. Voir l'article «Reuters drops photographer over doctored images», *The Guardian*, 7 août 2006. www.theguardian.com/media/2006/aug/07/reuters.pressandpublishing
- 2 Pendant le Printemps arabe, le blog «A Gay Girl in Damascus», prétendument écrit par une femme syro-américaine, s'est révélé être celui d'un personnage fictif créé par un étudiant universitaire écossais. Voir Mackey, R., «Gay Girl in Damascus' Blog a Hoax, American Says», *The New York Times*, 13 juin 2011. www.nytimes.com/2011/06/13/world/middleeast/13blogger.html
- 3 En juillet 2006, au cours du conflit israélo-libanais, deux ambulances de la Croix-Rouge libanaise ont été touchées par des tirs qui ont blessé plusieurs ambulanciers. Des communautés de blogueurs ont affirmé que les dégâts avaient été mis en scène en raison de l'apparence des véhicules. Les enquêtes ont confirmé que l'attaque s'était réellement produite. Voir Conseil des droits de l'homme de l'ONU, Rapport de la Commission d'enquête sur le Liban établi conformément à la résolution S-2/1 du Conseil des droits de l'homme, A/HRC/3/2, 23 novembre 2006, par.173, https://digitallibrary.un.org/record/587605/files/A_HRC_3_2-FR
- 4 Voir, par exemple, Rid, T., *Active Measures: The Secret History of Disinformation and Political Warfare*, 2020.
- 5 Voir, par exemple, West, D. M., «Fighting hurricane misinformation requires aggressive pushback», *Brookings*, 2024. www.brookings.edu/articles/fighting-hurricane-misinformation-requires-aggressive-push-back
- 6 Voir Organisation mondiale de la Santé, «Infodemic». www.who.int/health-topics/infodemic
- 7 De fausses informations selon lesquelles les antennes-relais 5G seraient liées à la propagation du Covid-19 ont conduit à des incendies criminels visant des infrastructures de télécommunications au Royaume-Uni ainsi qu'à des menaces contre des employés du secteur. Les autorités scientifiques ont systématiquement réfuté tout lien entre la technologie 5G et le virus. Voir Hern, A., «Arsonists Attack Phone Mast Serving NHS Nightingale Hospital», *The Guardian*, 14 avril 2020. www.theguardian.com/technology/2020/apr/14/arsonists-attack-phone-mast-serving-nhs-nightingale-hospital
- 8 Nichols, T., *The Death of Expertise: The Campaign Against Established Knowledge and Why It Matters*, 2024.
- 9 *Les Principes mondiaux des Nations Unies pour l'intégrité de l'information: Recommandations pour une action multipartite* ont été officiellement lancés le 24 juin 2024 à l'occasion d'une conférence de presse menée par le Secrétaire général des Nations Unies, M. António Guterres. <https://www.un.org/fr/information-integrity>
- 10 Voir Nations Unies, *Intégrité de l'information sur les plateformes numériques* (Notre Programme commun, Note d'orientation n° 8), Cabinet du Secrétaire général, 2023, p. 2. <https://www.un-ilibrary.org/content/papers/10.18356/30052432-31/read>
- 11 Département de la communication globale des Nations Unies, *Issue Brief 1 – From Principles to Practice: Strengthening Information Integrity*, 2025. www.un.org/sites/un2.un.org/files/un-information-integrity-issue-brief-01.pdf
- 12 IFRC, CICR et Croix-Rouge guinéenne, *Guinée: le Mouvement Croix-Rouge est inquiet des violences contre des volontaires participant à la lutte contre Ebola*, déclaration conjointe, 11 février 2015. <https://www.icrc.org/fr/document/guinee-le-mouvement-croix-rouge-est-inquiet-des-violences-contre-des-volontaires>
- 13 Voir par exemple www.mdpi.com/2673-5172/2/4/34; www.niemanlab.org/2015/09/what-happened-after-7-news-sites-got-rid-of-reader-comments; www.axios.com/2023/07/06/online-comments-medianews-group-newspapers-end.
- 14 Singer, PW. et Brooking, ET., *LikeWar: The Weaponization of Social Media*, 2018, p. 270.
- 15 Ayala Iacucci, A. *Misinformation, Disinformation, and Hate Speech in Humanitarian Contexts*, Grand Challenges Canada, Creating Hope in Conflict initiative, 2024, p. 38. https://humanitariangrandchallenge.org/wp-content/uploads/2024/11/MDH-Scoping-Study_Full-report.pdf. Cette étude préliminaire passe en revue l'ensemble des défis opérationnels, des approches stratégiques et des solutions innovantes pour lutter contre la désinformation, la désinformation et les discours de haine dans les situations humanitaires liées à des conflits.
- 16 Borges do Nascimento, I.J., Pizarro, AB., Almeida, JM., et al., «Infodemics and health misinformation: a systematic review of reviews», *Bull World Health Organ*, vol. 100, n° 9, 2022, 544–561 (doi:10.2471/BLT.21.287654), p. 557.
- 17 CICR, *Addressing Harmful Information in Conflict Settings: A Response Framework for Humanitarian Organizations*, 2025, p. 8. www.icrc.org/en/publication/addressing-harmful-information-conflict-settings-response-framework-humanitarian
- 18 *Ibid.*
- 19 Assemblée générale des Nations Unies, *Désinformation et liberté d'opinion et d'expression pendant les conflits armés – Rapport de la Rapporteuse spéciale sur la promotion et la protection du droit à la liberté d'opinion et d'expression* (A/77/288), août 2022, paragraphe 59. <https://docs.un.org/fr/a/77/288>
- 20 Radford, A., Tanno, S., et Kemp, O., «At least 290 killed in Air India crash», CNN, 12 juin 2025. <https://edition.cnn.com/world/live-news/ahmedabad-india-plane-crash-06-12-25>
- 21 Silva, A., «Social media flooded with racist comments after Air India crash, but Indians are speaking out», *ABC News*, 17 juin 2025. www.abc.net.au/news/2025-06-17/air-india-plane-crash-indian-racism-social-media-comments/105420614
- 22 Mellersh, N., «Anti-Syrian sentiment grows in Turkey following earthquake», *Infomigrants*, 16 février 2023. www.infomigrants.net/en/post/46835/antisyrain-sentiment-grows-in-turkey-following-earthquake; Ozduzen, O., Korkut, U., et Ozduzen, C., «“Refugees are not welcome”: Digital racism, online place-making and the evolving categorization of Syrians in Turkey», *New Media & Society*, vol. 23, n° 11, 2022, p. 3349–3369. <https://doi.org/10.1177/1461444820956341>
- 23 Carter, RT., «Racism and Psychological and Emotional Injury: Recognizing and Assessing Race-Based Traumatic Stress», *The Counseling Psychologist*, vol. 35, no 1, 2007, p. 13–105. <https://doi.org/10.1177/0011000006292033>; Hankerson, SH., Moise, N., et Wilson, DR., «The Intergenerational Impact of Structural Racism and Cumulative Trauma on Depression», *The American Journal of Psychiatry*, vol. 179, no 6, 2022, p. 434–440. <https://doi.org/10.1176/appi.ajp.21101000>; Polanco-Roman, L., DeLapp, RC., et Dackis, MN., «Racial/ethnic discrimination and suicide-related risk in a treatment-seeking group of ethnorracially minoritized adolescents», *Clinical Child Psychology and Psychiatry*, vol. 28, no 4, 2023, p. 1305–1320. <https://doi.org/10.1177/13591045221132682>
- 24 Rodriguez-Díaz, CE., et Lewellen-Williams, C., «Race and Racism as Structural Determinants for Emergency and Recovery Response in the Aftermath of Hurricanes Irma and Maria in Puerto Rico», *Health Equity*, vol. 4, no 1, 2020, p. 232–238. <https://doi.org/10.1089/heap.2019.0103>
- 25 Gatwiri, K., et Anderson, L., «Boundaries of Belonging: Theorizing Black African Migrant Experiences in Australia», *International Journal of Environmental Research and Public Health*, vol. 18, no 1, 2020, p. 38. <https://doi.org/10.3390/ijerph18010038>
- 26 Quiñones-Rosado, R., «Liberation Psychology and Racism», dans Comas-Díaz, L. et Torres Rivera, E., (Eds.), *Liberation Psychology*:

- Theory, Method, Practice, and Social Justice*, American Psychological Association, 2020, p 53–68. <https://doi.org/10.1037/0000198-004>
- 27 Perrigo, B., «Exclusive: New Global Safety Standards Aim to Protect AI's Most Traumatized Workers», *Time*, 19 juin 2025. <https://time.com/7295662/ai-workers-safety-rules>
- 28 Bengio, Y, Mindermann, S., Privitera, D., et al., *International Scientific Report on the Safety of Advanced AI*, International Scientific Panel on the Safety of Advanced AI, 2025. <https://internationalaisafetyreport.org>
- 29 *Ibid.*
- 30 *Ibid.*, p 62.
- 31 *Ibid.*, p 65.
- 32 *Ibid.*, p 69.
- 33 *Ibid.*
- 34 *Ibid.*, p 69.
- 35 Nations Unies, *Gouverner l'IA au bénéfice de l'humanité: Rapport final*, 2024, p 32. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_fr.pdf. L'Organe consultatif de haut niveau sur l'intelligence artificielle, établi par le Secrétaire général des Nations Unies et composé de 32 experts internationaux, vise à faire en sorte que le développement de l'IA respecte les droits fondamentaux et tienne compte des Objectifs de développement durable.
- 36 *Ibid.*, p 33, 35, 66.
- 37 Voir par exemple, Singer, PW., et Brooking, ET., *LikeWar: The Weaponization of Social Media*, 2018, p 262.
- 38 *Ibid.*
- 39 *Ibid.*, p 160.
- 40 Voir par exemple Slim, H., *Solferino 21: Warfare, Civilians and Humanitarians in the Twenty-First Century*, 2022.
- 41 Access Now, «Lives on hold: internet shutdowns in 2024», 24 février 2025. www.accessnow.org/internet-shutdowns-2024; DW, «Myanmar: Why is junta shutting down internet?», 27 février 2025. www.dw.com/en/myanmar-why-is-junta-shutting-down-internet/a-71770580; Myanmar Internet Project, Access Now, et al., «Myanmar must lift internet restrictions following devastating earthquake», déclaration conjointe, 31 mars 2025. www.accessnow.org/press-release/call-for-lifting-of-internet-restrictions-myanmar
- 42 Les «5 W» (qui, quoi, quand, où et pourquoi, «Who», «What», «When», «Where» et «Why» en anglais) constitue un cadre journalistique normalisé visant à garantir l'exhaustivité et la clarté de l'information. Le «How» (comment) est parfois ajouté comme sixième élément.
- 43 «A» pour «Actor» (acteur), «B» pour «Behaviour» (comportement), «C» pour «Content» (contenu), «D» pour «Degree» (degré), «E» pour «Effect» (effet). Pamment, J., *The EU's Role in Fighting Disinformation: Crafting A Disinformation Framework*, 2020.
- 44 Le dispositif DISARM (*Disinformation Analysis and Risk Management*) est un cadre open source mis au point par la Fondation DISARM qui s'appuie sur des méthodologies de cybersécurité pour aider les organisations à identifier, analyser et coordonner leurs réponses aux campagnes de désinformation grâce à une taxonomie commune des tactiques, techniques et procédures (voir www.disarm.foundation).
- 45 Le dispositif MITRE ATT&CK (Adversarial Tactics, Techniques and Common Knowledge) une base de connaissances ouverte mise au point par la Fondation MITRE qui documente les tactiques et techniques utilisées par les cybercriminels dans le cadre de cyberattaques (voir <https://attack.mitre.org>).
- 46 LinkAlong Sàrl, une entreprise Suisse fondée à l'École Polytechnique fédérale de Lausanne (l'une des principales institutions scientifiques et techniques européennes) associe l'intelligence artificielle avancée à l'expertise humaine pour repérer les signaux faibles, anticiper les risques et accélérer la prise de décisions éclairées.
- 47 Le sigle «URL» signifie «Uniform Resource Locator» (localisateur uniforme de ressources): il s'agit de l'adresse Web saisie dans un navigateur pour accéder à une page ou à une ressource spécifique sur Internet.
- 48 NewsGuard est une entreprise américaine qui emploie une équipe internationale de journalistes afin de produire des données sur la fiabilité de l'information à l'intention des consommateurs, des marques et des institutions. Elle tient à jour une base de données lisible par la machine qui répertorie les fausses informations circulant en ligne.



1030
1030

ETHIOPIAN RED CROSS SOCIETY