

ÉTUDE EXPLORATOIRE

**De la visibilité à la
vulnérabilité :
quand les réseaux
sociaux faillissent
aux défenseurs des
droits humains**

Mars 2026



RAPPORTEUSE SPÉCIALE
DE L'ONU SUR LES
DÉFENSEURS
DES DROITS
DE L'HOMME

SOLIDARITÉ ET ESPOIR

SOMMAIRE

Introduction	03
Méthodologie	04
Droit international et normes applicables	05
Défaillances de modération des contenus et perte de visibilité	06
Accession aux demandes répressives des gouvernements	11
Attaques en ligne et réponses insuffisantes des plateformes	13
Disparités de traitement selon les utilisateurs et les contenus	17
Difficultés à interagir avec les plateformes de réseaux sociaux	18
Conclusion	22
Recommandations	24

INTRODUCTION

Depuis leur création, les réseaux sociaux sont devenus des outils indispensables pour celles et ceux qui défendent les droits humains. Les gens y ont recours pour recueillir des informations, mettre en lumière les progrès et les défis, partager des rapports et de la documentation, développer et mobiliser des communautés et des réseaux, et relayer les voix des personnes marginalisées.

Cependant, dans le cadre de son travail et lors de consultations avec des défenseurs des droits humains, la Rapporteuse spéciale des Nations Unies sur la situation des défenseurs des droits humains (« la Rapporteuse spéciale ») a constaté que ces mêmes plateformes freinaient considérablement les activités de défense des droits humains. Qu'il s'agisse de méthodes de modération défailtantes, de censure algorithmique, de réponses inadéquates face au harcèlement en ligne ou de complaisance face aux demandes répressives de certains États, les plateformes de réseaux sociaux contribuent, intentionnellement ou non, à réduire au silence, à exclure et à mettre en danger les défenseurs des droits humains (DDH). [1]

Face à ces préoccupations, la Rapporteuse spéciale publie cette étude exploratoire examinant l'impact des réseaux sociaux sur le droit de défendre les droits humains. Ses constatations soulignent la nécessité urgente pour les entreprises gestionnaires de plateformes de réseaux sociaux d'assumer leurs responsabilités en matière de droits humains, et pour les États de respecter leurs obligations en la matière et de veiller à ce que les espaces numériques restent des environnements sûrs et propices pour celles et ceux qui défendent les droits humains.

[1] La Rapporteuse spéciale définit un défenseur des droits humains comme une personne qui, individuellement ou en association avec d'autres, œuvre pacifiquement à la promotion et la protection des droits humains conformément à la Déclaration sur les défenseurs des droits de l'homme.

MÉTHODOLOGIE

Cette étude exploratoire s'appuie sur des recherches et des consultations [2] menées par l'équipe de la Rapporteuse spéciale des Nations Unies sur les défenseurs des droits humains auprès d'organisations de la société civile qui travaillent aux côtés de personnes défendant et promouvant les droits humains partout dans le monde.

La portée de cette étude a été limitée par des contraintes de temps, de ressources et d'accès, et ne prétend pas être exhaustive. Elle rapporte toutefois les expériences des organisations qui ont pu contribuer au processus de consultation. Elle doit être considérée comme un aperçu indicatif des principales tendances et des risques, plutôt que comme une analyse exhaustive de l'ensemble des impacts de ces plateformes sur le droit de défendre les droits humains.

La Rapporteuse spéciale des Nations unies sur les défenseurs des droits de l'homme a fait part de ses préoccupations concernant les pratiques de [Meta Inc.](#) et de [X Corp.](#) et leur impact sur le droit de défendre les droits humains dans deux communications officielles adressées aux entreprises le 30 décembre 2025, les informant que ces communications et leurs réponses seraient incluses dans le présent document exploratoire. Aucune réponse n'avait été reçue à la date de publication du présent document.

[2] Les consultations se sont déroulées sous la règle de Chatham House.

1. DROIT INTERNATIONAL ET NORMES APPLICABLES

Adoptée par consensus lors de l'Assemblée générale des Nations Unies en décembre 1998, la Déclaration des Nations Unies sur les défenseurs des droits de l'homme [3] (« la Déclaration sur les DDH ») stipule à l'article 1 que « Chacun a le droit, individuellement ou en association avec d'autres, de promouvoir et de lutter pour la protection et la réalisation des droits de l'homme et des libertés fondamentales aux niveaux national et international. » Ce principe est renforcé à l'article 12.1, qui stipule que « Chacun a le droit, individuellement ou en association avec d'autres, de participer à des activités pacifiques contre les violations des droits de l'homme et des libertés fondamentales », tandis que l'article 6 dispose que chacun a le droit de partager des informations, des opinions et des connaissances sur l'ensemble des droits de l'homme, et d'attirer l'attention du public sur ces questions.

L'article 6 de la Déclaration sur les DDH s'appuie sur l'article 19 du Pacte international relatif aux droits civils et politiques, qui protège le droit à la liberté d'opinion et le droit à la liberté d'expression, définis comme « la liberté de rechercher, de recevoir et de diffuser des informations et des idées de toute espèce, sans considération de frontières, que ce soit oralement, par écrit ou sous forme imprimée, par les arts ou par tout autre moyen de son choix ». La liberté d'expression peut être restreinte en vertu de l'article 19, paragraphe 3, qui exige que les restrictions soient claires et précises (« prévues par la loi ») et nécessaires pour assurer « le respect des droits ou de la réputation d'autrui », « la protection de la sécurité nationale ou de l'ordre public », ou encore « la santé et la moralité publiques ».

En vertu du droit international relatif aux droits de l'homme, il incombe en premier lieu aux États de protéger les personnes contre les violations des droits humains commises sur leur territoire ou juridiction par des tiers, y compris les entreprises. Toutefois, selon les Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme, les entreprises ont elles-mêmes la responsabilité : a) d'éviter d'avoir des incidences négatives sur les droits de l'homme ou d'y contribuer par leurs propres activités ; et b) de prévenir ou d'atténuer les incidences négatives sur les droits de l'homme qui sont directement liées à leurs activités, produits ou services par leurs relations commerciales. Afin de s'acquitter de leur responsabilité en matière de respect des droits humains, les entreprises devraient disposer de : a) L'engagement politique de s'acquitter de leur

[3] A/RES/53/144, disponible en plusieurs langues à l'adresse suivante : <https://www.ohchr.org/en/special-procedures/sr-human-rights-defenders/declaration-human-rights-defenders-different-languages>

responsabilité en matière de respect des droits de l'homme; b) Une procédure de diligence raisonnable en matière de droits de l'homme pour identifier leurs incidences sur les droits de l'homme, prévenir ces incidences et en atténuer les effets, et rendre compte de la manière dont elles y remédient; et c) Des procédures permettant de remédier à toutes les incidences négatives sur les droits de l'homme qu'elles peuvent avoir ou auxquelles elles contribuent. La responsabilité de respecter les droits de l'homme constitue une norme de conduite générale applicable à toutes les entreprises et existe indépendamment de la capacité ou de la volonté des États de s'acquitter de leurs propres obligations en matière de droits de l'homme.

2. DÉFAILLANCES DE MODÉRATION DES CONTENUS ET PERTE DE VISIBILITÉ

Partout dans le monde, les défenseurs des droits humains font état de diverses défaillances en matière de modération des contenus sur les réseaux sociaux, avec de graves répercussions sur le droit de défendre les droits humains. Parmi ces défaillances figurent la suppression de contenus légitimes, le *shadow banning*, les blocages arbitraires, une application incohérente des standards de la communauté et, plus largement, un manque de transparence sur les processus internes.

2.1 SUPPRESSION DE CONTENUS LÉGITIMES RELATIFS AUX DROITS HUMAINS

Lors de la préparation de cette étude, plusieurs organisations nous ont fait part d'incidents au cours desquels des contenus légitimes relatifs aux droits humains auraient été supprimés au motif qu'ils enfreignaient les règles de la plateforme.

Une organisation a constaté que des publications relatives aux droits humains en **Palestine** avaient été supprimées par des plateformes appartenant à Meta en vertu de leurs politiques concernant les « individus et organisations dangereux » ou les « contenus violents », souvent sans explication claire ni possibilité de recours. Ces suppressions concernaient notamment des posts documentant des exécutions extrajudiciaires, des arrestations arbitraires et des manifestations pacifiques, et ont particulièrement affecté les défenseurs des droits humains et les journalistes, dont certains n'ont reçu aucune notification ni explication concernant cette censure. Des défenseurs des droits humains ont également signalé des incohérences : par exemple, des publications critiquant des contenus haineux ou extrémistes ont été supprimées alors que les contenus haineux ou

extrémistes originaux qu'elles dénonçaient sont restés en ligne [4]. Une organisation de défense des droits numériques a également régulièrement documenté des cas de censure et de suppression de contenus relatifs aux droits humains en Palestine par les plateformes de réseaux sociaux, notamment Facebook, X et Instagram [5]. De même, une autre organisation a signalé la suppression de contenus relatifs aux victimes de la guerre en **Syrie**, y compris des publications qui auraient pu contribuer aux efforts visant à obtenir justice au niveau international [6]. Les motifs généralement invoqués par les plateformes pour justifier la suppression de ces contenus incluent la glorification d'individus ou d'organisations dangereux, et le fait que ces contenus soient considérés comme violents ou explicites. Des photos de victimes, en particulier d'enfants, ont été signalées comme relevant de l'exploitation infantile.

Des cas similaires ont été observés dans d'autres régions et sur d'autres plateformes. Sur TikTok, des contenus liés aux violations des droits humains au **Xinjiang** et au **Tibet** ont été à plusieurs reprises signalés, supprimés ou restreints ces dernières années, au motif qu'ils enfreindraient les règles de la communauté [7][8]. Une organisation biélorusse de défense des droits humains a signalé que ses publications sur les victimes de la répression politique en **Biélorussie** étaient parfois signalées comme inappropriées ou à caractère politique sur Facebook et Instagram, les recours aboutissant à des résultats variables, sans véritable justification [9]. Au **Cambodge**, des journalistes ont vu leurs comptes Facebook suspendus après avoir publié des contenus sur les conflits fonciers, la corruption et la déforestation, et lors des élections nationales de 2023, plusieurs comptes de médias indépendants ont été bloqués, restreignant considérablement l'accès à une information indépendante [10]. Une organisation régionale de défense des droits humains a rapporté que l'une de ses publications sponsorisées portant sur les droits humains au sein de l'ASEAN avait été signalée comme inappropriée ou sensible et retirée des plateformes de réseaux sociaux aux **Philippines** et au **Cambodge**, sans justification claire et alors qu'elle avait été diffusée sans problème dans d'autres pays d'Asie [11].

[4] Contribution (A) reçue par la Rapporteuse spéciale

[5] Contribution (B) reçue par la Rapporteuse spéciale

[6] Contribution (C) reçue par la Rapporteuse spéciale

[7] *The Guardian*, "Revealed: how TikTok censors videos that do not please Beijing", septembre 2019 <https://www.theguardian.com/technology/2019/sep/25/revealed-how-tiktok-censors-videos-that-do-not-please-beijing>

[8] *Radio Free Asia*, "TikTok deletes videos related to Uyghur human rights violations", novembre 2024 <https://www.rfa.org/english/uyghur/2024/11/05/uyghur-tiktok-censors-abroad/>

[9] Contribution (D) reçue par la Rapporteuse spéciale

[10] Contribution (E) reçue par la Rapporteuse spéciale

[11] Contribution (F) reçue par la Rapporteuse spéciale

2.2 SHADOW BANNING ET RESTRICTIONS DE VISIBILITÉ

Le *shadow banning*, pratique consistant à rendre les publications d'un utilisateur invisibles aux yeux des autres utilisateurs sans avertissement préalable, est également fréquemment signalé par les défenseurs des droits humains.

Plusieurs organisations ont constaté que les comptes des défenseurs des droits humains palestiniens et les contenus liés aux droits humains en **Palestine** avaient fait l'objet d'un *shadow banning* notable sur les plateformes de Meta, Facebook et Instagram, ces pratiques s'étant intensifiées depuis le début du génocide perpétré par Israël [12], mais ayant également été signalées avant celui-ci [13]. Concrètement, cela signifie que les informations partagées par ces défenseurs des droits humains, qu'il s'agisse de documenter des violations des droits humains sur le terrain, de sensibiliser l'opinion publique internationale ou de collecter des fonds pour soutenir les victimes, sont vues par moins de personnes, ce qui nuit à l'efficacité de leur travail. Les utilisateurs concernés ne reçoivent aucune notification ni justification. De tels cas peuvent être difficiles à comprendre et à vérifier, car Meta fait preuve de peu de transparence quant à la mise en œuvre de ses algorithmes [14][15].

Des organisations engagées dans la défense de la liberté d'expression ont également constaté que le *shadow banning* touchait de manière disproportionnée les défenseuses des droits humains féministes au **Liban**, qui dénoncent la suppression de leurs publications, notamment celles rédigées en arabe, tandis que les contenus haineux visant les femmes restent en ligne [16][17]. Les défenseuses des droits humains engagées dans la défense des droits sexuels et reproductifs sont également concernées [18].

Dans un cas rapporté par l'une des organisations ayant participé à la consultation, un collectif de journalistes migrants au **Chili** a observé que certaines de ses publications Instagram consacrées à un événement antiraciste n'apparaissaient pas dans les fils d'actualité. Malgré des demandes d'assistance répétées, aucune explication claire n'a été fournie quant à la raison pour laquelle ces publications avaient été masquées, et leur visibilité n'a été rétablie qu'après de longs échanges. Ce manque de transparence a laissé

[12] Amnesty International, "Israel/OPT: Post-ceasefire: Israel's genocide in the occupied Gaza strip continues", 27 novembre 2025 <https://www.amnesty.org/en/documents/mde15/0527/2025/en/>

[13] SMEX, "Israel's Fierce Attack on Palestinian Narrative and Content", juin 2022 <https://smex.org/israels-fierce-attack-on-palestinian-narrative-and-content/>

[14] Contribution (A) reçue par la Rapporteuse spéciale

[15] Contribution (B) reçue par la Rapporteuse spéciale

[16] Contribution (B) reçue par la Rapporteuse spéciale

[17] SMEX, "Research Report: Confronting Structural Silencing: Challenges and Resistance Among Digital Feminist Activists in Lebanon", avril 2025 <https://smex.org/research-report-confronting-structural-silencing-challenges-and-resistance-among-digital-feminist-activists-in-lebanon/>

[18] Contribution (C) reçue par la Rapporteuse spéciale

le collectif dans l'incertitude quant aux motifs de ces restrictions et l'a dissuadé de publier davantage [19].

Le *shadow banning* a également été utilisé au détriment des personnes engagées dans la défense des droits LGBTIQ+. En 2024, des changements de politique [20] chez Meta ont limité via l'algorithme les « contenus politiques » sur Instagram et Threads, sans préciser exactement ce que cela recouvrait ni comment cette mesure serait appliquée. Dans les faits, ce changement a considérablement réduit la portée de publications crédibles et informatives de plusieurs comptes Instagram de défense des droits des personnes LGBTIQ+ [21], notamment ceux des organisations Human Rights Campaign et GLAAD aux **États-Unis**. Cette évolution est venue s'ajouter aux antécédents de Meta en matière de *shadow banning* de contenus LGBTIQ+, qui sont souvent signalés à tort comme étant sexuellement explicites [22].

2.3 DÉFAILLANCES DE LA MODÉRATION ALGORITHMIQUE ET AUTOMATISÉE

Un certain nombre des défaillances de modération mentionnées précédemment semblent résulter de l'utilisation de systèmes de modération automatisés ou basés sur l'intelligence artificielle (IA). Selon une ONG, ces technologies sont souvent trop peu avancées pour permettre de distinguer de manière fiable les contenus licites des contenus illicites, ce qui conduit à la suppression d'une grande quantité de contenus légitimes, y compris ceux publiés par des défenseurs des droits humains [23].

Une autre organisation nous a fait part d'un exemple où des filtres automatiques ont classé à tort des publications légitimes documentant la répression lors des élections de 2024 au **Venezuela** comme « incitation à la violence », ce qui a entraîné leur blocage [24]. Elle a également rapporté qu'un collectif de défense des droits reproductifs au Brésil avait vu son compte suspendu de manière définitive pour avoir partagé des informations sur le misoprostol, un médicament contre les ulcères utilisé depuis longtemps par les femmes

[19] Contribution (A) reçue par la Rapporteuse spéciale

[20] Instagram, "Continuing our Approach to Political Content on Instagram and Threads", février 2024 <https://about.instagram.com/blog/announcements/continuing-our-approach-to-political-content-on-instagram-and-threads>

[21] Accountable Tech, "Meta's Political Content Limit Causes Steep Drop in Reach for Accounts", juillet 2024 <https://accountabletech.org/research/metas-political-content-limit-causes-steep-drop-in-reach-for-accounts/>

[22] Human Rights Campaign, "Meta's New Policies: How They Endanger LGBTQ+ Communities and Our Tips for Staying Safe Online", janvier 2025 <https://www.hrc.org/news/metas-new-policies-how-they-endanger-lgbtq-communities-and-our-tips-for-staying-safe-online>

[23] Contribution (G) reçue par la Rapporteuse spéciale

[24] Contribution (A) reçue par la Rapporteuse spéciale

brésiliennes pour provoquer des avortements [25].

2.4 SUSPENSIONS DE COMPTES ET SIGNALEMENTS DE MASSE

Les personnes et les organisations de la société civile qui défendent les droits humains sont également confrontées à des suspensions de compte et à des restrictions provoquées par des campagnes coordonnées de signalements de masse menées par des trolls, des acteurs liés à l'État, ainsi que des groupes d'extrême droite et anti-droits.

En **Amérique centrale**, une organisation régionale a documenté l'utilisation de « bots » et de campagnes de signalements abusifs coordonnées, entraînant la suppression de contenus ou la perte de comptes de défenseuses des droits humains. Ces incidents surviennent souvent sans avertissement ni justification, et les procédures de recours sont opaques voire inaccessibles, en particulier pour les utilisateurs sans accès direct aux canaux de partenaires de confiance des plateformes [26]. Au **Bangladesh**, des défenseurs des droits humains féministes, *queer* et laïques ont été confrontés à des problèmes similaires lorsque des campagnes coordonnées de signalement de masse ont entraîné la suspension automatique de leurs comptes sur Facebook et Instagram, sans aucun moyen de recours significatif [27]. En **Iran**, une organisation a indiqué que les comptes Instagram de militantes iraniennes des droits des femmes avaient été désactivés ou soumis à un *shadow ban* à la suite de signalements massifs attribués aux autorités iraniennes [28]. En **Belgique**, un compte Instagram dédié à l'éducation antiraciste et géré par une jeune défenseuse des droits humains a été désactivé à sept reprises entre 2023 et 2025 en raison de signalements massifs émanant de groupes anti-droits, ce qui l'a finalement conduite à rendre le compte privé et à cesser d'y publier [29]. En **Malaisie**, une organisation rapporte que les comptes Facebook et Instagram d'une organisation de défense des droits des femmes et de deux de ses membres ont été suspendus sans explication à la suite d'une vague de harcèlement en ligne. Bien que leurs comptes aient été réactivés par la suite, cet incident a affecté leurs communications et l'organisation a dû procéder à un réexamen interne de sa stratégie numérique. [30]

[25] Contribution (A) reçue par la Rapporteuse spéciale

[26] Contribution (H) reçue par la Rapporteuse spéciale

[27] Contribution (A) reçue par la Rapporteuse spéciale

[28] Contribution (B) reçue par la Rapporteuse spéciale

[29] <https://www.instagram.com/p/DHvS29aomok/>

[30] Contribution (G) reçue par la Rapporteuse spéciale

3. ACCESSION AUX DEMANDES RÉPRESSIVES DES GOUVERNEMENTS

Tandis que les personnes qui utilisent les réseaux sociaux pour promouvoir et protéger les droits de humains doivent faire face aux ingérences de gouvernements cherchant à faire taire la dissidence et à contrôler les discours [31], certaines entreprises se plient fréquemment à des requêtes officielles de censure qui vont à l'encontre du droit international relatif aux droits humains ou facilitent la surveillance étatique.

3.1 ACCESSION AUX DEMANDES DE CENSURE

Dans les informations reçues par la Rapporteuse spéciale dans le cadre de la préparation de cette étude, plusieurs organisations ont fait état de cas où des gouvernements ont exercé des pressions sur les plateformes de réseaux sociaux afin qu'elles alignent leurs pratiques de modération des contenus sur des positions politiques spécifiques ou qu'elles censurent directement les défenseurs des droits humains. Ces pressions peuvent prendre la forme de menaces de limitation de débit (en réduisant la bande passante mise à disposition de la plateforme), de restrictions publicitaires ou de sanctions juridiques sous forme d'amendes.

En **Turquie**, où les réseaux sociaux sont devenus l'une des sources principales d'information et un outil essentiel pour le journalisme indépendant, le gouvernement a, ces dernières années, renforcé son emprise sur les plateformes par le biais d'une série d'amendements à la Loi Internet. Une organisation nous a rapporté des taux de conformité très élevés aux demandes de retrait de contenu émanant du gouvernement turc sur les principales plateformes de réseaux sociaux : 95 % pour TikTok, 85 % pour X et 79 % pour Instagram au second semestre 2024 [32]. Lors de l'arrestation du leader de l'opposition Ekrem İmamoğlu en mars 2025, le gouvernement a ordonné le blocage de centaines de comptes, notamment ceux de défenseurs des droits humains, de journalistes, de groupes d'étudiants, d'organisations de la société civile et de groupes de défense des droits des femmes. Tandis que X s'est conformé à bon nombre de ces demandes de blocage sous la menace d'une restriction de son débit, Meta a déclaré n'avoir pris aucune mesure [33] et aurait écopé d'une amende considérable [34].

[31] Voir [A/HRC/38/35](#), paragraphe 13; [A/80/341](#), paragraphes 12 et 76

[32] Contribution (G) reçue par la Rapporteuse spéciale

[33] Meta, "Case Studies - Content alleged to violate Turkish Law", mars 2025
<https://transparency.meta.com/reports/content-restrictions/case-studies>

[34] Associated Press, "Meta is fined by Turkey after refusing to restrict content on Facebook and Instagram" avril 2025 <https://apnews.com/article/meta-facebook-instagram-turkey-erdogan-59f16571e884fd6f0f9597a72771131f>

L'organisation a en outre signalé que Meta [35], TikTok [36], YouTube [37] et X [38] avaient tous été contraints de nommer des représentants locaux sous la menace de représailles juridiques croissantes, et ce en dépit des avertissements selon lesquels cette mesure permettrait au gouvernement turc d'exercer un contrôle accru sur leurs plateformes [39].

En Inde, *Bolta Hindustan*, un média indépendant consacré aux questions sociales et comptant 300 000 abonnés, a vu sa chaîne YouTube suspendue en avril 2024 à la suite d'une notification du ministère de l'Information et de la Radiodiffusion. Dans un premier temps, le service juridique de Google n'a fourni aucune justification de cette suspension au média [40]. Les défenseurs de la liberté de la presse ont dénoncé ce blocage, et le compte a par la suite été réactivé par YouTube [41]. De même, Twitter (aujourd'hui X) a été accusé de censurer les voix des Cachemiris à la demande du gouvernement indien. En 2019, le Comité pour la protection des journalistes (CPJ) a rapporté que Twitter avait empêché les utilisateurs basés en Inde d'accéder à des milliers de tweets partageant des informations sur la situation au **Jammu-et-Cachemire** pendant les black-out médiatiques et les coupures d'Internet, à la suite de mises en demeure envoyées par le gouvernement indien [42].

Au **Pakistan**, des diffusions en live de manifestations baloutches sur Facebook et YouTube ont fait l'objet de ralentissements inexplicables de la connexion Internet ou ont été supprimées sans motif [43].

3.2 FACILITATION DE LA SURVEILLANCE ÉTATIQUE

Le manque de transparence et de contrôle sur les données personnelles partagées par les défenseurs des droits humains sur leurs comptes de réseaux sociaux constitue une autre

[35] Facebook, "An Update on Facebook in Turkey", janvier 2021

<https://about.fb.com/news/2021/01/an-update-on-facebook-in-turkey/>

[36] TikTok, "An update on TikTok in Turkey", janvier 2021

<https://newsroom.tiktok.com/an-update-on-tiktok-in-turkey?lang=en>

[37] YouTube, "An update on YouTube in Turkey", décembre 2020

<https://blog.youtube/news-and-events/update-youtube-turkey/>

[38] *Balkan Insight*, "X Appoints Representative in Turkey, Bowing to Govt Pressure", mai 2024

<https://balkaninsight.com/2024/05/23/x-appoints-representative-in-turkey-bowing-to-govt-pressure/>

[39] Contribution (G) reçue par la Rapporteuse spéciale

[40] Contribution (F) reçue par la Rapporteuse spéciale

[41] *Medianama*, "YouTube Reinstates Blocked and Demonetised Channels", mai 2024

<https://www.medianama.com/2024/05/223-youtube-reinstates-blocked-demonetised-channels/>

[42] *Committee to Protect Journalists*, "India uses opaque legal process to suppress Kashmiri journalism, commentary on Twitter", octobre 2018

<https://cpj.org/2019/10/india-opaque-legal-process-suppress-kashmir-twitter/>

[43] Contribution (F) reçue par la Rapporteuse spéciale

source de préoccupation. Dans des cas recensés par l'une des organisations participant à cette étude, des données d'utilisateurs appartenant à des défenseurs des droits humains auraient été transmises aux autorités publiques à leur insu et sans leur consentement. Au **Guatemala**, par exemple, Meta a partagé les informations d'un défenseur des droits humains, qui concernaient non seulement le compte requis par les autorités, mais l'ensemble de ses comptes sur les réseaux sociaux. Le défenseur n'en a jamais été officiellement informé et ne l'a appris qu'après avoir été averti par une source fiable [44]. Cela soulève de sérieuses préoccupations, en particulier dans les contextes où les gouvernements visent délibérément les voix dissidentes. Les procédures par lesquelles les plateformes évaluent et répondent aux demandes des pouvoirs publics nationaux demeurent opaques. Par conséquent, les personnes qui défendent les droits humains sont susceptibles d'être soumises à de la surveillance, à des intimidations ou à des représailles sans savoir que leurs données privées ont été consultées ou divulguées, et sans avoir accès à aucun recours effectif.

4. ATTAQUES EN LIGNE ET RÉPONSES INSUFFISANTES DES PLATEFORMES

Les réseaux sociaux ont reproduit, sous diverses formes numériques, les schémas de harcèlement et d'abus auxquels de nombreux défenseurs des droits humains sont déjà confrontés dans le cadre de leur travail. Les violences en ligne sont devenues à la fois plus fréquentes et mieux coordonnées. Les défenseurs des droits humains sont régulièrement l'objet de campagnes de harcèlement, de diffamation et de désinformation relayées sur les réseaux sociaux, souvent à l'aide de *bots* automatisés [45]. Ces tactiques, parfois assistées par des outils d'intelligence artificielle, sont de plus en plus courantes. Face à cette situation, les plateformes n'ont pas su mettre en place des mécanismes de réponse adéquats pour protéger les utilisateurs victimes de ces attaques.

Pour illustrer l'ampleur du problème, *Global Witness* a souligné dans un rapport récent que 92 % des défenseurs des droits fonciers et environnementaux interrogés dans le cadre de

[44] Contribution (A) reçue par la Rapporteuse spéciale

[45] Résolution 58/23 adoptée par le Conseil des droits de l'homme le 4 avril 2005 <https://docs.un.org/en/A/HRC/RES/58/23> ; UN Women, "Tipping point: The chilling escalation of online violence against women in the public sphere", 2025, <https://www.unwomen.org/en/digital-library/publications/2025/12/tipping-point-the-chilling-escalation-of-violence-against-women-in-the-public-sphere-in-the-age-of-ai> ; Front Line Defenders, "Global Analysis 2024/2025", https://www.frontlinedefenders.org/sites/default/files/1609_fld_ga24-5_output.pdf

leur enquête [46] avaient fait l'objet de harcèlement en ligne de la part de divers acteurs, dont des particuliers, des bots et des campagnes coordonnées. Selon cette étude, à l'échelle mondiale, c'est sur Facebook que les défenseurs des droits humains subiraient le plus de harcèlement, suivi de X (Twitter) et d'Instagram [47].

4.1 HARCÈLEMENT EN LIGNE FONDÉ SUR LE GENRE ET L'IDENTITÉ

Les individus subissant déjà une forme ou une autre d'oppression sont généralement exposés à des risques plus importants et à des violations plus graves lorsqu'ils utilisent les réseaux sociaux pour défendre et promouvoir les droits humains. A travers le monde, les défenseurs des droits humains, en particulier les femmes et les défenseurs des droits des personnes LGBTIQ+, sont confrontés à la désinformation, aux campagnes de dénigrement, à la diffamation, aux *deepfakes*, au chantage, aux menaces, au harcèlement, au *doxxing* et à l'usurpation d'identité, des pratiques souvent ignorées ou insuffisamment prises en compte par les plateformes. Ces attaques visent à délégitimer, intimider, discréditer et ridiculiser les défenseurs et leur travail.

Une organisation en Amérique centrale a constaté que les attaques en ligne représentaient environ un quart de l'ensemble des attaques contre les défenseuses des droits humains recensées entre 2020 et juin 2025 dans la région, un chiffre qui n'a eu de cesse d'augmenter. Au **Salvador**, les attaques en ligne représentaient plus de la moitié de l'ensemble des attaques contre les défenseuses des droits humains. Près de 90 % de ces attaques ont été perpétrées sur les réseaux sociaux [48].

Une autre organisation a rapporté que la défenseuse **syrienne** des droits humains Hiba Ezzideen Al-Hajji avait été victime d'une campagne de diffamation sur Facebook visant à la discréditer, ainsi que sa famille et l'organisation dont elle fait partie, à la suite d'une publication dénonçant les mariages forcés. Elle aurait reçu de nombreuses menaces de mort à travers une avalanche de messages sur son propre compte, ainsi que sur la page de son organisation, appelant à des violences physiques à son encontre [49][50]. Quelques années auparavant, elle avait déjà été victime de diffamation et de menaces sur la plateforme Telegram [51].

Au **Bangladesh**, une organisation a souligné que la transition politique de 2024 avait

[46] Global Witness, "Toxic platforms, broken planet", 16 juillet 2025

<https://globalwitness.org/en/campaigns/digital-threats/toxic-platforms-broken-planet/>

[47] Global Witness, Ibid

[48] Contribution (H) reçue par la Rapporteuse spéciale

[49] Contribution (A) reçue par la Rapporteuse spéciale

[50] Voir [AL SYR 3/2025](#)

[51] Voir [AL OTH 129/2023](#), à laquelle les titulaires de mandat au titre des procédures spéciales n'ont reçu aucune réponse de Telegram

marqué un tournant dans les pratiques de harcèlement en ligne : d'incidents isolés, il s'est transformé en campagnes coordonnées, accompagnées d'une forte augmentation des cas de *doxing*, de diffamation par l'image, de discours de haine sous forme de mèmes, de signalements massifs des comptes de défenseurs sur Facebook et Instagram, ainsi que de la diffusion de fausses informations les visant. Ces attaques comportent souvent des insultes à caractère sexiste ou religieux, visant en particulier les féministes, les défenseurs des droits humains laïques et les défenseurs des droits humains LGBTQI+ [52].

Plusieurs organisations de défense des droits humains ont signalé que l'utilisation croissante de l'IA pour générer des *deepfakes* et des images manipulées avait accru la vulnérabilité des défenseurs des droits humains, en particulier des femmes, sur Internet [53][54][55]. C'est notamment le cas au **Venezuela**, où la défenseuse des droits humains Martha Lía Grajales a été la cible d'une campagne de diffamation reposant sur une vidéo générée par l'IA, qui prétendait à tort que son travail en faveur des droits humains cachait un complot contre le gouvernement. La vidéo a été largement partagée sur les réseaux sociaux et WhatsApp, notamment à l'aide de milices numériques spécialisées dans la mise en œuvre et l'amplification de ce type d'attaques. Une fois disséminé, ce type de contenu est difficile à contrer ou à supprimer [56].

Dans tous les exemples cités ci-dessus, les plateformes auraient tardé à réagir ou se seraient montrées inefficaces face à ces préjudices.

Plusieurs contributions ont souligné le rôle des plateformes de réseaux sociaux dans l'alimentation et l'amplification de contenus malveillants, notamment en raison de leurs algorithmes qui privilégient les contenus sensationnels, clivants et violents, augmentant ainsi la portée des harceleurs et des agresseurs [57][58][59]. L'une des organisations ayant participé à l'étude s'inquiète du fait que les algorithmes basés sur le profit actuellement utilisés par les principales plateformes de réseaux sociaux contribuent directement aux violences en ligne contre les femmes défenseuses des droits humains en **Amérique centrale**, qui se traduisent souvent par des violences physiques [60]. Selon l'enquête susmentionnée menée par *Global Witness*, 84 % des défenseurs de l'environnement en **Afrique**, en **Amérique latine** et en **Asie** qui ont déclaré avoir été pris pour cible en raison de leur activisme estiment que les violences en ligne ont contribué « directement » ou

[52] Contribution (A) reçue par la Rapporteuse spéciale

[53] Contribution (H) reçue par la Rapporteuse spéciale

[54] Contribution (J) reçue par la Rapporteuse spéciale

[55] Contribution (A) reçue par la Rapporteuse spéciale

[56] Contribution (A) reçue par la Rapporteuse spéciale

[57] Contribution (I) reçue par la Rapporteuse spéciale

[58] Contribution (H) reçue par la Rapporteuse spéciale

[59] Contribution (C) reçue par la Rapporteuse spéciale

[60] Contributions (B), (F), (H) et (J) reçues par la Rapporteuse spéciale

« dans une certaine mesure » aux violences hors ligne [61].

4.2 ABSENCE DE DISPOSITIFS EFFICACES DE TRAITEMENT DES PLAINTES

Dans l'ensemble des contributions reçues, les organisations ont souligné l'absence de mécanisme de plainte réactif, transparent ou à dimension humaine lorsque les défenseurs des droits humains tentent de signaler une violation. L'enquête de Global Witness a révélé que seuls 12 % des défenseurs des droits humains ayant signalé des abus et actes de harcèlement aux plateformes de réseaux sociaux se sont déclarés satisfaits de la réponse qui leur a été apportée [62]. Selon l'une des organisations ayant répondu à notre étude, certaines plateformes telles que TikTok et Telegram ne répondent jamais aux demandes d'assistance émanant des organisations spécialisées en **Amérique centrale**. Par ailleurs, de nombreuses procédures exigent une maîtrise de l'anglais et sont traitées par des bots ou des réponses automatisées [63]. Au **Bangladesh**, une organisation a constaté que plus de 90 % des signalements adressés à Facebook concernant le harcèlement en ligne de défenseurs des droits humains ont été soit ignorés, soit ont donné lieu à des mesures de représailles contre le plaignant plutôt que contre l'auteur des faits, y compris la suspension de comptes [64]. Une défenseuse des droits humains au **Liberia** a quant à elle souligné que l'absence de réponse des plateformes de réseaux sociaux face aux attaques en ligne contre les défenseuses des droits humains promouvant la participation politique des femmes dans le pays avait, dans certains cas, conduit à des agressions physiques à leur rencontre [65]. L'une des organisations, qui consacre du temps et des ressources à la gestion d'un service d'assistance en matière de sécurité numérique destiné aux défenseurs des droits humains, a indiqué que la plupart des demandes qu'elle traite finissent par être résolues. Cependant, le délai de réponse des plateformes, qui peut aller jusqu'à 14 jours, ainsi que l'absence occasionnelle de réponse ou les résolutions défavorables, contribuent à mettre davantage en danger les défenseurs des droits humains. En outre, il est important de noter que la plupart des défenseurs des droits humains n'ont pas accès à un tel niveau de soutien et de ressources, ce qui se traduit par un nombre relativement plus faible de résolutions positives dans leurs interactions avec les plateformes de réseaux sociaux. Des inquiétudes ont également été soulevées concernant des cas d'usurpation d'identité ou de harcèlement via les messageries instantanées, qui sont des problèmes courants pour les femmes défenseuses des droits humains au **Moyen-Orient** et en **Afrique du Nord** et peuvent entraîner de graves préjudices. Pourtant, l'organisation rapporte que les plateformes parviennent rarement à résoudre ces problèmes, se contentant de demander aux utilisateurs de remplir un formulaire sans résultat tangible [66].

[61] Global Witness, "Toxic platforms, broken planet", 16 juillet 2025

[62] Global Witness, Ibid

[63] Contribution (H) reçue par la Rapporteuse spéciale

[64] Contribution (A) reçue par la Rapporteuse spéciale

[65] Contribution (K) reçue par la Rapporteuse spéciale

[66] Contribution (C) reçue par la Rapporteuse spéciale

5. DISPARITÉS DE TRAITEMENT SELON LES UTILISATEURS ET LES CONTENUS

Plusieurs organisations ont identifié des disparités dans la manière dont les plateformes de réseaux sociaux traitent les défenseurs des droits humains et les contenus qu'ils publient, notamment en fonction de leur langue et de leur région [67].

5.1 DISPARITÉS LINGUISTIQUES

Le manque de compétences linguistiques suffisantes au sein des équipes et des outils de modération des contenus constitue une problématique majeure. Cela est particulièrement flagrant dans la modération des contenus rédigés dans des langues autres que l'anglais, le français ou l'espagnol, où les contenus malveillants ou haineux passent souvent inaperçus et ne sont pas traités.

Une étude menée par une organisation sur le harcèlement en ligne au **Bangladesh** a démontré que des plateformes telles que Facebook ne parviennent pas à détecter ni à traiter les contenus injurieux en bengali. Malgré des signalements répétés, les comptes malveillants restent souvent actifs, et les termes haineux tels que « Shahbagi » [68], utilisés de manière péjorative à l'encontre des défenseurs des droits humains et des militants, ne sont ni signalés ni supprimés [69]. En parallèle, la modération excessive des contenus dans d'autres langues engendre une dynamique différente, mais tout aussi néfaste. Plusieurs organisations ont observé que les publications en arabe, en particulier celles liées à la **Palestine**, à la **Syrie** et plus largement aux droits humains dans la région, font l'objet d'une modération plus stricte que les publications dans d'autres langues [70].

Dans l'ensemble, les contenus dans des langues autres que l'anglais, le français ou l'espagnol ont tendance à être supprimés ou signalés de manière disproportionnée, car les plateformes de réseaux sociaux manquent de capacités de modération dans les langues locales [71]. Une organisation de défense des droits humains au **Cambodge** a soulevé des préoccupations similaires, notant que les suppressions de contenus en khmer n'étaient pas accompagnées d'explications et qu'il n'existait aucune infrastructure de soutien pour les

[67] Contributions (A), (B), (C), (F) et (I) reçues par la Rapporteuse spéciale

[68] Terme apparu lors des manifestations de Shahbag en 2013, utilisé comme argot péjoratif pour désigner les personnes de gauche, libérales ou laïques, sous-entendant souvent de l'islamophobie et une adhésion au tribunal médiatique et à la justice populaire.

[69] Contribution (A) reçue par la Rapporteuse spéciale

[70] Contributions (A), (B) et (C) reçues par la Rapporteuse spéciale

[71] Contributions (A), (B), (C) et (E) reçues par la Rapporteuse spéciale

défenseurs des droits humains qui ne parlent que le khmer [72]. Selon une autre organisation, la langue ourdoue constitue un autre « angle mort » pour les plateformes lorsqu'elle est utilisée pour des discours de haine et les menaces à l'encontre des défenseurs des droits humains au **Pakistan** [73]. En **Iran**, le processus de modération des contenus en persan sur Instagram a également mis en évidence des lacunes dans l'application des politiques de modération de la plateforme pour les langues non occidentales [74].

5.2 DISPARITÉS RÉGIONALES

Plusieurs organisations ont également signalé d'importantes disparités régionales dans la manière dont les défenseurs des droits humains et leurs contenus sont traités par les plateformes de réseaux sociaux [75]. D'après des recherches menées par l'une des organisations participantes à la présente étude, alors que le nombre de signalements d'atteintes à l'encontre des défenseurs des droits de l'homme envoyés aux plateformes est relativement homogène d'une région du monde à l'autre, les défenseurs en **Afrique** avaient moins de chances d'obtenir une réponse à leur plainte que leurs homologues européens (50 % contre 72 % respectivement) [76]. Une autre organisation a mis en évidence des disparités similaires dans la manière dont les plateformes traitent les cas de harcèlement à l'encontre des défenseurs des droits humains selon les régions [77].

6. DIFFICULTÉS À INTERAGIR AVEC LES PLATEFORMES DE RÉSEAUX SOCIAUX

6.1 DÉNOUEMENTS DÉCEVANTS POUR LES DÉFENSEURS DES DROITS HUMAINS

Face aux restrictions, violations et biais susmentionnés, les défenseurs des droits humains eux-mêmes disposent d'un accès très limité à des voies de recours.

[72] Contribution (E) reçue par la Rapporteuse spéciale

[73] Contribution (B) reçue par la Rapporteuse spéciale

[74] Article 19, "Iran: Meta must overhaul Persian-language content moderation on Instagram", juin 2022 <https://www.article19.org/resources/iran-meta-persian-language-content-moderation-instagram/>

[75] Contributions (A), (B), (F) et (I) reçues par la Rapporteuse spéciale

[76] Contribution (I) reçue par la Rapporteuse spéciale

[77] Contribution (A) reçue par la Rapporteuse spéciale

Les recherches menées par une organisation montrent qu'en l'absence d'un soutien spécifique de la société civile pour accompagner les défenseurs des droits humains dans leurs plaintes auprès des plateformes de réseaux sociaux, l'efficacité de ces dernières reste très limitée. Les plateformes telles que Meta disposent des standards de communauté qui, souvent, ne reconnaissent pas de nombreux types de violences en ligne comme allant à l'encontre de leurs règlements, en particulier lorsque le contenu préjudiciable est présenté comme relevant du débat public ou de la liberté d'expression. Par conséquent, de nombreux signalements de harcèlement sont ignorés ou rejetés, laissant les défenseurs des droits humains sans recours significatif. Même lorsque les plateformes prennent des mesures, la réponse est souvent lente et ne correspond pas à la gravité de l'abus [78].

Une autre organisation a identifié que les principaux obstacles à l'accès des défenseuses des droits humains en **Amérique centrale** à des recours en cas d'attaques en ligne étaient l'absence de mécanismes adéquats, les difficultés pour y accéder, en particulier pour celles qui ne parlent pas anglais, les longs délais d'attente pour obtenir une réponse, l'absence de suivi, et le fait que les réponses étaient souvent automatisées et gérées par des *bots*, de sorte qu'il est pratiquement impossible d'obtenir une réponse humaine [79]. Par conséquent, certaines défenseuses choisissent de ne pas signaler les abus dont elles sont victimes. 65 % des défenseuses des droits humains ayant rapporté des attaques numériques à l'organisation en question ne les ont pas signalées aux plateformes elles-mêmes, principalement à cause de l'inefficacité de leurs mécanismes [80]. Les recherches menées par une autre organisation sur le harcèlement en ligne au **Bangladesh** ont révélé une tendance similaire : les défenseurs des droits humains interrogés ont rapporté que les mécanismes de signalement de Facebook étaient souvent complexes, opaques et décourageants, ce qui a conduit nombre d'entre eux à se désillusionner et à renoncer à déposer des plaintes officielles [81].

Dans certains cas impliquant des menaces directes ou l'utilisation non autorisée de l'image des défenseurs, les plateformes ont parfois réagi en supprimant des publications ou en désactivant les profils incriminés. Cependant, de telles interventions sont rares et inconsistantes. De même, les résolutions favorables sont rares dans le cadre des recours concernant des blocages ou des suspensions de compte.

L'opacité des procédures de signalement et de recours, le manque de transparence ainsi que l'absence d'interlocuteurs humains contribuent à ce que les défenseurs des droits humains obtiennent rarement les réponses et les résolutions qu'ils recherchent. Cela conduit de nombreux défenseurs à solliciter le soutien d'organisations de la société civile tierces pour tenter d'obtenir des résultats.

[78] Contribution (A) reçue par la Rapporteuse spéciale

[79] Contribution (H) reçue par la Rapporteuse spéciale

[80] Contribution (H) reçue par la Rapporteuse spéciale

[81] Contribution (A) reçue par la Rapporteuse spéciale

6.2 LA SOCIÉTÉ CIVILE COMME INTERMÉDIAIRE : UN MODÈLE IMPARFAIT ET INTENABLE

Certaines organisations de défense des droits humains établies maintiennent des canaux de communication directs avec les plateformes de réseaux sociaux afin de leur signaler des cas et d'apporter leur soutien aux défenseurs des droits humains. Si ces canaux ont permis d'aboutir à davantage de résolutions favorables et d'accélérer les processus, les organisations de la société civile les trouvent souvent peu fiables ou irréguliers [82]. Une organisation rapporte que ses échanges avec les plateformes de réseaux sociaux pour le compte des défenseurs s'avèrent souvent infructueux et ne mènent à aucun changement systémique significatif [83]. De même, une autre organisation confirme que les résultats ne sont jamais garantis. Lorsqu'elle a tenté d'obtenir la réactivation du compte d'un défenseur des droits humains **kazakh** en établissant une ligne de communication directe entre ce dernier et Meta, aucune suite n'a été donnée à la demande malgré des tentatives répétées [84].

Même des initiatives telles que le programme « Trusted Partner » de Meta, créé pour favoriser la collaboration entre la plateforme et les organisations de la société civile, ont produit des résultats inégaux. Une organisation a indiqué que, dans les cas où le contenu de défenseurs des droits humains était supprimé ou signalé, elle tentait d'assurer un suivi via ces canaux d'assistance officiels. Si les réponses provenant du canal « Trusted Partner » de Meta étaient plus fréquentes que celles de X ou de Telegram, elles consistaient souvent en des réponses automatisées dépourvues de justification substantielle. Même lorsque des explications étaient fournies, elles reconnaissaient ou comprenaient rarement le contexte général de la situation, et le fait de faire remonter le problème n'a jamais garanti sa résolution [85].

Dans l'ensemble, le dialogue entre les organisations de la société civile et les plateformes semble caractérisé par un manque de constance et de transparence ainsi que par un désengagement croissant de la part des entreprises. Malgré cela, les ONG s'efforcent de combler les lacunes laissées par les plateformes et assument de plus en plus un rôle d'intermédiaire, en documentant les préjudices subis, en dialoguant avec les plateformes et en défendant les intérêts des utilisateurs [86]. Il ne s'agit toutefois pas d'un modèle viable, car il fait peser la charge de la protection et de l'assistance sur la société civile, tandis qu'au sein des plateformes, l'absence de mécanismes accessibles et efficaces pour les défenseurs des droits humains en danger persiste [87].

[82] Contributions (A), (B), (C), (F) et (I) reçues par la Rapporteuse spéciale

[83] Contribution (I) reçue par la Rapporteuse spéciale

[84] Contribution (B) reçue par la Rapporteuse spéciale

[85] Contribution (A) reçue par la Rapporteuse spéciale

[86] Contributions (A), (B) et (C) reçues par la Rapporteuse spéciale

[87] Contribution (A) reçue par la Rapporteuse spéciale

6.3 ÉROSION DES ENGAGEMENTS DES PLATEFORMES EN FAVEUR DES DROITS HUMAINS

Ces dernières années, dans ce qui ne peut être considéré que comme un signe de l'affaiblissement de leur engagement en faveur des droits humains, les principales entreprises propriétaires de réseaux sociaux ont commencé à réduire progressivement leur collaboration avec les défenseurs des droits humains et les ONG.

Malgré de nombreuses années de collaboration directe avec les plateformes de réseaux sociaux, à la fois de manière informelle et en tant que membre officiel de divers partenariats, l'une des organisations ayant participé à cette étude nous a indiqué que cette collaboration était devenue de moins en moins productive, avec des délais de réponse s'étant considérablement allongés en 2025. Des contenus malveillants restaient parfois en ligne pendant 10 à 15 jours après avoir été signalés, même dans des cas de menaces graves à l'encontre de défenseuses des droits humains au **Moyen-Orient** et en **Afrique du Nord**. Il a également été signalé que les plateformes étaient bien moins enclines à réagir aux pressions publiques qu'auparavant [88]. D'autres organisations de la société civile ont confirmé que l'engagement des plateformes en faveur des droits humains avait considérablement diminué au cours des trois dernières années, avec un intérêt décroissant à l'égard des contributions de la société civile [89].

Ces changements sont survenus parallèlement au recours par les entreprises de réseaux sociaux à des licenciements massifs de personnel clé et d'équipes consacrées à l'intégrité, à la sécurité des utilisateurs, aux droits humains et à la lutte contre la désinformation, ainsi qu'au démantèlement d'organes et de programmes consultatifs axés sur ces questions. Parmi les plateformes concernées figurent X [90], Meta [91], YouTube [92] et TikTok [93]. De tels changements structurels ont considérablement affaibli les voies de dialogue dont disposaient auparavant les défenseurs des droits humains et les organisations de la société

[88] Contribution (C) reçue par la Rapporteuse spéciale

[89] Contributions (I) et (J) reçues par la Rapporteuse spéciale

[90] *Amnesty International*, "Global: Twitter's decision to disband safety council threatens wellbeing of users", décembre 2022 <https://www.amnesty.org/en/latest/news/2022/12/global-twitters-decision-to-disband-safety-council-threatens-wellbeing-of-users/>; *Electronic Frontier Foundation*, "Monetization, Not Human Rights or Vulnerable Communities, Matter Most at Twitter Under Musk", novembre 2022 <https://www.eff.org/deeplinks/2022/11/twitters-monetizable-users-not-human-rights-matter-most-under-musks-rein-leaving>; Contribution (F) reçue par la Rapporteuse spéciale

[91] *Free Press*, "Big Tech Backslide: How Social-Media Rollbacks Endanger Democracy Ahead of the 2024 Elections", décembre 2023 <https://www.freepress.net/big-tech-backslide-how-social-media-rollbacks-endanger-democracy-ahead-2024-elections>; Contribution (F) reçue par la Rapporteuse spéciale

[92] *Free Press*, *ibid*

[93] *Business and Human Rights Centre*, "Germany: TikTok content moderators strike over mass layoffs & plans to replace them with AI", août 2025 <https://www.business-humanrights.org/en/latest-news/deutschland-content-moderatorinnen-protestieren-gegen-stellenabbau-geplante-ersetzung-durch-ki-algorithmus/>

civile, et ont compromis la capacité des plateformes à traiter des cas individuels ainsi que les problèmes structurels [94].

Dans le même temps, les plateformes de réseaux sociaux les plus largement utilisées ont abrogé des politiques essentielles qui contribuaient à la protection des droits des utilisateurs. Meta, X et YouTube ont toutes affaibli les mesures de protection de la vie privée des utilisateurs afin de permettre l'entraînement de leurs outils d'IA. Meta a revu à la baisse le recours au *fact-checking* par des tiers au profit d'une modération dite « communautaire » ou automatisée, ce qui soulève des inquiétudes quant à la propagation de la désinformation. X et YouTube ont également assoupli leurs politiques relatives à la désinformation électorale [95].

Bien que Meta publie depuis 2022 des rapports annuels décrivant en détail la manière dont l'entreprise identifie et atténue les risques liés aux droits humains sur ses plateformes, son dernier rapport [96] ne traite pas réellement des défis susmentionnés. Les autres plateformes principales, telles que X, TikTok ou YouTube, ne publient pas de rapports dédiés et spécifiques sur les droits humains.

7. CONCLUSION : RÉPERCUSSIONS SUR LA PRÉSENCE EN LIGNE DES DÉFENSEURS ET LEUR STRATÉGIE ORGANISATIONNELLE

L'effet cumulatif des différentes formes de répression digitale mentionnées dans cette étude se traduit par un rétrécissement de l'espace civique en ligne, les défenseurs des droits humains étant progressivement réduits au silence sur Internet. Confrontés à des restrictions de visibilité, au harcèlement ou à la surveillance via les réseaux sociaux, et en l'absence de soutien efficace ou adéquat de la part des plateformes pour remédier à ces violations, un grand nombre de défenseurs des droits humains font état d'un climat de peur et d'anxiété pour eux-mêmes et leurs communautés, tel que le démontrent les contributions reçues dans le cadre de la présente étude [97] ainsi que les échanges de la Rapporteuse spéciale avec les défenseurs des droits humains au cours de son mandat.

[94] Contribution (C) reçue par la Rapporteuse spéciale

[95] *Free Press*, ibid

[96] Meta, "2024 Human Rights Report", décembre 2025 <https://humanrights.fb.com/wp-content/uploads/2025/12/2024-Meta-Human-Rights-Report.pdf>

[97] Contributions (A), (H), (I) et (J) reçues par la Rapporteuse spéciale

Cela conduit parfois les défenseurs à choisir de limiter leur visibilité en ligne, d'éviter certains sujets sensibles ou de quitter purement et simplement les réseaux. Au **Salvador**, des défenseuses des droits humains ont renoncé à s'exprimer publiquement en raison du cyber-harcèlement continu dont elles faisaient l'objet [98]. Au **Pakistan**, une étude réalisée en 2023 par *Media Matters for Democracy* a révélé que 8 journalistes femmes sur 10 s'autocensuraient en ligne par crainte d'attaques numériques et physiques [99]. Au **Cambodge**, une organisation a constaté qu'un tiers des journalistes évitaient désormais de publier sur les réseaux sociaux des contenus critiques à l'égard des autorités [100]. Ailleurs en **Asie**, le recours à l'autocensure sur les sujets politiquement sensibles demeure courant [101]. Près de la moitié des défenseurs des droits fonciers et environnementaux interrogés par *Global Witness* ont indiqué que les abus dont ils avaient été victimes en ligne avaient entraîné une perte de productivité et une diminution de leur capacité à mener des campagnes sur ces questions [102].

Tout cela se traduit globalement par une diminution de la présence en ligne et de l'engagement public des défenseurs des droits humains, en particulier celles et ceux issus de communautés marginalisées et/ou travaillant sur des questions sensibles. Cette situation contribue à façonner un écosystème digital renforçant les rapports de force existants, au risque de faire disparaître complètement des espaces en ligne des communautés déjà surexposées aux menaces.

Certains défenseurs et organisations des droits humains explorent des plateformes alternatives, plus autonomes, où la sûreté et la confidentialité priment sur la portée. D'autres ont adapté leur stratégie en matière de présence sur les réseaux sociaux, en décidant notamment de limiter leur empreinte numérique, d'utiliser des pseudonymes ou un langage codé et d'éviter les publications en temps réel [103]. Cependant, les plateformes de réseaux sociaux grand public restent essentielles pour la plupart des défenseurs. Les nouvelles plateformes ont encore une portée très limitée et comptent beaucoup moins d'utilisateurs actifs, et l'écosystème numérique actuel reste dominé par une poignée de méga-plateformes cherchant à générer des profits. Tout cela contraint les défenseurs des droits humains à déterminer un équilibre entre visibilité et sécurité. Plusieurs organisations nous ont fait part des difficultés rencontrées par les défenseurs pour se désengager complètement sans risquer de compromettre l'impact de leur travail. Le manque d'alternatives sûres conduit ainsi les défenseurs à rester continuellement exposés aux risques. [104]

[98] Contribution (H) reçue par la Rapporteuse spéciale

[99] Contribution (J) reçue par la Rapporteuse spéciale

[100] Contribution (E) reçue par la Rapporteuse spéciale

[101] Contribution (F) reçue par la Rapporteuse spéciale

[102] *Global Witness*, *ibid.*

[103] Contributions (A), (C), (F) et (I) reçues par la Rapporteuse spéciale

[104] Contributions (A), (B) et (C) reçues par la Rapporteuse spéciale

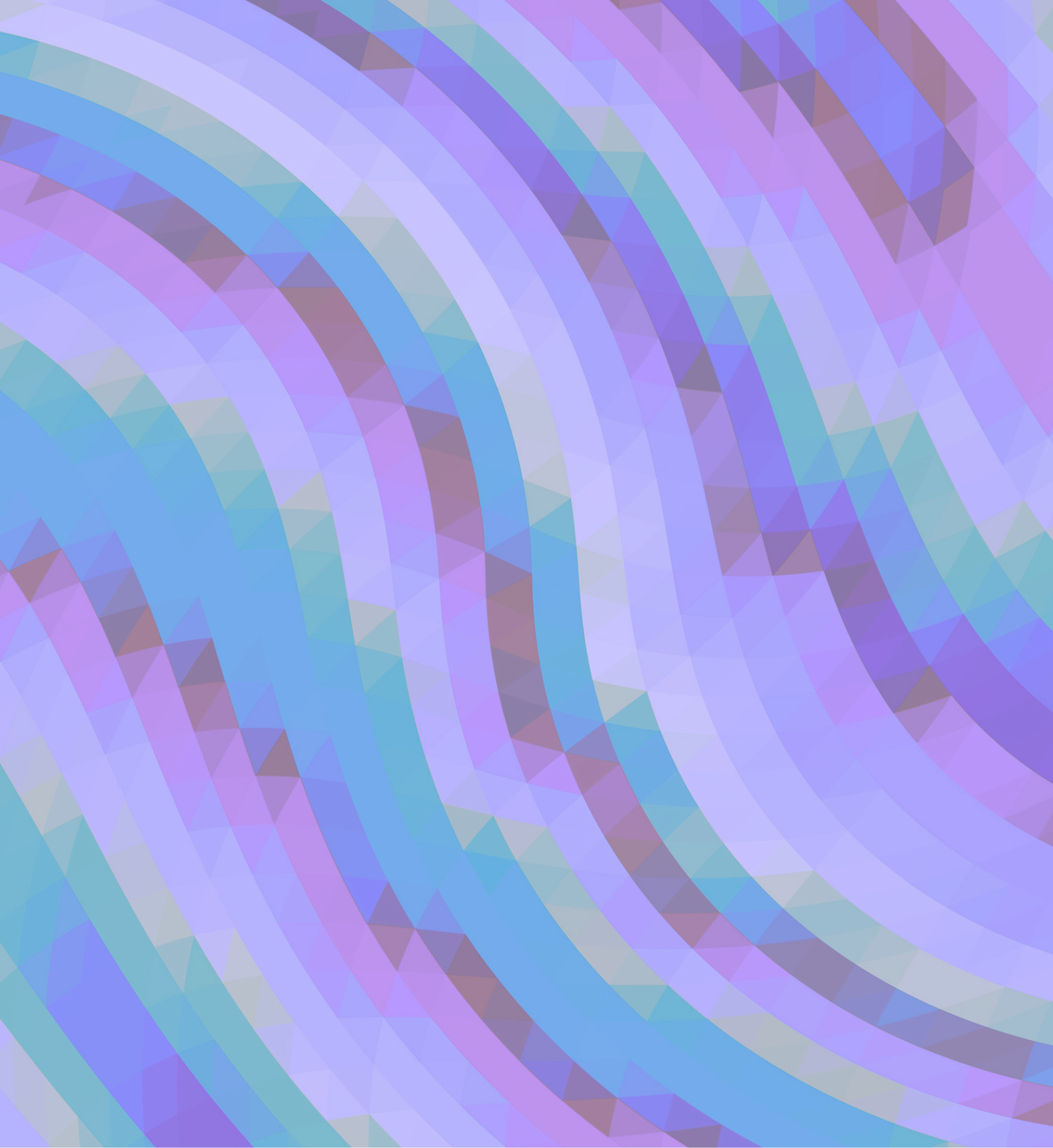
RECOMMANDATIONS

À L'ATTENTION DES ENTREPRISES GESTIONNAIRES DE PLATEFORMES DE RÉSEAUX SOCIAUX :

- Veiller à ce que les **politiques de modération et de curation des contenus** soient strictement conformes au droit international des droits humains et ne portent pas atteinte au droit de défendre les droits humains ;
- S'abstenir de développer et d'utiliser l'**IA** dans le cadre de la modération et de la suppression de contenus sans vérification et validation humaines par des tiers ;
- Améliorer la **transparence** des décisions de modération de contenu et des procédures de recours ;
- Renforcer les **outils de confidentialité et de sécurité**, notamment par un chiffrement plus robuste, des signalements anonymes et des mesures de protection contre le doxxing et la surveillance ;
- Évaluer l'**efficacité des mécanismes de signalement** des abus en ligne et investir dans des dispositifs d'assistance fiables, accessibles et gérés par des **personnes humaines** afin d'apporter une aide aux défenseurs des droits humains qui en ont besoin, avec des équipes formées aux langues, cultures et contextes politiques concernés ;
- Garantir des **délais de réponse plus courts**, en particulier pour les partenaires de confiance et les cas urgents ;
- Investir dans des outils et des protocoles de modération tenant compte des **questions de genre** et intégrer dans les directives communautaires des sections spécifiques consacrées au harcèlement et aux abus fondés sur le genre ;
- Mettre en place des **audits externes des algorithmes** afin de limiter la propagation de la désinformation et des contenus haineux ou injurieux, et s'engager à concevoir des systèmes inclusifs qui ne marginalisent pas les contenus relatifs aux droits humains ;
- Refuser de se conformer aux **demandes de censure** émanant des États qui sont contraires aux normes internationales en matière de droits humains ;
- Consigner et **publier les demandes de retrait** émanant des États ainsi que les réponses des plateformes ;
- Revenir sur les **reculs récents en matière de politiques relatives aux discours de haine, à la désinformation et à la vie privée**, ainsi qu'en matière de dotation en personnel des équipes « Intégrité et sûreté » et « Droits humains » ;
- Mettre en place et renforcer les **canaux de communication avec les défenseurs des droits humains et la société civile**, en particulier dans les régions où il n'existe pas de dispositifs structurés, et veiller à les associer à la conception de politiques et de mécanismes de soutien qui répondent à leurs besoins.

À L'ATTENTION DES ÉTATS :

- Veiller à ce que les entreprises de réseaux sociaux **respectent la liberté d'expression** et les **normes en matière de droits humains** ;
- Envisager des **solutions régulatrices** pour contrer le monopole des plateformes dominant le marché ;
- Reconnaître la **violence en ligne** comme s'inscrivant dans le continuum plus large des attaques contre les défenseurs des droits humains ;
- **Enquêter** sur les violences en ligne à l'encontre des défenseurs des droits humains et **engager des poursuites** contre leurs auteurs ;
- Adopter des **cadres juridiques et politiques clairs** qui engagent la responsabilité des plateformes de réseaux sociaux pour leur rôle en matière de harcèlement en ligne, en particulier lorsque ces actes menacent la sécurité des défenseurs des droits humains ;
- S'abstenir de contraindre les plateformes de réseaux sociaux à mettre en œuvre des mesures de censure ou de surveillance.



HRC-SR-DEFENDERS@UN.ORG
WWW.SRDEFENDERS.ORG



SOLIDARITÉ ET ESPOIR