

POUSSÉ·E·S VERS LES TÉNÈBRES

COMMENT LE FIL « POUR TOI » ENCOURAGE L'AUTOMUTILATION
ET LES IDÉES SUICIDAIRES

AMNESTY
INTERNATIONAL



Amnesty International est un mouvement rassemblant 10 millions de personnes qui fait appel à l'humanité en chacun et chacune de nous et milite pour que nous puissions toutes et tous jouir de nos droits humains.

Notre vision est celle d'un monde dans lequel les dirigeants et dirigeantes tiennent leurs promesses, respectent le droit international et sont tenus de rendre des comptes.

Indépendante de tout gouvernement, de toute idéologie politique, de tout intérêt économique et de toute religion, Amnesty International est essentiellement financée par ses membres et des dons de particuliers. Nous avons la conviction qu'agir avec solidarité et compassion aux côtés de personnes du monde entier peut rendre nos sociétés meilleures.

© Amnesty International 2023

Sauf exception dûment mentionnée, le contenu de ce document est sous licence internationale 4.0 Creative Commons (paternité, pas d'utilisation commerciale, pas de modification). <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> Pour plus d'informations, veuillez consulter la page relative aux autorisations sur notre site : www.amnesty.org/fr.

Lorsqu'une entité autre qu'Amnesty International est détentrice du copyright, le matériel n'est pas sous licence Creative Commons.

Édition originale publiée en 2020
par Amnesty International Ltd
Peter Benenson House, 1 Easton Street
London WC1X 0DW, Royaume-Uni

Index: POL 40/7350/2023

Langue originale : Anglais

amnesty.org



Photo de couverture : © Luisa Balaban

**AMNESTY
INTERNATIONAL**



AVERTISSEMENT

Le présent rapport porte sur des questions sensibles, notamment l'automutilation et le suicide, et contient des images explicites provenant de contenus diffusés sur TikTok qui concernent ces sujets. Vous trouverez les coordonnées de lignes d'assistance téléphonique et d'organisations qui apportent un soutien émotionnel gratuit aux quatre coins de la planète dans le guide d'Amnesty International intitulé **Sauver le monde sans s'effondrer (volume 2) : Manuel du bien-être pour les jeunes militant·e·s.**¹

REMERCIEMENTS

Amnesty International tient à remercier ses partenaires techniques, à savoir l'Algorithmic Transparency Institute (National Conference on Citizenship) et AI Forensics, ainsi que les personnes, organisations, militant·e·s et spécialistes qui ont facilité et orienté ces recherches, notamment : S. Bryn Austin, Zeyna Awan, Jerome Cleofas, Julien Cornebise, Kevin Gachee, In Touch Community Services, Lillian Kariuki, KERI: Caring for Activists, Amanda Raffoul, Marc Reyes, Aliya Shah, Talang Dalisay, Ian Wafula et Youth for Mental Health Coalition.

Amnesty International est également reconnaissante aux enfants et aux jeunes du Kenya et des Philippines qui lui ont fait part de leur expérience et de leurs idées lors des recherches menées aux fins du présent rapport.

1. Amnesty International, Sauver le monde sans s'effondrer (volume 2) : Manuel du bien-être pour les jeunes militant·e·s (index : ACT 10/3231/2020), janvier 2021, [amnesty.org/fr/documents/act10/3231/2020/fr/](https://www.amnesty.org/fr/documents/act10/3231/2020/fr/), p. 99.

SOMMAIRE

Glossaire	5
1. RÉSUMÉ	6
2. MÉTHODE	11
2.1 Objet et portée de ce rapport	11
2.2 Axe de recherche	12
2.3 Méthodologie de recherche	13
3. CONTEXTE	16
3.1 Passer à côté de l'essentiel : lacunes en matière de recherche et obstacles commerciaux entravant l'étude des risques que présentent les plateformes de réseaux sociaux du point de vue des droits humains des enfants et des jeunes	16
3.2 Entre politiques et populisme : des efforts limités pour réglementer les plateformes de réseaux sociaux	18
3.3 Aggravation des problèmes de santé mentale chez les enfants et les jeunes	22
4. CADRE DE PROTECTION DES DROITS HUMAINS	24
4.1 Le droit au respect de la vie privée à l'ère des réseaux sociaux	25
4.2 Le droit à la liberté de pensée	27
4.3 Le droit à la santé	29
4.4 L'intérêt supérieur de l'enfant et le droit d'être entendu-e	31
4.5 L'obligation des entreprises de respecter les droits humains	31
5. UNE CONCEPTION ADDICTIVE PAR NATURE	34
5.1 Les stratégies d'encouragement de l'engagement de TikTok	36
5.2 Les conséquences des réseaux sociaux addictifs pour la santé des jeunes	38
6. DESCENTE EN SPIRALE	42
6.1 Comment le fil « Pour Toi » de TikTok pousse à l'extrême un système intrinsèquement dangereux	42
6.2 Des risques accrus pour les jeunes souffrant de problèmes de santé mentale	46
6.3 Exploration systématique du système de recommandation de TikTok	47
6.4 Mesurer et catégoriser les contenus nocifs aux fins de l'étude	50
6.5 Principales conclusions	53
6.6 L'intérêt pour la santé mentale déclenche l'effet de « spirale »	54
6.7 Exemples de publications recommandées	63
6.8 Les limites du projet	67
7. LES ÉCHECS DES ENTREPRISES	69
7.1 Manque de diligence requise	69
7.2 Des réponses inadéquates à des risques graves	73
8. CONCLUSION ET RECOMMANDATIONS	76
Annexe	82

GLOSSAIRE

CDC	Centres de contrôle et de prévention des maladies (États-Unis)
CdE	Conseil de l'Europe
CDE	Convention relative aux droits de l'enfant
CDH (notes de bas de page)	Conseil des droits de l'homme [ONU]
CEDH (notes de bas de page)	Cour européenne des droits de l'homme
DSA	Règlement de l'Union européenne sur les services numériques
DUDH	Déclaration universelle des droits de l'homme
DSA	The European Union's Digital Services Act
Effet de « spirale »	Expression couramment utilisée dans l'étude des plateformes de réseaux sociaux ; dans ce rapport, elle désigne le fait de resserrer les recommandations de contenu sur le fil « Pour toi » de TikTok autour d'un seul sujet, en partant du principe que ce contenu suscitera de vives réactions émotionnelles et poussera l'utilisateur ou l'utilisatrice à poursuivre le visionnage
Géants du numérique (ou géants technologiques)	Expression courante désignant les principales entreprises de technologies de l'information qui façonnent et dominent Internet
IA	Intelligence artificielle : Il n'existe pas de définition universellement admise du terme « intelligence artificielle » (ou « AI »). Le Haut-Commissariat des Nations Unies aux droits de l'homme utilise ce terme pour désigner tout un éventail de processus et de technologies permettant à des ordinateurs d'aider ou de remplacer des humains dans des tâches spécifiques, comme la prise de décisions et la résolution de problèmes, y compris l'apprentissage automatique et l'apprentissage en profondeur. ²
ICESCR	International Covenant on Economic, Social and Cultural Rights
LGBTI	Lesbiennes, gays et personnes bisexuelles, transgenres ou intersexes
Métadonnées	Données relatives à des fichiers et des communications numériques, par exemple données d'horodatage, de localisation
Moissonnage	Collecte automatique de données en ligne à grande échelle
PIDCP	Pacte international relatif aux droits civils et politiques
PIDESC	Pacte international relatif aux droits économiques, sociaux et culturels
Principes directeurs de l'OCDE	Principes directeurs de l'OCDE à l'intention des entreprises multinationales
Principes directeurs des Nations unies	Principes directeurs relatifs aux entreprises et aux droits de l'homme [ONU]
TDAH	Trouble du déficit de l'attention avec hyperactivité
VPN	Réseau privé virtuel

2. Haut-Commissariat des Nations Unies aux droits de l'homme (HCDH), *Le droit à la vie privée à l'ère du numérique*, 15 septembre 2021, A/HRC/48/31, note de bas de page 2.

1. RÉSUMÉ

« Avez-vous déjà regardé une plaquette de comprimés en pensant à tout avaler ? Avez-vous déjà simplement souhaité en finir ? Avez-vous déjà pensé au soulagement que vous ressentiriez, à toute cette douleur qui disparaîtrait ? » Sur fond de jeu d'arcade où de petits personnages pixélisés circulent dans un labyrinthe, une voix produite par intelligence artificielle (IA) prononce ces phrases lors d'une séquence vidéo d'une demi-minute. Alors qu'elle visionne et revoit cette séquence, puis passe à la suivante, la personne de 13 ans qui parcourt ces vidéos ne ressent rien. Et pour cause : il s'agit d'un compte automatisé créé et programmé par Amnesty International et l'Algorithmic Transparency Institute pour simuler et étudier la réalité numérique dans laquelle les enfants et les jeunes qui présentent des problèmes de santé mentale comme l'anxiété ou la dépression. À chaque heure passée sur le fil « Pour toi » de TikTok, ce sont toujours plus de vidéos « recommandées » qui montrent des enfants et des jeunes en pleurs, ou seuls dans le noir, accompagnées de textes exprimant des pensées dépressives ou de voix sans visage décrivant leur souffrance, leurs actes d'automutilation et leurs pensées suicidaires.

Au cours des trois dernières années, marquées par la pandémie de COVID-19, TikTok est devenu une plateforme mondiale, attirant des centaines de millions d'enfants et de jeunes. Sa page « Pour toi » et le système de recommandation par algorithme sur laquelle elle se fonde ont joué un rôle crucial, rendant la plateforme omniprésente dans la vie des enfants et des jeunes. Le rapport *I feel exposed: Caught in TikTok's Surveillance Web*, publié en accompagnement au présent document, met en évidence le modèle profondément discriminatoire et intrusif de surveillance à des fins mercantiles qui sous-tend le modèle d'activité de TikTok. Ce rapport complémentaire porte sur la surveillance des utilisateurs et utilisatrices, les décisions en matière de conception et les recommandations de contenu personnalisées que TikTok exploite pour obtenir la « participation » de ses utilisateurs et utilisatrices, ainsi que l'impact préjudiciable de ces outils dans le monde réel pour les enfants et les jeunes.

Le système de recommandation de TikTok, autrement dit l'ensemble d'algorithmes qui analysent les intérêts et les modes de participation des utilisateurs et utilisatrices, pour les mettre en correspondance avec de nouvelles vidéos afin que ces personnes demeurent investies émotionnellement et que leurs yeux restent fixés sur le fil, n'a pas été conçu pour produire une « spirale » de contenu dépressif. Cependant, si de jeunes utilisateurs et utilisatrices montrent un intérêt net pour un contenu relatif à la santé mentale, c'est à ce type de contenu que le système les associera en vue d'optimiser leur participation. Si le compte automatique créé aux fins de l'étude était celui d'un être humain, mieux à même de distinguer ce qui est pertinent de ce qui ne l'est pas, le système produirait ces recommandations encore plus rapidement.

Voici comment « Luis » (prénom modifié), un étudiant de 21 ans vivant à Manille et souffrant de troubles bipolaires diagnostiqués, a décrit son expérience avec le fil « Pour toi » de TikTok :

« Quand je n'avais pas le moral, je pense que 80 % [des contenus] étaient liés à la santé mentale. C'est une sorte d'engrenage. C'est une spirale infernale, parce que ça commence avec une seule vidéo. Si une vidéo parvient à capter ton attention, même si tu ne l'aimes pas, elle t'est de nouveau présentée quand tu ouvres TikTok la fois suivante et, parce qu'elle te semble familière, tu la regardes à nouveau et alors, sa fréquence d'apparition dans ton fil augmente exponentiellement ».

Ce rapport se fonde sur une étude documentaire, une enquête de cadrage, des discussions de groupe et des entretiens avec plus de 300 enfants et jeunes de deux des pays où le taux d'utilisation des réseaux sociaux est le plus élevé au monde (le Kenya et les Philippines) ainsi que sur une enquête technique, menée conjointement avec l'Algorithmic Transparency Institute et AI Forensics, qui ont joué le rôle de partenaires techniques. À cet effet, plus de 30 comptes automatisés ont été utilisés afin de comparer les résultats entre le Kenya et les États-Unis, outre les expériences réalisées manuellement au Kenya, aux Philippines et aux États-Unis.

Pour l'audit automatisé, l'équipe de recherche a créé 40 comptes automatisés avec quatre personas prédéfinies présentant divers degrés d'intérêt pour les contenus liés à la santé mentale, de façon à imiter différents comportements adoptés par des enfants sur TikTok. Chaque compte a été paramétré pour fonctionner par sessions uniques de près de 60 minutes par jour pendant 10 jours. Les comptes ont été divisés en quatre sous-groupes correspondant chacun à un certain comportement de défilement des contenus, chaque sous-groupe visionnant des vidéos associées à un ensemble donné de termes et de hashtags et passant le contenu sans rapport avec ceux-ci : 20 comptes ont été paramétrés pour imiter les comportements de personnes de 13 ans aux États-Unis, tandis que 20 autres simulaient les comportements d'enfants au Kenya (dont 11 ont été inclus dans l'analyse). L'âge correspondant aux comptes (13 ans) a été choisi afin de pouvoir examiner les recommandations proposées à la tranche d'âge la plus jeune autorisée à utiliser TikTok, sachant que la plateforme applique aux comptes concernés des mesures de sécurité propres aux adolescent-e-s. Dans le cadre de l'expérience complémentaire réalisée manuellement, des interactions d'une heure entre l'équipe de recherche et le fil « Pour toi » de comptes nouvellement créés imitant eux aussi le comportement d'enfants âgés de 13 ans ont été enregistrées.

D'après les recherches d'Amnesty International, TikTok a porté à leur maximum les caractéristiques addictives des choix de conception et des stratégies de participation pour lesquels ont opté des entreprises de réseaux sociaux concurrentes, encourageant un nivellement vers le bas au sein d'une poignée de géants du secteur rivalisant pour obtenir le plus grand nombre d'utilisateurs et utilisatrices et le taux de participation le plus élevé. TikTok a adopté cette ligne de conduite en dépit des éléments scientifiques de plus en plus nombreux qui mettent en évidence les graves risques liés à une utilisation addictive des réseaux sociaux, en particulier pour la santé des enfants et des jeunes, notamment les problèmes de sommeil et d'attention, et même des modifications de la structure cérébrale analogues à celles observées chez les personnes souffrant d'une dépendance à la drogue.

Outre son caractère addictif, le fil « Pour toi » de TikTok comporte encore d'autres risques pour les enfants et les jeunes ayant déjà des problèmes de santé mentale. Une enquête technique réalisée par Amnesty International, l'Algorithmic Transparency Institute (National Conference on Citizenship) et AI Forensics montre que les enfants et les jeunes qui regardent des contenus relatifs à la santé mentale sur la page « Pour toi » de la plateforme peuvent facilement tomber dans des « spirales » de contenus potentiellement préjudiciables, notamment des vidéos qui idéalisent et encouragent les pensées dépressives, l'automutilation et le suicide.

Au bout de cinq ou six heures passées sur la plateforme, presque une vidéo sur deux présentées aux comptes automatisés, programmés pour simuler le comportement d'un enfant de 13 ans (au Kenya et aux États-Unis) présentant un intérêt pour la santé mentale, était relative à ce sujet et potentiellement nocifs, soit un volume 10 fois plus important que celui présenté aux comptes n'ayant indiqué aucun

intérêt pour le sujet. Une analyse manuelle de 540 vidéos recommandées à un échantillon de ces comptes automatisés a montré une progression régulière : 17 % des vidéos présentées la première heure étaient classées comme potentiellement préjudiciables contre 44 % des contenus proposés la 10e heure (sur la base de sessions d'une heure réparties sur 10 jours).

Amnesty International a observé un effet de « spirale » encore plus rapide, associé à un taux encore plus élevé de contenu potentiellement dangereux lorsque les chercheurs et chercheuses ont visionné manuellement des vidéos liées à la santé mentale suggérées à un public censé être composé d'enfants de 13 ans au Kenya, aux Philippines et aux États-Unis. Parmi les recommandations proposées à un compte situé aux Philippines, la première vidéo accompagnée du hashtag « #depressionanxiety » (sic) (« dépressionanxiété ») et montrant un garçon en état de détresse a été suggérée au bout de 67 secondes de défilement du contenu recommandé sur la page « Pour toi ». À partir de 12 minutes de défilement, 58 % des publications recommandées avaient un rapport avec l'anxiété, la dépression, l'automutilation et/ou le suicide, et étaient catégorisées comme pouvant avoir des effets néfastes sur les enfants et les jeunes souffrant déjà de problèmes de santé mentale.

Dans le cadre de l'expérience manuelle menée aux États-Unis, la quatrième vidéo présentée était taguée « #paintok » (« douleurtok »). Son contenu portait sur le texte suivant : « Quand tu te rends compte qu'on ne t'a jamais fait passer avant quoi que ce soit de ta vie, mais que tu es juste cette personne qui comble le vide dans les vies des autres jusqu'à ce qu'ils n'aient plus besoin de toi. » À partir de la 20e vidéo (soit au bout de moins de trois minutes de connexion), 57 % des vidéos avaient un rapport avec des problèmes de santé mentale, avec au moins neuf publications qui idéalisaient, banalisaient ou encourageaient le suicide en seulement une heure.

Dans le cadre de ces expériences manuelles, c'est le compte kenyan qui a vu la progression la plus lente vers un fil rempli de contenus relatifs à la dépression. Néanmoins, une fois ce stade atteint (20 minutes après le début de l'expérience), 72 % des vidéos recommandées au cours des 40 minutes suivantes étaient en lien avec des difficultés de santé mentale, avec au moins cinq références aux pensées suicidaires ou au désir de mort de la personne ayant créé le contenu. Absolument aucune des vidéos relatives à la santé mentale n'avait été produite par un-e professionnel-le de la santé mentale ni une organisation de santé mentale reconnue.

Le cas de Molly Russell, l'adolescente britannique de 14 ans qui s'est donné la mort après avoir visionné du contenu lié à la dépression sur Instagram, montre de la manière la plus tragique qui soit qu'exposer des jeunes présentant des symptômes dépressifs à un fil constitué d'une quantité importante de publications qui évoquent, banalisent voire idéalisent les pensées dépressives, l'automutilation et le suicide sur les réseaux sociaux peut aggraver des problèmes de santé mentale préexistants chez les jeunes utilisateurs et utilisatrices et encourager ainsi des actes dangereux et même dévastateurs dans la vie réelle. De telles interférences avec les pensées et les émotions d'une personne constituent une atteinte au droit à la liberté de pensée et au droit à la santé.

Compte tenu de la fragilité émotionnelle des enfants et des jeunes adultes, qui a fait l'objet de beaucoup d'études, et des nombreux éléments rapportés par la société civile et les médias, TikTok doit savoir et s'être rendu compte que son système algorithmique de recommandation risque de faire entrer les jeunes utilisateurs et utilisatrices dans une « spirale » de publications potentiellement dangereuses.

Pour remplir ses obligations aux termes des Principes directeurs relatifs aux entreprises et aux droits de l'homme [Organisation des Nations unies (ONU)], TikTok doit mettre en place comme il se doit une procédure de diligence raisonnable en matière de droits humains qui lui permette de cerner ses incidences réelles et potentielles sur les droits humains, de les prévenir et de les atténuer, ainsi que de rendre compte de la manière dont il y remédie. Dans le cadre de cette procédure, l'entreprise doit

déceler les risques pour les enfants et les jeunes qui sont inhérents à la conception de sa plateforme, compte tenu des éléments de plus en plus nombreux qui attestent les risques systémiques liés aux systèmes algorithmiques de recommandation dont ont fait état les médias, les organisations de la société civile et les institutions internationales, et prendre des mesures pour les atténuer. La réponse de l'entreprise aux questions d'Amnesty International fait apparaître des mesures disparates comme la redirection de certaines recherches vers des ressources d'aide, la possibilité de « rafraîchir » le fil et des partenariats avec des organisations de soutien en matière de santé mentale. La réaction de TikTok témoigne de l'absence de procédures adéquates face au caractère systémique de ces risques, et les mesures et outils de sécurité mis en place par l'entreprise sont loin d'être à la hauteur de l'ampleur des risques systémiques décrits dans ce rapport.

Quoi qu'il en soit, l'action isolée d'une seule entreprise ne saurait suffire à contenir les effets d'un modèle économique fondamentalement incompatible avec les droits humains, en particulier le droit au respect de la vie privée, le droit à la liberté de pensée et le droit à la santé. Par conséquent, il faut que les États réglementent efficacement les activités des « géants du numérique » comme TikTok, conformément au droit international relatif aux droits humains et aux normes connexes, de façon à protéger les droits des enfants et des jeunes.

Amnesty International appelle TikTok à mettre en œuvre de toute urgence les recommandations suivantes :

- Transition vers un modèle d'activité respectueux des droits qui ne se fonde pas sur un suivi intrusif des données. Pour commencer, TikTok doit veiller à ce que ses politiques et processus de diligence nécessaire en matière de droits humains prennent en compte les risques systémiques et les incidences généralisées sur les droits humains qui sont liés à son modèle d'activité, en particulier en ce qui concerne le droit au respect de la vie privée, le droit à la liberté de pensée et le droit à la santé.
- Il faut que TikTok cesse de maximiser « la participation des utilisateurs et utilisatrices » aux dépens de la santé et des autres droits humains de ceux-ci. Dans le cadre de son processus de diligence nécessaire en matière de droits humains, TikTok doit déceler, en coopération avec les utilisateurs et utilisatrices, y compris les enfants et les jeunes, ainsi qu'avec des spécialistes indépendants, les éléments de conception qui encouragent une utilisation addictive de la plateforme et des comparaisons sociales, et les remplacer par une expérience axée sur la « sécurité par la conception » et l'intérêt supérieur de l'enfant.
- Pour respecter la vie privée des utilisateurs et utilisatrices et leur donner une réelle capacité de choix et de contrôle, un écosystème de réseaux sociaux sans profilage devrait être la norme et non pas une option parmi d'autres. Les algorithmes utilisés par TikTok et les autres plateformes en ligne pour organiser le contenu ne devraient donc pas être basés par défaut sur le profilage (par exemple, en fonction du temps de visionnage ou de la participation), et doivent reposer sur une démarche volontaire d'inscription plutôt que de désinscription, le consentement pour l'inscription devant être spécifique et donné en connaissance de cause (notamment grâce à l'utilisation d'un langage accessible aux enfants), librement et sans ambiguïté.
- Il faut que TikTok cesse de collecter des données personnelles sensibles et d'opérer des déductions quant aux intérêts, à l'état émotionnel ou au bien-être des utilisateurs et utilisatrices en fonction du temps de visionnage et de la participation en vue de « personnaliser » les recommandations de contenu et de cibler les publicités. Au lieu de se livrer à une surveillance généralisée pour adapter les fils aux intérêts des utilisateurs et utilisatrices, TikTok doit permettre à ceux-ci de faire part de leurs intérêts de manière délibérée (par exemple, il pourrait être demandé aux utilisateurs et utilisatrices de communiquer des intérêts en particulier s'ils/elles souhaitent

obtenir des recommandations personnalisées) et seulement sur la base d'un consentement libre, spécifique et donné en connaissance de cause.

Pour concrétiser les droits des enfants et des jeunes, les États doivent :

- Empêcher les entreprises de conditionner l'accès à leurs services au fait d'« accepter » la collecte, le traitement et le partage de leurs données personnelles à des fins de ciblage de contenu, de marketing ou de publicité ;
- Réglementer les activités des entreprises de réseaux sociaux conformément au droit international relatif aux droits humains et aux normes connexes afin de veiller à ce que les algorithmes utilisés par les plateformes en ligne pour organiser le contenu ne soient pas fondés par défaut sur le profilage, et que ces plateformes demandent une démarche volontaire d'inscription plutôt que de désinscription, le consentement pour l'inscription devant être spécifique et donné librement, en connaissance de cause et sans ambiguïté. La collecte et l'utilisation de données personnelles sensibles obtenues par déduction (par exemple, le fait de baser les recommandations sur le temps de visionnage et les mentions « Like », qui permettent de déduire des informations sensibles) pour personnaliser les publicités et les recommandations de contenu doivent être interdites.

2. MÉTHODE

2.1 OBJET ET PORTÉE DE CE RAPPORT

Ce rapport expose les risques pour les droits humains qui sont inhérents à l'utilisation des recommandations de contenu ciblées que font les grandes plateformes de réseaux sociaux et à la concurrence qu'elles se livrent pour capter l'attention du jeune public, l'accent étant mis sur la conception addictive de TikTok et la façon dont cette plateforme amplifie les contenus liés à la dépression et à l'automutilation.

Il est l'un des deux rapports d'Amnesty International portant sur les incidences en matière de droits humains du modèle économique fondé sur la surveillance qui est celui de TikTok et d'autres plateformes de réseaux sociaux – modèle reposant sur un suivi intrusif des utilisateurs et utilisatrices que l'organisation a décrit comme « un pacte faustien, car ils [les utilisateurs et utilisatrices] ne peuvent jouir de leurs droits en ligne qu'en souscrivant à un système reposant sur des atteintes aux droits humains³ » – face à des enfants (personnes de moins de 18 ans) et des jeunes (15-24 ans, également appelés « jeunes utilisateurs et utilisatrices » ou « jeune public » dans ce rapport).

Le rapport qui accompagne celui-ci, intitulé *I feel exposed: Caught in TikTok's Surveillance Web*, met en lumière les atteintes commises par des entreprises technologiques à l'encontre du droit des enfants au respect de la vie privée, lesquelles sont inhérentes au modèle d'activité des plateformes des « géants du numérique⁴ ». Ce rapport complémentaire concerne les effets délétères sur les enfants et les jeunes que peuvent avoir dans la vie réelle la surveillance des utilisateurs, les décisions en matière de conception et les recommandations de contenu personnalisées que les entreprises de réseaux sociaux exploitent pour optimiser la « participation » de ses utilisateurs et utilisatrices. Il détaille les risques que TikTok présente pour les enfants et les jeunes, en particulier ceux qui souffrent de problèmes de santé mentale comme la dépression, l'anxiété et l'automutilation.

Il s'appuie sur des travaux antérieurs d'Amnesty International, notamment son rapport de 2022 intitulé *The Social Atrocity – Meta and the Right to Remedy for the Rohingya*, qui exposait le rôle de Facebook dans l'amplification des publications incitant à la violence, à la haine et à la discrimination à l'encontre du peuple rohingya, contribuant au bout du compte au nettoyage ethnique perpétré par les forces de sécurité du Myanmar en 2017.⁵

Ensemble, ces rapports contribuent à la constitution d'une base croissante d'éléments en faveur de la campagne mondiale d'Amnesty International pour la responsabilité des entreprises et à la réparation

3. Amnesty International, *Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights* (index : POL 30/1404/2019), 21 novembre 2019, <https://www.amnesty.org/en/documents/pol30/1404/2019/en>, p. 5.

4. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, (index : POL 40/7349/2023), 7 novembre 2023.

5. Amnesty International, *The Social Atrocity: Meta and the Right to Remedy for the Rohingya* (index : ASA 16/5933/2022), 29 septembre 2022, <https://www.amnesty.org/en/documents/asa16/5933/2022/en>.

des atteintes aux droits humains découlant du modèle d'activité fondé sur la surveillance pour lequel ont opté Meta, Google, TikTok et d'autres « géants du numérique ».

2.2 AXE DE RECHERCHE

Toutes les entreprises de réseaux sociaux rivalisent pour obtenir un nombre toujours croissant d'utilisateurs et utilisatrices et la plus grande part de leur temps d'écran. Ce rapport est axé sur TikTok car, au cours des trois dernières années, marquées par la pandémie de COVID-19, l'entreprise est devenue une plateforme mondiale, attirant des centaines de millions d'enfants et de jeunes, dont beaucoup déclarent aujourd'hui qu'il s'agit de leur plateforme préférée, selon des études de marché.⁶ Du fait de la croissance rapide de TikTok, le nombre d'utilisateurs et utilisatrices actifs a globalement doublé entre décembre 2019 et septembre 2021;⁷ ce mois-là, l'entreprise a annoncé avoir dépassé le milliard d'utilisateurs et utilisatrices mensuels.⁸ La majorité des utilisateurs et utilisatrices de TikTok seraient des enfants et des jeunes.⁹ En août 2020, il a été indiqué que TikTok avait « classé plus d'un tiers de ses 49 millions d'utilisateurs et utilisatrices quotidiens aux États-Unis dans la catégorie des 14 ans et moins », et que nombre de ces enfants avaient probablement moins de 13 ans, l'âge minimum requis par l'entreprise pour accéder à sa version adulte dans ce pays et la plupart des autres où la plateforme est présente.¹⁰

L'essor rapide de TikTok, qui lui a permis de dominer le marché, ainsi que les éléments de conception particuliers que l'entreprise utilise pour maximiser le temps passé par les utilisateurs et utilisatrices, suscitent des inquiétudes croissantes quant au fait que l'entreprise expose des enfants et des jeunes à une plateforme addictive et potentiellement dangereuse.¹¹

Bien que TikTok soit utilisé presque partout dans le monde, les études disponibles concernant son impact (et celui d'autres plateformes de réseaux sociaux) sur les enfants et les jeunes, y compris les dangers potentiels, portent très majoritairement sur l'Australie, l'Europe et les États-Unis.¹² Près de

6. Quostodio, "Social Media Annual Data Report 2021: Living and Learning in a Digital World", Chapter 2, Social Media, <https://quostodio.com/en/social-media-quostodio-annual-data-report-2021/>.
7. We Are Social and Hootsuite, "Digital 2022 Global Overview Report", 26 janvier 2022, <https://wearesocial.com/hk/blog/2022/01/digital-2022-another-year-of-bumper-growth>.
8. Reuters, "TikTok hits 1 billion monthly active users globally – company", 27 septembre 2021, <https://www.reuters.com/technology/tiktok-hits-1-billion-monthly-active-users-globally-company-2021-09-27>.
9. Meltwater, "54 TikTok stats you need to know [2023]", 30 décembre 2022, <https://www.meltwater.com/en/blog/tiktok-statistics> ; *The Guardian*, « What TikTok does to your mental health: "It's embarrassing we know so little" », 30 octobre 2022, <https://www.theguardian.com/technology/2022/oct/30/tiktok-mental-health-social-media> ; Reuters Institute, "Digital News Report 2023", juin 2023, <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/dnr-executive-summary>.
10. New York Times, "A third of TikTok's U.S. users may be 14 or under, raising safety questions", 14 août 2020, <https://www.nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html> ; TikTok offre une « expérience de l'application limitée » assortie de « protections supplémentaires en matière de sécurité et de respect de la vie privée » aux utilisateurs et utilisatrices de moins de 13 ans, voir TikTok, "TikTok for younger users", 13 décembre 2019, <https://www.newsroom.tiktok.com/en-us/tiktok-for-younger-users>. L'article du *New York Times* fait référence à des allégations formulées par un ancien employé de TikTok selon lequel l'entreprise aurait permis à des vidéos provenant d'enfants de moins de 13 ans de « rester en ligne pendant des semaines ». Étant donné que la version de l'application destinée aux moins de 13 ans ne permet pas de partager des vidéos, cela indique que les personnes concernées ont utilisé la version adulte, probablement en mentant sur leur âge. En 2019, TikTok a accepté de payer une amende de 5,7 millions de dollars des États-Unis face aux allégations de la Commission fédérale du commerce (FTC) selon lesquelles l'entreprise avait collecté illégalement des informations sur des enfants de moins de 13 ans. *Wired*, "FTC Hits TikTok With Record \$5.7 Million Fine Over Children's Privacy", 27 février 2019, <https://www.wired.com/story/tiktok-ftc-record-fine-childrens-privacy/> ; pour en savoir plus sur la politique de TikTok en matière d'âge, voir TikTok, "Guardian's Guide", <https://www.tiktok.com/safety/en/guardians-guide/> (consulté le 26 septembre 2023).
11. Voir, par exemple : *Wall Street Journal*, "The corpse bride diet": How TikTok inundates teens with eating-disorder videos", 17 décembre 2021, <https://www.wsj.com/articles/how-tiktok-inundates-teens-with-eating-disorder-videos-11639754848> ; Sophia Petrillo, "What makes TikTok so addictive?: An analysis of the mechanisms underlying the world's latest social media craze", 13 décembre 2021, *Brown Undergraduate Journal of Public Health*, numéro de 2021-2022, <https://www.sites.brown.edu/publichealthjournal/2021/12/13/tiktok/> ; BBC, "Inside TikTok's real-life frenzies - from riots to false murder accusations", 21 septembre 2023, <https://www.bbc.co.uk/news/technology-66719572> ; *The Guardian*, "What TikTok does to your mental health: 'It's embarrassing we know so little'", 30 octobre 2022, <https://www.theguardian.com/technology/2022/oct/30/tiktok-mental-health-social-media>.
12. Sakshi Ghai, Lucia Magis-Weinberg et autres, "Social media and adolescent well-being in the Global South", août 2022, *Current Opinion in Psychology*, volume 46, <https://pubmed.ncbi.nlm.nih.gov/35439684/>.

90 % des jeunes du monde vivent dans des pays en développement,¹³ mais ils sont extrêmement sous-représentés dans les études et les débats publics sur les réseaux sociaux. Afin d'élargir la base d'éléments géographiques, les recherches menées dans le cadre de ce rapport se sont concentrées sur l'utilisation de TikTok par des enfants et des jeunes du Kenya et des Philippines, deux pays choisis en raison du temps exceptionnellement long qu'y passent les utilisateurs et utilisatrices âgés de 16 à 64 ans sur les réseaux sociaux (un peu plus de trois heures par jour au Kenya et un peu plus de quatre heures aux Philippines, alors que la moyenne mondiale est de deux heures et demie).¹⁴ Les États-Unis ont été inclus dans l'enquête technique (voir la section « Méthodologie de recherche » ci-après) afin de déterminer si TikTok a opéré des choix de conception différents dans un pays où le débat public sur les risques des réseaux sociaux pour les adolescents occupe le devant de la scène, et qui peut également être considéré comme un marché clé pour l'entreprise.

Bien que les recherches aient été axées en particulier sur les répercussions de la plateforme en matière de santé mentale, elles n'avaient pas vocation à prouver que TikTok a causé ou aggravé des problèmes de santé mentale chez les différents participant-e-s. En effet, cela aurait nécessité une connaissance approfondie de chaque facteur ayant une incidence sur la santé et le bien-être de ces personnes et l'opinion d'un-e spécialiste de la santé mentale, ce qui, pour la plupart des jeunes participant-e-s, n'aurait pas été possible en raison de l'accès extrêmement limité à des services de santé mentale gratuits ou tout du moins abordables (pour en savoir plus, voir la section 3.3). En associant des méthodes qualitatives et une étude technique du système de recommandation de TikTok, les recherches menées par Amnesty International aux fins du présent rapport permettent néanmoins de brosser un tableau suffisamment clair pour justifier que des mesures urgentes soient prises en vue d'atténuer ces risques.

2.3 MÉTHODOLOGIE DE RECHERCHE

Les recherches menées aux fins du présent rapport, qui se sont déroulées entre octobre 2022 et juillet 2023, comprenaient les volets suivants :

1. Une recherche documentaire approfondie, dont une étude des sources pertinentes issues du monde universitaire, des Nations unies, d'organisations non gouvernementales (ONG) et des médias, ainsi que d'autres sources secondaires.
2. Une étude de cadrage sous la forme d'un questionnaire en ligne, distribué sur les réseaux sociaux par Amnesty International et des organisations partenaires et auquel ont répondu 550 enfants et jeunes adultes âgés de 13 à 24 ans dans 45 pays entre octobre et novembre 2022. Le questionnaire portait sur l'utilisation des principales plateformes de réseaux sociaux, les expériences vécues sur ces plateformes, les éléments appréciés ou détestés, les réactions à des expériences négatives et les idées en faveur du changement, afin de mieux comprendre le vécu, les préoccupations et les comportements liés aux réseaux sociaux .
3. Cinq entretiens semi-structurés en présentiel et 20 discussions de groupe auxquels ont participé plus de 300 enfants et jeunes âgés de 14 à 24 ans (dont 180 étaient des utilisateurs et utilisatrices de TikTok) au Kenya (mars 2023) et aux Philippines (avril et mai 2023). Au Kenya,

13. UNESCO, "Thematic factsheet: youth and empowerment", 31 janvier 2023, <https://www.unesco.org/en/youth-and-empowerment>.

14. We Are Social, "Digital 2022 Global Overview Report", 26 janvier 2022, <https://wearesocial.com/hk/blog/2022/01/digital-2022-another-year-of-bumper-growth>. L'équipe de recherche d'Amnesty International n'a pas pu trouver de statistiques comparables à l'échelle mondiale qui concernent spécifiquement les enfants et les jeunes.

les entretiens en présentiel et les discussions de groupe avec des élèves, des jeunes militant-e-s, des étudiant-e-s et des jeunes diplômé-e-s se sont déroulés dans la capitale, Nairobi, ainsi que dans les villes de Kisumu et Mombasa et dans le comté de Machakos. Aux Philippines, les entretiens en présentiel ont eu lieu à Manille, la capitale, et dans la province de Batangas. En outre, neuf entretiens en ligne ont été organisés avec des enfants et des jeunes dans les régions des Philippines qui n'étaient pas accessibles à l'équipe de recherche pour des raisons pratiques ou des questions de sécurité. L'équipe de recherche d'Amnesty International s'est entretenue avec 37 enfants et jeunes aux Philippines, dont 34 ont évoqué leurs expériences sur TikTok ainsi que d'autres plateformes de réseaux sociaux.¹⁵

Au Kenya, l'organisation s'est appuyée sur le réseau « Écoles amies des droits humains » d'Amnesty International Kenya et sur ses entités locales pour trouver des jeunes à interroger. Aux Philippines, un appel à participation a été diffusé par l'intermédiaire du réseau Jeunes d'Amnesty International Philippines et grâce à des réseaux de services de soutien en matière de santé mentale et des organisations de jeunes faisant campagne en faveur de la santé mentale.

Cette démarche d'approche, l'accent mis sur les entretiens ou les discussions en très petits groupes et le nombre accru de jeunes adultes de notre cohorte aux Philippines ayant, dans une certaine mesure, accès à des services de soutien en matière de santé mentale ont probablement contribué à ce que bien davantage de participant-e-s signalent eux-mêmes des problèmes passés ou actuels de santé mentale que dans le groupe du Kenya, composé principalement d'adolescent-e-s. Les résultats des groupes de discussion organisés au Kenya en mars 2023 ont permis, par une approche plus ciblée, d'inviter, avec l'aide d'organisations de santé mentale et du réseau de jeunes militant-e-s d'Amnesty International, des participant-e-s qui avaient le sentiment que leur utilisation des réseaux sociaux avait eu des conséquences sur leur santé mentale.

Tous les entretiens et les groupes de discussion organisés par l'équipe de recherche d'Amnesty International se sont déroulés en anglais, avec parfois une interprétation lorsque les participant-e-s ont exprimé leurs pensées ou leur vécu en swahili et en dholuo (Kenya) ou en tagalog (Philippines). Les questions portaient sur leur utilisation de différentes plateformes de réseaux sociaux, leur expérience de ces plateformes, leurs opinions sur les démarches visant à réglementer l'utilisation des réseaux sociaux par les enfants, l'impact de leur utilisation des réseaux sociaux sur leur vie et les conséquences présumées de leur expérience des plateformes sur leur état émotionnel et leur santé mentale.

Le consentement des parents ou des tuteurs/tutrices a été demandé et obtenu pour tous les participant-e-s de moins de 18 ans et le consentement éclairé de chaque enfant a été recueilli. Les participant-e-s de plus de 18 ans ont donné leur consentement en connaissance de cause avant l'entretien ou la discussion en groupe. Afin de protéger l'identité des participant-e-s, nous utilisons des prénoms d'emprunt tout au long du rapport pour tous les participant-e-s autres que les spécialistes qui ont autorisé la publication de leur nom.

4. Quatorze entretiens avec des psychologues spécialisés dans l'adolescence, des responsables de campagne en faveur de la santé mentale (des jeunes), des spécialistes de la santé publique et des enseignant-e-s, au Kenya, aux Philippines et aux États-Unis, réalisés entre février et mai 2023.
5. Le fait de choisir pour sujet d'étude les enfants et les jeunes impose nécessairement des limites d'ordre éthique quant à la méthode employée pour déterminer comment les systèmes algorithmiques interagissent avec les utilisateurs et utilisatrices. Pour éviter d'exposer les jeunes

15. Amnesty International. "We are totally exposed": Young people share concerns about social media's impact on privacy and mental health in global survey, 7 février 2023, <https://www.amnesty.org/en/latest/news/2023/02/children-young-people-social-media-survey-2>.

utilisateurs et utilisatrices à des contenus dangereux ou de compromettre leur droit à la vie privée, tout en produisant les résultats les plus représentatifs possible, nous avons inclus dans l'étude une enquête technique en deux parties, effectuée en juin et juillet 2023 :

- Un audit algorithmique automatisé du système de recommandation qui alimente la page « Pour toi » de TikTok en utilisant des comptes simulant la participation d'adolescent-e-s de 13 ans au Kenya et aux États-Unis.
- Des captures d'écran effectuées sur trois comptes gérés manuellement, créés pour représenter des utilisateurs ou utilisatrices de 13 ans au Kenya, aux Philippines et aux États-Unis, afin de suivre l'amplification des contenus relatifs à la santé mentale sur la page « Pour toi » de TikTok.

Des informations complémentaires sur la méthode utilisée pour notre enquête technique sont fournies au chapitre 6.

6. Une lettre de recherche demandant des informations quant au processus de diligence requise en matière de droits humains, aux procédures de protection des données et aux mesures de promotion du bien-être et de la santé mentale des enfants et jeunes adultes parmi les utilisateurs et utilisatrices a été envoyée, ainsi qu'une lettre de droit de réponse présentant les principales conclusions de ce rapport. Les réponses de TikTok, datées du 12 juillet et du 29 octobre 2023, sont disponibles en Annexe 3 et 4.

3. CONTEXTE

3.1 PASSER À CÔTÉ DE L'ESSENTIEL : LACUNES EN MATIÈRE DE RECHERCHE ET OBSTACLES COMMERCIAUX ENTRAVANT L'ÉTUDE DES RISQUES QUE PRÉSENTENT LES PLATEFORMES DE RÉSEAUX SOCIAUX DU POINT DE VUE DES DROITS HUMAINS DES ENFANTS ET DES JEUNES

L'enquête judiciaire ouverte en octobre 2022 sur la mort de Molly Russell, une adolescente britannique, a été la première à désigner officiellement les réseaux sociaux comme ayant « probablement contribué » au décès d'une personne.¹⁶ Dans les mois précédant la mort de Molly, le fil Instagram de cette adolescente, qui a perdu la vie à l'âge de 14 ans, était alimenté en grande partie par du contenu portant sur l'automutilation, la dépression et le suicide, ou en faisant l'apologie.¹⁷ Les services médico-légaux ont conclu que sa mort était due à « un acte que la victime s'est infligé à elle-même du fait d'une dépression et des effets délétères de contenus en ligne », dont certains « tend[ai]ent à présenter l'automutilation et le suicide comme une conséquence inévitable d'un état dont il est impossible de se remettre », dans un contexte où les systèmes de recommandation proposaient à Molly une « vision limitée et irrationnelle sans aucune référence de normalité pour la contrebalancer¹⁸ ». Ils ont recommandé que l'État britannique et les entreprises de réseaux sociaux « se penchent sur l'utilisation des algorithmes dans le cadre de la fourniture de contenu¹⁹ ». Il s'agit, à ce jour, de l'une des mises en cause les plus fermes émanant d'une autorité publique dans le débat concernant l'impact des réseaux sociaux sur la santé mentale des adolescent·e·s.

Quoi qu'il en soit, les entreprises de réseaux sociaux sont réputées pour compliquer la tâche aux chercheurs et chercheuses indépendants qui tentent d'étudier leurs systèmes algorithmiques.²⁰ Amnesty International et d'autres organisations de la société civile ont demandé que des mesures

-
16. North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 octobre 2022, https://www.judiciary.uk/wp-content/uploads/2022/10/Molly-Russell-Prevention-of-future-deaths-report-2022-0315_Published.pdf.
 17. *Politico*, "Digital bridge: Platforms on the hook — Transatlantic AI rulebook — Let's talk data transfers", 6 octobre 2022, <https://www.politico.eu/newsletter/digital-bridge/platforms-on-the-hook-transatlantic-ai-rulebook-lets-talk-data-transfers/> ; une porte-parole de Meta qui s'est exprimée lors de l'enquête a déclaré que le contenu que Molly avait visionné était « nuancé et complexe » et qu'il était important de permettre aux utilisateurs et utilisatrices ayant des pensées suicidaires de s'exprimer en ligne. BBC, "Molly Russell: Instagram posts seen by teen were safe, Meta says". 26 septembre 2022, <https://www.bbc.co.uk/news/uk-england-london-63034300> ; pour consulter la réponse détaillée de Meta et les principales interrogations soulevées, voir Meta, "Response from Meta", 6 décembre 2022, <https://www.judiciary.uk/wp-content/uploads/2022/10/2022-0315-Response-from-META.pdf>.
 18. North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 octobre 2022 (déjà cité).
 19. BBC News, "Molly Russell: Coroner's report urges social media changes", 14 octobre 2022, <https://www.bbc.co.uk/news/uk-england-london-63254635>.
 20. Algorithm Watch, "AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook", 13 août 2021, <https://www.algorithmwatch.org/en/instagram-research-shut-down-by-facebook>. Algorithm Watch, "A guide to the EU's new rules for researcher access to platform data", 7 décembre 2022, <https://www.algorithmwatch.org/en/instagram-research-shut-down-by-facebook>.

réglementaires soient prises pour empêcher les entreprises d'entraver les recherches indépendantes et d'utilité publique sur le fonctionnement et l'impact des plateformes de réseaux sociaux, en formulant l'avertissement suivant :

« De nombreuses plateformes [de réseaux sociaux] font activement obstacle aux recherches dans leurs conditions générales, par des mesures techniques, ainsi que par des manœuvres d'intimidation ou en menaçant d'intenter des actions en justice. Plus particulièrement, les efforts des plateformes pour empêcher le moissonnage [collecte automatique de données à grande échelle] ont eu un effet dissuasif sur les chercheurs et chercheuses qui tentaient d'amener celles-ci à rendre des comptes.²¹»

Néanmoins, étant donné que le nombre d'utilisateurs et utilisatrices et le temps passé par les enfants sur les réseaux sociaux augmentent d'année en année, les articles de presse et le discours public sur le rôle de ces plateformes dans la vie des enfants se font de plus en plus critiques, dénonçant le tort que celles-ci peuvent causer aux jeunes. Cela est particulièrement notable en Europe et aux États-Unis, où le débat public a été alimenté par la divulgation en 2021 de documents internes, dits « Facebook Papers », qui contenaient des résultats d'enquête communiqués à l'entreprise en 2019, selon lesquels Instagram « aggrav[ait] les problèmes d'image corporelle chez une adolescente sur trois²²».

Avant 2021, il existait déjà un vif débat parmi les psychologues et les spécialistes de la santé publique sur la question de savoir si les réseaux sociaux contribuaient à la progression de la dépression et de l'automutilation chez les jeunes.²³ La plupart des études disponibles ont une limite principale : elles portent sur l'effet des réseaux sociaux en général ou, de manière encore plus globale, sur l'effet du temps d'écran chez les adolescent·e·s américains et européens. Il y a une absence criante d'éléments similaires pour d'autres parties du monde.²⁴ En réalité, les plateformes de réseaux sociaux diffèrent les unes des autres de manière importante et évoluent au fil du temps, tout comme les enfants et les jeunes. S'agissant des risques et de la réglementation, la question qu'il faudrait poser n'est pas de savoir si tous les réseaux sociaux nuisent à tous les enfants ni simplement à un enfant « moyen », dans l'hypothèse où une telle variable existerait, mais plutôt de savoir quels aspects particuliers des plateformes pourraient comporter des risques pour *certain*s utilisateurs et utilisatrices ou groupes d'utilisateurs/utilisatrices.

Les entreprises de réseaux sociaux mettent en avant les bienfaits de leurs plateformes pour la vie des enfants et des jeunes, auxquels adhèrent d'ailleurs de nombreux jeunes, y compris une grande partie de celles et ceux ayant participé aux recherches dans le cadre de ce rapport. Les jeunes

-
21. Le moissonnage peut présenter des risques considérables en matière de respect de la vie privée en fonction du contexte, du type de données recueillies et de leur utilisation. Par conséquent, il doit être limité. Cependant, Amnesty International et d'autres ont dénoncé l'interdiction totale du moissonnage imposée par les entreprises de médias sociaux comme un obstacle indu à la recherche sur les risques systémiques. En l'absence de cadres régissant l'accès aux données dans le contexte de la recherche, tels que celui qui sera bientôt mis en œuvre en Europe grâce aux nouvelles obligations incombant aux grandes plateformes en ligne en vertu du règlement de l'Union européenne (UE) sur les services numériques, des chercheurs et chercheuses n'ont parfois eu d'autre choix que de réaliser leurs études d'utilité publique en enfreignant les conditions de service des plateformes concernées. Mozilla Foundation, Amnesty International et autres, « Response to the European Commission's call for evidence for a Delegated Regulation on data access provided for in the Digital Services Act », mai 2023, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3423646_en. * Bien que le Règlement de l'UE sur les services numériques interdise les publicités ciblées à l'intention des mineurs. Pour de plus amples informations sur les pratiques de TikTok en matière de collecte de données et de publicité, voir Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 7 novembre 2023 (déjà cité).
 22. *The Guardian*, "Facebook aware of Instagram's harmful effect on teenage girls, leak reveals", 14 septembre 2021, <https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals>. Meta a rejeté les éléments contenus dans les « Facebook Papers », en opposant qu'ils étaient « fondés sur des documents mal définis et privés de tout contexte », voir *Washington Post*, "A whistleblower's power: Key takeaways from the Facebook Papers", 26 octobre 2021, <https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/>.
 23. Pour une analyse approfondie et actualisée, voir Jonathan Haidt et autres, "Social Media and Mental Health: A Collaborative Review", New York University, <https://docs.google.com/document/d/1w-HOfseF2wF9YlpXwUUtP65-olnkPyWcgF5BiAtBEyO/edit> (consulté le 12 septembre 2023).
 24. Sakshi Ghai et autres, "Social media and adolescent well-being in the Global South" (déjà cité) ; Sakshi Ghai, Luisa Fassi, Faisal Awadh et Amy Orben, "Lack of sample diversity in research on adolescent depression and social media use: A scoping review and meta-analysis", 7 février 2023, Clinical Psychological Science, <https://www.doi.org/10.1177/21677026221114859>.

participant-e-s vantaient le caractère divertissant de TikTok, la manière dont la plateforme encourage la créativité et le fait qu'elle leur donne un sentiment d'appartenance. De nombreuses études ont mis en évidence les avantages des communautés en ligne, constituées autour de réseaux sociaux, en particulier pour les jeunes marginalisés en quête de personnes qui ont un vécu similaire au leur, notamment les jeunes LGBTI et les jeunes souffrant de problèmes de santé mentale.²⁵ Cependant, ces avantages n'exemptent pas les entreprises de réseaux sociaux de leur responsabilité de repérer, d'évaluer, de prévenir et d'atténuer les dangers potentiels, et de remédier, si nécessaire et selon qu'il convient, aux préjudices réels qui découlent de l'utilisation de ces plateformes.

Toutefois, alors qu'elles font la promotion des avantages que procurent leurs produits, les entreprises de réseaux sociaux ne fournissent pas aux utilisateurs et utilisatrices des évaluations transparentes des risques. En outre, elles compliquent considérablement la tâche aux chercheurs et chercheuses indépendants qui tentent d'enquêter sur le fonctionnement de leurs plateformes et les risques potentiels qu'elles comportent pour les utilisateurs et utilisatrices.²⁶ En l'absence de la réglementation nécessaire pour permettre aux chercheurs et chercheuses de réaliser leurs études de manière efficace et dans le respect de la vie privée des utilisateurs et utilisatrices, des projets de recherche ont été étouffés et des chercheurs et chercheuses menacés d'une action en justice.²⁷ En outre, le rythme effréné de la plateforme vidéo de TikTok pose des problèmes supplémentaires en matière de recherche par rapport à d'autres plateformes car l'analyse à grande échelle de vidéos est plus difficile et nécessite davantage de ressources que pour des publications écrites.

3.2 ENTRE POLITIQUES ET POPULISME : DES EFFORTS LIMITÉS POUR RÉGLEMENTER LES PLATEFORMES DE RÉSEAUX SOCIAUX EFFORTS TO REGULATE SOCIAL MEDIA PLATFORMS

Dans ce contexte, il n'est peut-être pas surprenant que ce soient les recherches *internes* d'une entreprise de réseaux sociaux qui aient alimenté un discours public et des débats politiques plus critiques concernant l'impact de ces réseaux sur les enfants et les jeunes. Comme indiqué plus haut, les « Facebook Papers » dévoilés en 2021 contenaient des preuves accablantes du rôle d'Instagram dans le renforcement de l'image corporelle négative chez les adolescent-e-s.²⁸ Ces éléments et d'autres issus des documents divulgués ont fait la une des médias dans le monde entier et intensifié les pressions exercées sur les responsables politiques pour surmonter les obstacles de longue date qui empêchaient la modernisation des cadres réglementaires applicables aux plateformes, lesquels étaient alors en retard d'au moins deux décennies.²⁹

-
25. John Naslund, Ameya Bondre et autres, "Social media and mental health: Benefits, risks, and opportunities for research and practice", 20 avril 2020, *Journal of Technology in Behavioral Science*, volume 5, <https://www.doi.org/10.1007/s41347-020-00134-x>; Matthew N. Berger, Melody Taba et autres, "Social media use and health and well-being of lesbian, gay, bisexual, transgender, and queer youth: Systematic review", 21 septembre 2022, *Journal of Medical Internet Research*, volume 24, numéro 9, <https://www.jmir.org/2022/9/e38449>.
 26. Pour en savoir plus sur les recommandations d'Amnesty International et d'autres organisations de la société civile sur l'identification et l'atténuation des risques systémiques, voir People vs Big Tech, "Briefing: Fixing Recommender Systems: From identification of risk factors to meaningful transparency and mitigation", 25 août 2023, <https://www.peoplevsbig.tech/briefing-fixing-recommender-systems-from-identification-of-risk-factors-to-meaningful> (consulté le 25 août 2023).
 27. Algorithm Watch, "AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook", 13 août 2021, <https://www.algorithmwatch.org/en/instagram-research-shut-down-by-facebook>; Algorithm Watch, "A guide to the EU's new rules for researcher access to platform data", décembre 2022, <https://www.algorithmwatch.org/en/dsa-data-access-explained/>.
 28. *The Guardian*, "Facebook aware of Instagram's harmful effect on teenage girls, leak reveals", septembre 2021 (déjà cité).
 29. Le principal texte législatif des États-Unis régissant la responsabilité des plateformes en ligne, la Loi relative à la décence des communications, a été promulgué en 1996. Avant le DSA de 2022, l'UE s'était dotée en 2000 d'une directive sur le commerce électronique.

Néanmoins, peu de progrès ont été accomplis s'agissant des risques systémiques liés aux grandes plateformes de médias sociaux, et peu d'États ou d'organisations régionales ont adopté une législation en la matière. Ni le Kenya ni les Philippines, deux des pays sur lesquels porte ce rapport, n'ont encore adopté de législation en ce sens.³⁰

Parmi les quelques États ou organisations régionales ayant pris des mesures figure l'UE, dont le Règlement sur les services numériques (DSA), adopté en juillet 2022, est devenu la première législation importante de niveau régional sur les « géants de la haute technologie ». Cet instrument vise à limiter les effets délétères des plateformes de réseaux sociaux, notamment en interdisant les « publicités ciblées » intrusives à l'intention des enfants.³¹ Il impose aux grandes plateformes en ligne, comme TikTok, d'évaluer les risques systémiques qu'elles présentent, notamment en matière de santé publique et vis-à-vis des enfants, de prendre des mesures pour atténuer ces risques et de se soumettre à des audits indépendants visant à contrôler dans quelle mesure elles se conforment à ces obligations (bien que nombre de questions demeurent quant à la nature précise et à l'application de ces mesures³²). Aux termes du DSA, TikTok et les autres très grandes plateformes sont également tenues de proposer aux utilisateurs et utilisatrices au moins un fil qui ne se fonde pas sur le profilage. Amnesty International estime que cette approche reposant sur une démarche volontaire de la part des utilisateurs et utilisatrices est une « occasion manquée » car modifier le paramétrage de façon à protéger davantage la vie privée est souvent un processus fastidieux et les personnes concernées ont tendance à conserver les réglages par défaut.³³

L'efficacité du DSA reste encore à prouver : certaines des obligations qu'il impose aux très grandes plateformes en ligne comme la présentation d'évaluations des risques ont pris effet en août 2023, mais le DSA n'entrera pleinement en vigueur qu'en 2024. Bien que cet instrument vise à atténuer certains risques liés au modèle d'activité fondé sur la surveillance qu'ont adopté les entreprises de réseaux sociaux, par exemple en prohibant la publicité ciblée à l'intention des enfants, il n'interdit pas ce modèle d'activité en tant que tel ; il est donc nécessairement limité. Néanmoins, il est le texte le plus ambitieux à ce jour en matière de réglementation des plateformes. Bien que le champ d'application du DSA soit limité aux activités des plateformes dans l'UE, on peut néanmoins espérer qu'il ait des répercussions positives au-delà, étant donné que les changements introduits au sein de l'UE peuvent être plus facilement étendus ou reproduits dans des propositions de loi concernant d'autres pays.

Au Royaume-Uni, la Loi relative à la sécurité en ligne, adoptée en septembre 2023, impose de nouvelles obligations aux plateformes en vue de créer un environnement en ligne plus sûr, avec des protections particulières pour les moins de 18 ans.³⁴ Ce texte a été élaboré pour renforcer la protection de la vie privée et les obligations visant à limiter la collecte de données personnelles qui figuraient dans

30. Toutefois, ces deux États réglementent la collecte et l'utilisation de données personnelles, y compris au moyen de services en ligne. Le rapport qui accompagne celui-ci, intitulé *I feel exposed: Caught in TikTok's Surveillance Web*, contient une analyse de la législation kenyane dans ce domaine. Les Philippines ont adopté une réglementation sur la confidentialité des données (Data Privacy Act) en 2012, laquelle comprend des principes également établis dans le Règlement général sur la protection des données (RGPD) de l'UE, tels que la limitation de la finalité et l'obligation de recueillir le moins de données possible. Voir International Association of Privacy Professionals, "GDPR matchup: The Philippines' Data Privacy Act and its Implementing Rules and Regulations", juillet 2017, <https://iapp.org/news/a/gdpr-matchup-the-philippines-data-privacy-act-and-its-implementing-rules-and-regulations/>.

31. UE, Règlement 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (Règlement sur les services numériques), https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3A%3A2022%3A277%3ATOC&uri=uriserv%3A0J.L_.2022.277.01.0001.01.FRA.

32. Pour une analyse détaillée du Règlement sur les services numériques, voir Amnesty International, *Adoption de la loi de l'Union européenne sur les services numériques : quelles implications pour les droits humains et pour les modèles économiques aux effets nocifs des géants du numérique* (index : POL 30/5830/2022), 7 juillet 2022, <https://www.amnesty.org/fr/documents/pol30/5830/2022/fr> ; Euractiv, "Europe enters patchy road to audit online platforms' algorithms", 7 juillet 2023, <https://www.euractiv.com/section/platforms/news/europe-enters-patchy-road-to-audit-online-platforms-algorithms/>.

33. Alfred Ng, "Default settings for privacy – we need to talk", 21 décembre 2019, <https://www.cnet.com/tech/tech-industry/default-settings-for-privacy-we-need-to-talk/>.

34. Ministère du numérique, de la culture, des médias et du sport du Royaume-Uni, Online Safety Bill (telle que modifiée en commission), <https://bills.parliament.uk/bills/3137>.

le Code britannique des droits de l'enfant, qui n'autorisait initialement les services en ligne à recourir au « profilage » pour les enfants que s'ils étaient « dotés de mesures adéquates pour protéger l'enfant de tout effet délétère (en particulier le fait de se voir proposer du contenu préjudiciable à sa santé ou son bien-être³⁵) ». La Loi relative à la sécurité en ligne érige en infraction le fait de publier du contenu qui encourage « des actes graves d'automutilation³⁶ ». Elle impose aux plateformes d'empêcher les enfants d'accéder à du contenu qui « sans atteindre le seuil correspondant à une qualification pénale, promeut ou encourage le suicide, l'automutilation ou les troubles alimentaires ou en donne le “mode d'emploi”³⁷ », ce qui a suscité une importante controverse. Les Samaritains et un certain nombre d'autres organisations de santé mentale estimaient que le fait de limiter ces règles aux enfants était trop restrictif, tandis que d'autres groupes ont mis en garde contre le fait d'inciter les plateformes à réduire indûment la liberté d'expression des utilisateurs et utilisatrices pour tenter de se conformer aux mesures qui prescrivent de restreindre le contenu « légal mais dangereux³⁸ ».

Aux États-Unis, la nouvelle réglementation se trouve dans une impasse au niveau fédéral.³⁹ Néanmoins, plusieurs États ont adopté des lois réglementant les espaces en ligne. La Californie, par exemple, s'est dotée d'une loi relative à la prise en compte de l'âge dans la conception des plateformes, qui obligera celles « susceptibles d'être utilisées par des enfants » à offrir à ces derniers un degré de confidentialité plus élevé dans les réglages par défaut et interdira d'employer les informations relatives à un enfant « d'une manière dont l'entreprise sait, ou a toutes les raisons de savoir, qu'elle est matériellement préjudiciable à la santé physique, à la santé mentale ou au bien-être d'un enfant⁴⁰ ». Ce texte a cependant été bloqué par un juge fédéral en septembre 2023 en attendant la tenue du procès intenté par un groupe du secteur technologique, ce dernier estimant que la loi est contraire au premier amendement.⁴¹ Ce texte, à l'instar du DSA, obligera les entreprises à réaliser une évaluation de l'impact en matière de protection des données non publiques afin de déterminer si des éléments de conception ou des algorithmes de la plateforme pourraient représenter un danger pour des enfants. Bien que de nombreux défenseur-e-s des droits des enfants aient réservé un bon accueil à la loi, qu'ils considèrent comme un pas vers un espace en ligne plus sûr,⁴² les défenseur-e-s du respect de la vie privée et de la liberté d'expression ont soulevé des problèmes de confidentialité s'agissant des répercussions négatives que pourrait avoir un usage élargi des mesures de vérification de l'âge.⁴³

Entre-temps, l'Utah a promulgué en mars 2023 une loi qui confère aux parents un accès et un contrôle sans précédent en ce qui concerne les comptes de leurs enfants sur les réseaux sociaux, ce qui suscite des inquiétudes au sujet du respect de la vie privée des enfants. Les défenseur-e-s des

-
35. Information Commissioner's Office (ICO), Age-appropriate design code, <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/code-standards/> (consulté le 13 juillet 2023).
 36. Ministère du numérique, de la culture, des médias et du sport du Royaume-Uni, Online Safety Bill (déjà cité).
 37. Ministère du numérique, de la culture, des médias et du sport du Royaume-Uni, "A Guide to the Online Safety Bill", 16 décembre 2022, <https://www.gov.uk/guidance/a-guide-to-the-online-safety-bill>.
 38. *The Guardian*, "Online safety bill must protect adults from self-harm content, say charities", 14 octobre 2022, <https://www.theguardian.com/technology/2022/oct/14/online-safety-bill-must-protect-adults-self-harm-content-charities-samaritans>.
 39. Parmi les différents projets de loi présentés au niveau fédéral en vue de la réglementation des plateformes, seule la Loi relative à la sécurité des enfants en ligne (KOSA) semble pouvoir être adoptée prochainement. La KOSA associe des éléments du Code de conception californien et du DSA européen, en ce qu'elle oblige toutes les grandes plateformes susceptibles d'être utilisées par des personnes de moins de 17 ans à publier des rapports sur leurs activités de mise en conformité, afin d'appliquer des « mesures dans [leur] conception et le fonctionnement de [leurs] produits et services pour prévenir et atténuer » les préjudices et proposer aux jeunes des moyens de se désinscrire des fils basés sur des recommandations personnalisées. Voir Congress.gov, Text - S.1409 - 118th Congress (2023-2024): Kids Online Safety Act, 2 mai 2023, <https://www.congress.gov/bills/118th-congress/senate-bill/1409/text>.
 40. Alexander Misakian et Tiffany Young, "California enacts the California Age-Appropriate Design Code Act", 20 septembre 2022, <https://www.foley.com/en/insights/publications/2022/09/california-enacts-age-appropriate-design-code-act>.
 41. The Verge, "Court blocks California's online child safety law", 18 septembre 2023, <https://www.theverge.com/2023/9/18/23879489/california-age-appropriate-design-code-act-blocked-unconstitutional-first-amendment-injunction>.
 42. Voir par exemple : Fairplay, "Josh Golin statement regarding California Age Appropriate Design Code", 23 mai 2022, <https://fairplayforkids.org/josh-golin-ca-aadc/>.
 43. Iapp, "California Age Appropriate Design Code final passage brings mixed reviews", 31 août 2022, <https://iapp.org/news/a/california-age-appropriate-design-code-final-passage-brings-mixed-reviews/>.

droits des enfants ont émis des doutes quant à la question de savoir si transférer la responsabilité aux parents permettrait de protéger les enfants des dangers systémiques.⁴⁴ L'Arkansas et la Louisiane ont malgré tout emboîté le pas à l'Utah en adoptant des textes de loi similaires début 2023.⁴⁵ En mai 2023, le Montana est devenu le premier État des États-Unis à interdire TikTok pour d'autres motifs liés à la sécurité des données, compte tenu de l'origine chinoise de la plateforme, bien que des spécialistes se demandent dans quelle mesure une interdiction à l'échelle d'un État est techniquement réalisable.⁴⁶

La Chine, à son tour, a imposé des contrôles à ses grandes entreprises technologiques, notamment ByteDance, fondateur de TikTok et de sa version chinoise, Douyin. Les Règles relatives à l'administration des algorithmes de recommandation des services d'information en ligne, entrées en vigueur en mars 2022, associent des exigences de transparence, le contrôle des utilisateurs et utilisatrices et des obligations en matière d'évaluation des risques – qui ne sont pas sans rappeler les dispositions du DSA de l'UE – à la surveillance à caractère répressif que la Chine exerce sur les médias, prescrivant ainsi que les systèmes de recommandation algorithmiques « présentent des informations conformes aux valeurs communes » et « préviennent ou réduisent les controverses et les différends⁴⁷ ». Afin que les plateformes ne « mettent pas en danger la sécurité nationale ni l'intérêt social de la population », l'Administration chinoise du cyberspace dispose d'un large accès aux informations sur les systèmes algorithmiques des plateformes, bien que les entretiens entre l'Administration et ByteDance aient, semble-t-il, mis en évidence un manque de compétences techniques au niveau des fonctionnaires chargés de comprendre et d'interroger les données.⁴⁸ Probablement en réaction à une campagne plus générale des pouvoirs publics visant à réduire le temps consacré par les enfants aux divertissements en ligne, la plateforme Douyin a elle-même imposé une limite de 40 minutes par jour aux enfants de moins de 14 ans en 2021.⁴⁹

D'autres pays comme le Brésil, l'Inde et la Turquie ont tenté de limiter le pouvoir des plateformes des « géants technologiques » dans le but affiché d'endiguer la propagation de contenu illégal et de fausses informations. Cependant, des militant-e-s de la société civile et des journalistes ont soulevé de graves préoccupations au sujet de cette mainmise du gouvernement et de l'impact de telles mesures sur la liberté d'expression en ligne.⁵⁰

À l'échelle mondiale, il est nécessaire et urgent de continuer à tracer des voies pour gérer les risques systémiques exposés ici, ce dans le respect des droits humains.

-
44. NPR, "Utah's new social media law means children will need approval from parents", 24 mars 2023, <https://www.npr.org/2023/03/24/1165764450/utahs-new-social-media-law-means-children-will-need-approval-from-parents>.
 45. CNN, "Arkansas governor signs sweeping bill imposing a minimum age limit for social media usage", 12 avril 2023, <https://edition.cnn.com/2023/04/12/tech/arkansas-social-media-age-limit/index.html> ; CNN, "Louisiana lawmakers approve parental consent bill for kids' social media use and other online services", 8 juin 2023, <https://edition.cnn.com/2023/06/08/tech/louisiana-parental-consent-bill-social-media/index.html>.
 46. NPR, "Montana banned TikTok. Whatever comes next could affect the app's fate in the U.S.", 18 mai 2023, <https://www.npr.org/2023/05/18/1176940592/montana-ban-tiktok-lawsuit-constitution>.
 47. Friedrich Ebert Stiftung, *China's Regulations on Algorithms: Contexts, impacts and comparisons with the Union européenne*, janvier 2023, <https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf>.
 48. Matt Sheehan et Sharon Du, "What China's Algorithm registry reveals about AI governance", 9 décembre 2022, <https://carnegieendowment.org/2022/12/09/what-china-s-algorithm-registry-reveals-about-ai-governance-pub-88606>.
 49. BBC, "China: Children given daily time limit on Douyin – its version of TikTok", 20 septembre 2021, <https://www.bbc.co.uk/news/technology-58625934>.
 50. Global Voices, "How India's new internet regulations will change social media, online news and video streaming", 16 mars 2021, <https://globalvoices.org/2021/03/16/how-indias-new-internet-regulations-will-change-social-media-online-news-and-video-streaming/>; CNN, "Critics fear Turkey's new social media law could hurt freedom of expression. Here's how", 29 juillet 2020, <https://edition.cnn.com/2020/07/29/europe/turkey-social-media-law-intl/index.html> ; MIT Technology Review, "Brazil's 'fake news' bill won't solve its misinformation problem", 10 septembre 2020, <https://www.technologyreview.com/2020/09/10/1008254/brazil-fake-news-bill-misinformation-opinion>.

3.3 AGGRAVATION DES PROBLÈMES DE SANTÉ MENTALE CHEZ LES ENFANTS ET LES JEUNES

Le débat concernant les effets des réseaux sociaux sur la santé et le bien-être des enfants et des jeunes est sous-tendu par la progression avérée de l'anxiété, de la dépression et de l'automutilation chez les adolescent·e·s aux États-Unis et dans d'autres pays anglophones dans les années 2010, laquelle s'est accélérée pendant la pandémie de COVID-19.⁵¹ Bien que les éléments attestant une tendance similaire dans d'autres régions du monde soient limités, en partie faute de données, l'ampleur des besoins non satisfaits en matière de santé mentale dans le monde est de plus en plus reconnue.⁵² Selon la base de données Global Health Data Exchange, environ une personne âgée de 10 à 19 ans sur sept dans le monde souffre de problèmes de santé mentale, et l'Organisation mondiale de la santé (OMS) a signalé que l'anxiété et la dépression étaient particulièrement répandues.⁵³ Les spécialistes de la santé publique estiment que la proportion de jeunes présentant des problèmes de santé mentale est probablement plus haute en Afrique subsaharienne que dans les pays à revenu élevé, bien que le manque de données représentatives pour certaines parties de la région et les déclarations probablement inférieures à la réalité en raison de l'absence de services abordables et de la stigmatisation sociale entravent la compréhension et la prise en charge de la question.⁵⁴

Au Kenya et aux Philippines, où Amnesty International a réalisé des études qualitatives aux fins du présent rapport, l'accès à des services de santé mentale gratuits ou tout du moins abordables demeure très limité et les structures qui les prêtent se situent souvent dans des grandes villes ou à proximité.⁵⁵ Aux Philippines, des jeunes ayant participé à l'étude ont déclaré que, bien qu'ils/elles aient pu dans certains cas voir leur pathologie mentale diagnostiquée, ils/elles n'avaient pas les moyens de se procurer les médicaments prescrits et/ou avaient des difficultés à trouver des thérapeutes qui puissent assurer leur prise en charge à long terme.⁵⁶

Bien que les soins en matière de santé mentale soient davantage disponibles aux États-Unis, les psychiatres et thérapeutes spécialisés dans la prise en charge des adolescent·e·s peinent, semble-t-il, à satisfaire la demande accrue de soins pour des troubles de l'humeur et des pensées suicidaires.⁵⁷ D'après une enquête menée par les Centres de contrôle et de prévention des maladies (CDC), 30 % des lycéennes aux États-Unis « ont envisagé sérieusement de se suicider » en 2021 (contre 19 % en 2011) et 13 % ont déclaré avoir fait une tentative de suicide.⁵⁸ Dans ce pays, les services d'urgence ont signalé deux fois plus de troubles alimentaires chez les adolescentes en 2022 qu'avant la pandémie.⁵⁹ Dans tout l'ensemble de données des CDC, les violences sexuelles, la tristesse et le risque de suicide étaient plus fréquents chez les jeunes femmes et les adolescent·e·s LGBTI⁶⁰ (lesbiennes,

-
51. Pour une présentation générale des éléments disponibles : Jonathan Haidt, Zach Rausch et Jean Twenge (en cours). *Social Media and Mental Health: A Collaborative Review* (déjà cité).
 52. Organisation mondiale de la santé, "World mental health report: Transforming mental health for all", juin 2022, [who.int/publications/item/9789240049338](https://www.who.int/publications/item/9789240049338).
 53. Organisation mondiale de la santé, "Factsheet: Mental health of adolescents" (déjà cité).
 54. Elsabe Brits, "High mental health burden for Africa's youth", 20 octobre 2021, <https://www.nature.com/articles/d44148-021-00097-y> ; Ismail Temitayo Gbadamosi, Isaac Tabiri Henneh et autres, "Depression in Sub-Saharan Africa", 17 mars 2022, IBRO Neuroscience Reports, volume 12, pubmed.ncbi.nlm.nih.gov/35746974/.
 55. Constatation fondée sur les conversations avec des organisations de santé mentale, des psychologues et des organisations faisant campagne en faveur de la santé mentale au Kenya et aux Philippines, notamment In Touch Community Services, le docteur Jerome Cleofas et le docteur Marc Reyes, entre mars et mai 2023.
 56. Entretien avec une jeune femme de 20 ans à Manille, 5 mai 2023 ; entretien avec un jeune homme trans de 23 ans à Manille, 6 mai 2023 ; groupe de discussion, notamment avec une jeune femme de 22 ans, à Manille, 6 mai 2023.
 57. *New York Times Magazine*, "How do you actually help a suicidal teen?", 17 mars 2023, <https://www.nytimes.com/2023/05/17/magazine/suicide-teens.html>.
 58. CDC, "CDC report shows concerning increases in sadness and exposure to violence among teen girls and LGBQ+ youth", mis à jour le 9 mars 2023, <https://www.cdc.gov/nchhstp/newsroom/fact-sheets/healthy-youth/sadness-and-violence-among-teen-girls-and-LGBQ-youth-factsheet.html>.
 59. CDC Morbidity and Mortality Weekly Report, "Pediatric emergency department visits associated with mental health conditions before and during the COVID-19 pandemic — United States, January 2019-January 2022", 25 février 2022, Morbidity and Mortality Weekly Report, volume 21, numéro 8, <https://dx.doi.org/10.15585/mmwr.mm7108e2>.

gays et personnes bisexuelles ou intersexes) que chez les jeunes hommes et les adolescent-e-s hétérosexuels.⁶¹

Toutes ces données soulignent qu'il est urgent de s'intéresser de plus près à l'utilisation des réseaux sociaux par les jeunes du monde entier présentant des problèmes de santé mentale et aux répercussions des stratégies commerciales de maximisation de la participation sur ce groupe d'utilisateurs et utilisatrices vulnérables.

60. Aucune donnée sur les personnes transgenres n'a été prise en compte dans l'enquête.

61. CDC, "Pediatric emergency department visits associated with mental health conditions before and during the COVID-19 pandemic — United States, January 2019-January 2022", 25 février 2022, Morbidity and Mortality Weekly Report (déjà cité).

4. CADRE DE PROTECTION DES DROITS HUMAINS

Dans son rapport de 2019 intitulé *Les géants de la surveillance. Le modèle économique de Facebook et Google menace les droits humains*, Amnesty International démontrait que Meta (anciennement Facebook) et Google (dont le propriétaire est Alphabet) ont construit un modèle économique reposant sur la violation du droit de leurs utilisateurs et utilisatrices au respect de la vie privée, faisant de la collecte de données personnelles de milliards de personnes un moyen de générer des revenus publicitaires.⁶² Depuis, des centaines de millions d'autres utilisateurs et utilisatrices ont rejoint de nouvelles plateformes de réseaux sociaux qui emploient le même modèle économique fondé sur la surveillance pour cibler et exploiter le public dont la présence en ligne augmente le plus rapidement : les enfants et les jeunes.⁶³

Le rapport de 2023 d'Amnesty International intitulé *I feel exposed: Caught in TikTok's Surveillance Web* explique que TikTok recueille et analyse les données des enfants et des jeunes afin d'identifier une base d'utilisateurs et utilisatrices qui peut être la cible de publicités personnalisées et d'étendre cette base.⁶⁴ Il analyse en détail la manière dont ce modèle économique fondé sur la surveillance enfreint le droit au respect de la vie privée, entravant la capacité des personnes à contrôler leurs informations personnelles, ce qui bafoue le droit à la liberté de pensée, particulièrement le droit de ne pas rendre publiques ses idées et opinions.

Un grand nombre de ces préoccupations sont liées à des questions abordées dans ce rapport, notamment le fait que TikTok utilise la richesse des données personnelles collectées et les conclusions que l'entreprise en tire pour envoyer du contenu ciblé à des utilisateurs et utilisatrices afin de veiller à ce qu'ils continuent d'utiliser la plateforme, ainsi que le fait que son système algorithmique de recommandation de contenu peut exacerber les risques pour les enfants et les jeunes adultes souffrant de problèmes de santé mentale préexistants.

Ce chapitre résume le cadre de protection des droits humains applicable en ce qui concerne ces préoccupations.

62. Amnesty International, *Les géants de la surveillance* (déjà cité).

63. Union internationale des télécommunications, « Faits et chiffres 2022 : le point sur la connectivité mondiale dans un contexte de récession économique », 30 novembre 2022, <https://www.itu.int/hub/2022/11/facts-and-figures-2022-global-connectivity-statistics> ; *Forbes*, "What The Rise Of TikTok Says About Generation Z", 7 juillet 2020, <https://www.forbes.com/sites/forbestechcouncil/2020/07/07/what-the-rise-of-tiktok-says-about-generation-z/>

64. TikTok permet aux publicitaires d'envoyer des publicités ciblées à des utilisateurs et utilisatrices de tout juste 13 ans dans tous les États couverts par la politique de respect de la vie privée de TikTok, à l'exception de l'Union européenne, la région la plus réglementée, et deux autres pays (la Suisse et le Royaume-Uni). Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 2023 (déjà cité).

4.1 LE DROIT AU RESPECT DE LA VIE PRIVÉE À L'ÈRE DES RÉSEAUX SOCIAUX

« La protection de la vie privée est essentielle pour le pouvoir d'action, la dignité et la sécurité des enfants et pour l'exercice des droits de l'enfant », Comité des droits de l'enfant, Observation générale n° 25 (2021) sur les droits de l'enfant en relation avec l'environnement numérique.⁶⁵

Bien que le droit au respect de la vie privée soit reconnu par de nombreux instruments, il n'a probablement jamais été aussi menacé que par la montée des réseaux sociaux maintenant largement dominants dans le monde. Le droit au respect de la vie privée consacre le droit de chacun-e de ne pas faire l'objet « d'immixtions arbitraires ou illégales dans sa vie privée, sa famille, son domicile ou sa correspondance », ce droit devant être protégé par la loi.⁶⁶ Le principal instrument de protection des droits des enfants, la Convention relative aux droits de l'enfant, prévoit que « [l]'enfant a droit à la protection de la loi contre de telles immixtions ou de telles atteintes⁶⁷ ». Reconnaisant que l'innovation technologique a créé de nombreuses nouvelles menaces pour le droit au respect de la vie privée depuis la création de ces instruments, le Comité des droits de l'homme a déclaré que ces protections comprennent « le rassemblement et la conservation, par des autorités publiques, des particuliers ou des organismes privés, de renseignements concernant la vie privée d'individus sur des ordinateurs, dans des banques de données et selon d'autres procédés⁶⁸ ».

Le Haut-Commissariat des Nations unies aux droits de l'homme (HCDH) a également précisé la définition du respect de la vie privée dans le contexte de l'environnement numérique :

“Privacy can be considered as the presumption that individuals should have an area of « Le respect de la vie privée peut être considéré comme la présomption que tout individu a droit à un espace dans lequel il peut s'épanouir, interagir et jouir d'une liberté en toute autonomie, une “sphère privée” dans laquelle il est libre d'interagir ou non avec d'autres personnes et peut échapper à l'intervention de l'État et à toute intervention excessive non sollicitée d'une tierce partie. Dans l'environnement numérique, la confidentialité des renseignements personnels, c'est-à-dire de l'information qui existe ou qui peut être extrapolée au sujet d'une personne, de sa vie et des décisions fondées sur cette information, revêt une importance particulière.⁶⁹”

Le droit au respect de la vie privée englobe trois notions qui sont étroitement liées : le droit de chaque personne de ne pas subir d'immixtion dans sa vie privée, le droit de contrôler les informations la concernant, et le droit à un espace permettant d'exprimer librement son identité. Amnesty International estime que la nature du modèle économique de TikTok fondé sur la surveillance entrave chacun de ces trois éléments dans une mesure qui bafoue la nature même de la vie privée.

LE DROIT DE CHAQUE PERSONNE DE NE PAS SUBIR IMMIXTION DANS SA VIE PRIVÉE

La haute-commissaire aux droits de l'homme a reconnu que « [l]e simple fait de produire ou de collecter des données relatives à l'identité, à la famille ou à la vie privée d'une personne porte déjà atteinte au droit à la vie privée, car de tels actes ont pour effet de faire perdre à la personne concernée

65. Comité des droits de l'enfant, Observation générale n° 25 sur les droits de l'enfant en relation avec l'environnement numérique, 2 mars 2021, UN Doc. CRC/C/GC/25, § 67.

66. Article 12 de la Déclaration universelle des droits de l'homme (DUDH) et article 17 du Pacte international relatif aux droits civils et politiques (PIDCP).

67. Article 16 de la Convention relative aux droits de l'enfant.

68. Comité des droits de l'homme des Nations unies, Observation générale n° 16 : Droit au respect de la vie privée, de la famille, du domicile et de la correspondance et protection de l'honneur et de la réputation (article 17), 8 avril 1988, § 10.

69. Rapport du Haut-Commissaire des Nations Unies aux droits de l'homme, Le droit à la vie privée à l'ère du numérique, 3 août 2018, UN Doc. A/HRC/39/29, § 5.

une part du contrôle qu'elle exerce sur des informations qui pourraient constituer un risque pour sa vie privée.⁷⁰» La protection du droit au respect de la vie privée couvre non seulement le contenu des communications, « mais aussi [les] métadonnées [données liées à des fichiers numériques, des communications, etc. comme les données d'horodatage ou de localisation] car, lorsqu'elles sont analysées et agrégées, ces données "peuvent donner des indications sur la conduite d'un individu, ses relations sociales, ses préférences privées et son identité qui vont bien au-delà de ce que l'on obtient en accédant au contenu d'une communication privée"⁷¹».

L'immixtion dans la vie privée d'une personne n'est autorisée par le droit international relatif aux droits humains que si elle n'est ni arbitraire, ni illégale. Les mécanismes de protection des droits humains ont systématiquement interprété ces dispositions comme renvoyant aux principes généraux de légalité, de nécessité et de proportionnalité.⁷² Cependant, le Comité des droits de l'enfant a conclu que « [I]es pratiques numériques telles que le traitement automatisé des données, le profilage, le ciblage comportemental, la vérification obligatoire de l'identité, le filtrage des informations et la surveillance de masse, deviennent courantes » dans la vie des enfants.⁷³ Comme l'explique plus en détail le rapport *I feel exposed: Caught in TikTok's Surveillance Web*, les plateformes de réseaux sociaux dominantes, dont TikTok, instaurent des modes de surveillance de masse exercée par des entreprises qui sont par nature injustifiés et disproportionnés et ne peuvent en aucun cas constituer une immixtion autorisée dans la vie privée.⁷⁴

LE DROIT DE CONTRÔLER SES DONNÉES PERSONNELLES

Le deuxième élément du droit au respect de la vie privée prévoit que les personnes ont le droit de contrôler leurs données personnelles, ce que l'on nomme également l'« auto-détermination informationnelle⁷⁵», afin de décider quand et comment leurs données personnelles peuvent être partagées avec d'autres. Cela constitue le fondement de la réglementation sur la protection des données. La Cour européenne des droits de l'homme a conclu que la protection des données à caractère personnel joue un rôle fondamental pour l'exercice du droit au respect de la vie privée⁷⁶ et que le respect de la vie privée consacre donc le droit à une forme d'auto-détermination informationnelle.⁷⁷

LE DROIT À UN ESPACE PERMETTANT D'EXPRIMER LIBREMENT SON IDENTITÉ

Un large consensus a enfin été atteint quant au fait que le respect de la vie privée est également essentiel pour la création et la protection de l'espace nécessaire à la construction de nos identités.⁷⁸ Le Comité des droits de l'homme des Nations unies a défini la vie privée comme « une sphère de la vie d'une personne où elle peut exprimer librement son identité⁷⁹». Cela reflète une reconnaissance

70. Rapport du Haut-Commissaire des Nations Unies aux droits de l'homme : Le droit à la vie privée à l'ère du numérique, 3 août 2018 (déjà cité), § 7.

71. Rapport du Haut-Commissaire des Nations Unies aux droits de l'homme : Le droit à la vie privée à l'ère du numérique, 3 août 2018 (déjà cité), § 6.

72. Rapport du Haut-Commissariat des Nations Unies aux droits de l'homme : Le droit à la vie privée à l'ère du numérique, doc. ONU A/HRC/27/37, 30 juin 2014, § 21-27.

73. Comité des droits de l'enfant, Observation générale no 25 (déjà cité), § 68.

74. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web* (déjà cité).

75. Le terme « auto-détermination informationnelle » a été employé pour la première fois par la Cour constitutionnelle allemande dans le cadre d'une décision sur les données personnelles recueillies pendant le recensement de 1983. La Cour avait conclu qu'il revient à chaque personne de décider quand et dans quelles limites les informations sur sa vie privée doivent être communiquées à d'autres. Cour constitutionnelle fédérale, Arrêt du 15 décembre 1983, 1 BvR – Rn. 1-215, https://www.bverfg.de/e/rs19831215_1bvr020983.html

76. Cour européenne des droits de l'homme, *S et Marper c. Royaume-Uni*, requêtes 30562/04 et 30566/04, Arrêt de la Grande chambre, 4 décembre 2008, <https://hudoc.echr.coe.int/eng#%7B%22itemid%22%3A%22001-90052%22%7D>

77. Cour européenne des droits de l'homme, *Satakunnan Markkinapörssi Oy et Satamedia Oy c. Finlande*, requête 931/13, Arrêt de la Grande chambre, 27 juin 2017, <https://hudoc.echr.coe.int/eng#%7B%22itemid%22%3A%22001-175218%22%7D>, § 137.

78. Voir par exemple Philip Agre and Marc Rotenberg (éditeurs), *Technology and Privacy: The New Landscape*, 1998.

79. Décision du Conseil des droits de l'homme, *Coeriel et Aurik c. the Netherlands*, 31 décembre 1994, doc. ONU CCPR/C/52/D/453/1991, § 10.2.

que le sentiment d'identité d'une personne est socialement construit et dynamique : les personnes découvrent et dévoilent différentes facettes d'elles-mêmes dans différents contextes, que ce soit avec des ami-e-s, à l'école, au travail ou en public, et ces identités évoluent et s'adaptent en permanence.

Les plateformes de réseaux sociaux comme Instagram et TikTok exercent une influence considérable sur les jeunes, puisque ce sont les principaux espaces d'exploration et de représentation de leur identité. Pour participer à cet espace, les enfants et les jeunes partagent volontairement, mais sont également poussés à publier des informations personnelles par des publications, des stories, des réels ou des commentaires, ainsi qu'à révéler leurs préférences et intérêts personnels par des mentions « J'aime » ou en relayant le contenu d'autres personnes. Et cela ne couvre que ce que les autres utilisateurs et utilisatrices de ces plateformes peuvent voir. Les entreprises en question ont une vision bien plus étendue de la vie privée des jeunes qui utilisent leurs plateformes, grâce aux données sur les contenus que les enfants et les jeunes regardent ou auxquels ils réagissent, à la fréquence et la durée de ces interactions, à leur localisation, à des données sur les heures auxquelles ils se connectent, sur les appareils qu'ils utilisent et sur les achats qu'ils réalisent, pour ne citer que quelques exemples.⁸⁰

Bien que des recherches montrent que les adolescent-e-s plus jeunes ne sont pas toujours pleinement conscients de la surveillance dont ils font l'objet, ces jeunes sont toutefois plutôt conscients de la pression sociale inhérente aux plateformes qui reposent sur le fait que les utilisateurs et utilisatrices se voient et sont vus et des limites que cela crée pour leur sentiment de pouvoir explorer leur identité.⁸¹ Les entretiens qu'Amnesty International a menés avec des jeunes ont permis d'obtenir de vastes informations sur les manières complexes dont nombre de ces jeunes contrôlent qui, dans leur environnement social, peut voir quels types de publications, souvent par l'utilisation de plusieurs comptes sur une même plateforme avec différents niveaux de visibilité et de représentation publique ou d'anonymat.⁸²

4.2 LE DROIT À LA LIBERTÉ DE PENSÉE

Le droit à la liberté de pensée est protégé par l'article 18(1) du Pacte international relatif aux droits civils et politiques (PIDCP) et est un droit absolu, ce qui signifie que personne, en aucun cas, ne peut interférer avec nos pensées et croyances personnelles. Contrairement à la liberté d'expression, c'est-à-dire le fait d'exprimer son opinion, qui peut être limitée par la loi lorsque cela est nécessaire et proportionné à un but légitime,⁸³ par exemple pour protéger la dignité et les droits d'autres personnes, les pensées, croyances et opinions dans notre tête nous appartiennent pleinement et ne peuvent pas être diffusées ou manipulées contre notre gré. La Convention relative aux droits de l'enfant protège le droit à la liberté de pensée des enfants dans son article 14, qui impose aux États parties de « respecter le droit de l'enfant à la liberté de pensée, de conscience et de religion ».

Étroitement lié à ce droit, le droit à la liberté d'opinion, qui est également un droit absolu, est protégé par l'article 19(1) du PIDCP, qui prévoit que « [n]ul ne peut être inquiété pour ses opinions ».

Les droits aux libertés d'opinion et de pensée couvrent trois éléments :

80. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 7 novembre 2023 (déjà cité).

81. Mariya Stoilova, Sonia Livingstone et autres, "Digital by default: children's capacity to understand and manage online data and privacy", 3 novembre 2020, *Media and Communication*, volume 8, numéro 4, <https://eprints.lse.ac.uk/107114/>; Kathryn C. Montgomery, Jeff Chester et autres, "Children's privacy in the big data era: Research opportunities", 1er novembre 2017, *Pediatrics*, Volume 140, numéro 2, <https://doi.org/10.1542/peds.2016-1758O>.

82. De nombreux jeunes adultes aux Philippines ont indiqué en particulier utiliser des comptes publics originaux (« OG ») et des comptes « poubelle » ou « coups de gueule » (des comptes suivis par un nombre plus limité de personnes, utilisés pour donner une version plus honnête et sans filtre de la vie, des sentiments et des opinions des jeunes, souvent sur Instagram, mais pas seulement).

83. PIDCP, article 19 ; Convention relative aux droits de l'enfant, article 13.

- Le droit de ne pas révéler ses pensées et opinions ;
- Le droit de ne pas subir de manipulation de ses pensées et opinions ; et
- Le droit de ne pas être sanctionné en raison de ses pensées et opinions.⁸⁴

Lorsque le PIDCP et d'autres traités relatifs aux droits humains ont été rédigés, personne n'aurait pu prévoir une menace aussi vaste et généralisée pour la liberté de pensée que celle que représente la technologie moderne actuelle. Le modèle économique des entreprises de réseaux sociaux fondé sur la surveillance, qui implique une collecte massive de données personnelles sur tous leurs utilisateurs et utilisatrices (ainsi que la collecte de certaines données non personnelles de personnes qui n'ont même pas créé de compte⁸⁵), représente une menace directe sans précédent pour ces droits. Après avoir collecté les données personnelles des utilisateurs et utilisatrices, les entreprises de réseaux sociaux les utilisent pour analyser les personnes, les répartir en groupes et faire des prévisions sur leurs centres d'intérêt, leurs caractéristiques et, finalement, leur comportement – principalement dans le but de se servir de ces informations pour générer des revenus publicitaires.⁸⁶

Le rapport *I feel exposed: Caught in TikTok's Surveillance Web* explique que la visée lucrative de la collecte et l'agrégation d'autant de données personnelles que possible sur les utilisateurs et utilisatrices des réseaux sociaux (permettant de tirer des conclusions à partir de publications, de mentions « J'aime », de réactions, d'abonnements, de liens, de durée de visionnage, de commentaires, de localisations, de données sur les appareils et bien d'autres, afin de déterminer les intérêts et les orientations sexuelle et politique d'une personne, pour ne citer que quelques caractéristiques) bafoue le premier élément de ce droit : celui de ne pas révéler ses pensées et opinions.⁸⁷

Ce rapport abordera plus en détail la manière dont TikTok utilise la richesse de données personnelles et les conclusions qu'elles permettent de tirer pour envoyer du contenu ciblé aux utilisateurs et utilisatrices afin d'entretenir leur utilisation de la plateforme, enfreignant ainsi potentiellement le droit de ne pas subir de manipulation de ses pensées et opinions. Le Comité des droits de l'enfant aborde cet argument en déclarant qu'il

« [e]ngage les États parties à adopter une réglementation sur la protection des données et des normes de conception qui recensent, définissent et interdisent les pratiques qui altèrent ou compromettent le droit de l'enfant à la liberté de pensée et de croyance dans l'environnement numérique, par exemple par l'analyse ou l'inférence émotionnelle, ou à mettre à jour les réglementations et normes existantes. Les systèmes automatisés peuvent être utilisés pour tirer des conclusions sur l'état intérieur d'un enfant. Les États parties devraient veiller à ce que les systèmes automatisés ou les systèmes de filtrage de l'information ne soient pas utilisés pour agir sur le comportement ou les émotions des enfants ou les influencer ou pour limiter les perspectives ou le développement des enfants⁸⁸».

84. Vermeulen, "Freedom of thought, conscience and religion (article 9)", dans *Theory and Practice of the European Convention on Human Rights, 4e édition.*, eds P. van Dijk, F. van Hoof, A. van Rijn and L. Zwaak (Cambridge: Intersentia Press), 751–772 cité dans McCarth-Jones, *The Autonomous Mind: "The Right to Freedom of Thought in the Twenty-First Century"*, Front. Artif. Intell., 26 septembre 2019, Sec. Technology and Law, <https://doi.org/10.3389/frai.2019.00019> ; Evelyn Aswad, "Losing the freedom to be human", 8 juillet 2020, Columbia Human Rights Law Review, Volume 52, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635701

85. Vice, "TikTok is watching you – even if you don't have an account", 21 janvier 2021, <https://www.vice.com/en/article/jgqbmktiktok-data-collection> ; Wired, "All the ways TikTok tracks you and how to stop it", 23 octobre 2021, <https://www.wired.co.uk/article/tiktok-data-privacy>

86. Amnesty International, *Les géants de la surveillance*, (déjà cité).

87. Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, 7 novembre 2023 (déjà cité).

88. Comité des droits de l'enfant, Observation générale no 25 (déjà cité), § 62.

Le Conseil de l'Europe a également signalé que « les niveaux très subtils, subconscients et personnalisés de la persuasion algorithmique peuvent avoir des effets significatifs sur l'autonomie cognitive des citoyens et leur droit à se forger une opinion et à prendre des décisions indépendantes⁸⁹».

Les conclusions d'Amnesty International, présentées dans les sections suivantes, montrent non seulement que TikTok emploie des pratiques de conception addictives, mais aussi que les systèmes algorithmiques de recommandation de contenu de la plateforme, facteur clé de son ascension mondiale,⁹⁰ exposent les enfants et les jeunes utilisateurs et utilisatrices ayant des problèmes de santé mentale antécédents à de graves préjudices pour la santé mentale.

SYSTÈMES ALGORITHMIQUES DE RECOMMANDATION

Les systèmes permettant la création et la personnalisation de fils d'actualité sur les réseaux sociaux sont souvent évoqués comme un seul « algorithme ». En réalité, TikTok et d'autres plateformes de réseaux sociaux s'appuient sur une multitude d'outils automatisés pour collecter, combiner, trier et générer des données et pour prendre des décisions quant au contenu à afficher à un utilisateur ou une utilisatrice spécifique à un moment précis. La complexité de ces outils d'interaction, et les obstacles considérables à l'accès des chercheurs et chercheuses aux systèmes et données pertinents, sont un obstacle considérable à l'« audit » (c'est-à-dire l'examen) des systèmes de recommandation, comme celui du fil « Pour toi » de TikTok.

4.3 LE DROIT À LA SANTÉ

Les États ont l'obligation de prendre des mesures en vue de progressivement permettre la pleine réalisation du droit de toute personne au meilleur état de santé physique et mentale susceptible d'être atteint. Le droit à la santé est garanti par plusieurs traités relatifs aux droits humains.⁹¹ Le Comité des droits économiques, sociaux et culturels a déclaré que des établissements et des services de santé suffisants doivent être disponibles, accessibles et abordables pour tous les segments de la population.⁹²

Pourtant, la santé mentale reste négligée et les ressources qui y sont allouées restent insuffisantes, ce qui a poussé l'ancien rapporteur spécial sur le droit qu'a toute personne de jouir du meilleur état de santé physique et mentale possible à faire la déclaration suivante dans son rapport de 2020 au Conseil des droits de l'homme de l'ONU :

« Il ne peut y avoir de santé sans santé mentale. Depuis plusieurs dizaines d'années, on sait qu'il existe des liens forts entre l'esprit, le corps et l'environnement. Or, à l'aube de la troisième décennie de ce millénaire, la parité entre santé mentale et santé physique n'a été atteinte nulle part dans le monde, ce qui demeure un obstacle majeur au développement humain. Le message important à tirer de cet échec collectif est que si l'on n'accorde pas l'attention voulue aux droits de l'homme, aucun investissement dans la santé mentale ne portera ses fruits.⁹³»

89. Conseil de l'Europe, Déclaration du Comité des Ministres sur les capacités de manipulation des processus algorithmiques, 13 février 2019, https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4c

90. Alex Hern, "How TikTok's algorithm made it a success: 'It pushes the boundaries'", 24 octobre 2022, <https://www.theguardian.com/technology/2022/oct/23/tiktok-rise-algorithm-popularity>.

91. Article 25 de la DUDH, article 12 du PIDESC, article 24 de la Convention relative aux droits de l'enfant. Voir également des traités régionaux de protection des droits humains, notamment la Charte africaine des droits de l'homme et des peuples et la Charte sociale européenne, l'article 16 de la Charte africaine des droits de l'homme et des peuples, et la Charte sociale européenne (révisée).

92. Comité des droits économiques, sociaux et culturels, Observation générale n° 14, 11 août 2000, Doc. ONU E/C.12/2000/4.

93. Rapport du Rapporteur spécial sur le droit qu'a toute personne de jouir du meilleur état de santé physique et mentale possible (rapporteur spécial sur le droit à la santé), 15 juillet 2020, Doc. ONU A/HRC/44/48, § 1.

Il a également prévenu en juin 2020 que la pandémie de COVID-19 avait aggravé la « négligence historique de la prise en charge digne de la santé mentale » et que « l'association des fermetures des établissements scolaires et des conditions de confinement a eu des conséquences particulières pour le stress, l'anxiété et les problèmes de santé mentale des enfants. C'est particulièrement inquiétant étant donné le manque général de reconnaissance et de sensibilisation à la nature et l'ampleur des problèmes de santé mentale chez les adolescent-e-s dans de nombreux pays⁹⁴»

Les technologies numériques ont indéniablement joué un rôle positif dans la réduction des obstacles aux informations et services en matière de santé, par exemple en permettant aux psychologues de s'entretenir avec des patient-e-s à distance pendant les confinements, en permettant aux jeunes de rester en contact avec leurs réseaux de soutien et en donnant aux groupes marginalisés comme les adolescent-e-s LGBTI la possibilité de découvrir leur identité et de s'exprimer.⁹⁵ Mais il est également de plus en plus reconnu qu'en l'absence de réglementation étatique efficace, les plateformes de réseaux sociaux, en particulier, ont exposé les enfants et les jeunes à des risques systémiques comme ceux décrits dans le présent rapport.⁹⁶ Comme l'a noté le Comité des droits de l'enfant dans son Observation générale n° 25 :

« L'environnement numérique peut contenir des informations véhiculant des stéréotypes de genre, des informations discriminatoires, racistes, violentes, pornographiques ou abusives, ainsi que des récits mensongers, de la désinformation et des informations encourageant les enfants à se livrer à des activités illégales ou nocives. Ces informations peuvent provenir de sources multiples, notamment d'autres utilisateurs, de créateurs de contenus commerciaux, de délinquants sexuels ou de groupes armés qualifiés de terroristes ou d'extrémistes violents. Les États parties devraient protéger les enfants contre les contenus nocifs et douteux et veiller à ce que les entreprises concernées et les autres fournisseurs de contenus numériques élaborent et appliquent des directives permettant aux enfants d'accéder en toute sécurité à des contenus diversifiés, en reconnaissant le droit de l'enfant à l'information et son droit à la liberté d'expression, tout en protégeant les enfants contre les matériels nocifs, conformément à leurs droits et au développement de leurs capacités.⁹⁷» [Mise en italique par nos soins].

Se concentrant davantage sur les risques pour la santé des enfants et des jeunes, l'administrateur de la santé publique des États-Unis a déclaré en mai 2023 :

« Presque tous les adolescent-e-s aux États-Unis utilisent les réseaux sociaux, et pourtant, nous n'avons pas assez d'éléments pour conclure que c'est suffisamment sûr pour eux. Nos enfants sont, sans le savoir, devenus les participant-e-s d'une expérience de plusieurs décennies... Nous devons reconnaître le corpus croissant de recherches sur les potentiels préjudices, renforcer notre compréhension collective des risques liés à l'utilisation des réseaux sociaux et prendre de toute urgence des mesures pour limiter les préjudices et protéger la santé mentale et le bien-être des enfants et des adolescent-e-s pendant ces étapes de développement cruciales.⁹⁸»

94. HCDH, "COVID-19 has exacerbated the historical neglect of dignified mental health care, especially for those in institutions: UN expert", 23 juin 2020, <https://www.ohchr.org/en/news/2020/06/covid-19-has-exacerbated-historical-neglect-dignified-mental-health-care-especially>

95. Benjamin Hanckel et Shiva Chandra, "How young LGBTQIA+ people used social media to thrive during COVID lockdowns", *The Conversation*, 15 mars 2021, <https://theconversation.com/how-young-lgbtqia-people-used-social-media-to-thrive-during-covid-lockdowns-156130>

96. Voir par exemple le Rapport de la Rapporteuse spéciale sur le droit qu'a toute personne de jouir du meilleur état de santé physique et mentale possible : Innovation et technologies numériques et droit à la santé, avril 2023, <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G23/074/90/PDF/G2307490.pdf?OpenElement> ; Comité des droits de l'enfant, Observation générale n° 25, 2 mars 2021, Doc. ONU CRC/C/GC/25

97. Comité des droits de l'enfant, Observation générale n° 25, 2 mars 2021, Doc. ONU CRC/C/GC/25, § 54.

98. Administrateur de la santé publique des États-Unis, "Social Media and Youth Mental Health", mai 2023, <https://www.hhs.gov/surgeongeneral/priorities/youth-mental-health/social-media/index.html>

Cette réaction urgente requiert tant un travail réglementaire des gouvernements, conformément à leurs obligations au regard du droit international relatif aux droits humains de réaliser le droit à la santé et de prendre des mesures visant à protéger les droits humains dans le cadre des activités des entreprises, que des mesures immédiates des entreprises de réseaux sociaux en vue de respecter les droits humains et de se conformer à leurs responsabilités concernant le devoir de diligence en matière de droits humains (évoqué plus en détail dans la section suivante).

4.4 L'INTÉRÊT SUPÉRIEUR DE L'ENFANT ET LE DROIT D'ÊTRE ENTENDU·E

L'un des principes essentiels de la Convention relative aux droits de l'enfant est le principe de l'« intérêt supérieur de l'enfant ». L'article 3(1) prévoit : « Dans toutes les décisions qui concernent les enfants, qu'elles soient le fait des institutions publiques ou privées de protection sociale, des tribunaux, des autorités administratives ou des organes législatifs, l'intérêt supérieur de l'enfant doit être une considération primordiale.⁹⁹»

L'article 12 de la Convention relative aux droits de l'enfant prévoit que l'enfant a « le droit d'exprimer librement son opinion sur toute question l'intéressant, les opinions de l'enfant étant dûment prises en considération eu égard à son âge et à son degré de maturité.¹⁰⁰» Les capacités évolutives de l'enfant doivent être prises en considération en parallèle du principe d'intérêt supérieur et du droit de l'enfant de faire entendre ses opinions.

Appliquant ces principes à l'environnement numérique, le Comité des droits de l'enfant recommande que les États veillent à ce que, « dans toutes les décisions concernant la fourniture, la réglementation, la conception, la gestion et l'utilisation de l'environnement numérique, l'intérêt supérieur de chaque enfant soit une considération primordiale¹⁰¹ » et à ce que « les fournisseurs de services numériques offrent des services adaptés au développement des capacités des enfants.¹⁰² »

4.5 L'OBLIGATION DES ENTREPRISES DE RESPECTER LES DROITS HUMAINS

Les entreprises ont l'obligation de respecter les droits humains, où qu'elles opèrent dans le monde et dans tous leurs domaines d'activité. Il s'agit d'une norme de conduite générale largement reconnue et inscrite dans les normes internationales relatives à la responsabilité des entreprises en matière de droits humains, notamment dans les Principes directeurs relatifs aux entreprises et aux droits de l'homme des Nations unies (Principes directeurs de l'ONU) et dans les Principes directeurs à l'intention des entreprises multinationales de l'Organisation de coopération et de développement économiques¹⁰³ (Principes directeurs de l'OCDE). La responsabilité qu'ont les entreprises de respecter

99. Convention relative aux droits de l'enfant, article 3(1)

100. Convention relative aux droits de l'enfant, article 12.

101. Comité des droits de l'enfant, Observation générale no 25 (déjà cité), § 12.

102. Comité des droits de l'enfant, Observation générale no 25 (déjà cité), § 20.

103. Principes directeurs relatifs aux entreprises et aux droits de l'homme : Mise en œuvre du cadre de référence « protéger, respecter et réparer » des Nations unies, 2011, approuvés par la résolution 17/4 du Conseil des droits de l'homme des Nations unies : Les droits de l'homme et les sociétés transnationales et autres entreprises, adopté le 16 juin 2011, doc. ONU A/HRC/RES/17/4 ; OCDE, Principes directeurs pour les entreprises multinationales, 2011, <https://www.oecd.org/fr/gouvernementdentreprise/mne/>. Conformément aux Principes directeurs des Nations unies, la responsabilité qu'ont les entreprises de respecter les droits humains est indépendante des obligations propres aux États en la matière et prévaut sur le respect des lois et règlements nationaux qui protègent les droits fondamentaux. Voir les Principes directeurs des Nations unies, principe 11 et commentaire associé.

les droits humains est indépendante des obligations propres aux États en la matière et prévaut sur le respect des lois et règlements nationaux qui protègent les droits fondamentaux.¹⁰⁴

Cette responsabilité impose aux entreprises d'éviter d'être à l'origine d'atteintes aux droits humains ou d'y contribuer par leurs propres activités, et de remédier aux effets néfastes auxquels elles ont contribué, notamment en remédiant à toute atteinte. Les Principes directeurs des Nations unies prévoient que, pour y parvenir, les entreprises doivent mettre en œuvre de façon continue et volontariste une procédure de diligence raisonnable en matière de droits humains leur permettant d'identifier leurs incidences sur les droits humains, de les prévenir et d'en atténuer les effets, ainsi que de rendre compte de la manière dont elles y remédient. Si, au cours de la procédure de diligence requise, l'entreprise se rend compte qu'elle commet des atteintes aux droits humains ou y contribue, elle doit alors faire cesser ou prévenir ces atteintes.¹⁰⁵

Les États ont l'obligation de respecter et de protéger les droits humains dans le cadre des activités des entreprises. Ils ont l'obligation de « protéger lorsque des tiers, y compris des entreprises, portent atteinte aux droits de l'homme sur leur territoire et/ou sous leur juridiction. Cela exige l'adoption de mesures appropriées pour empêcher ces atteintes, et lorsqu'elles se produisent, enquêter à leur sujet, en punir les auteurs, et les réparer par le biais de politiques, de lois, de règles et de procédures judiciaires.¹⁰⁶»

En ce qui concerne spécifiquement les enfants, le Comité des droits de l'enfant a déclaré que les États doivent « adopter des mesures propres à garantir le respect des droits de l'enfant par les entreprises¹⁰⁷ » et a présenté des mesures qui doivent être prises par les États pour empêcher les entreprises de causer des atteintes aux droits humains ou d'y contribuer et pour « enquêter et se prononcer sur les violations des droits de l'enfant qui ont été causées par une entreprise ou auxquelles une entreprise a contribué et [...] offrir réparation ». Le Comité a insisté sur le fait qu'un État est « responsable des violations des droits de l'enfant causées par une entreprise ou auxquelles une entreprise a contribué lorsqu'il n'a pas pris les mesures nécessaires, appropriées et raisonnables pour prévenir et réparer ces violations ou lorsqu'il a de toute autre manière collaboré à leur commission ou les a tolérées¹⁰⁸».

La transparence et l'obligation de rendre compte publiquement de la manière dont une entreprise a identifié, prévenu ou atténué des conséquences néfastes de ses activités pour les droits humains sont essentielles à une diligence requise efficace. Les Principes directeurs des Nations unies prévoient que les entreprises doivent « connaître les droits de l'homme et montrer qu'elles les respectent¹⁰⁹ » et « communiquer, en assurant un certain degré de transparence et de responsabilité aux individus ou aux groupes susceptibles d'être touchés et aux autres acteurs pertinents.¹¹⁰ » Le Comité des droits de l'enfant a également indiqué que les États doivent encourager, le cas échéant, les entreprises à publier des informations sur les mesures qu'elles ont prises pour lutter contre les répercussions néfastes de leurs activités pour les droits humains des enfants.¹¹¹

Cette obligation de transparence contraste clairement avec la réalité de l'influence toujours plus forte des algorithmes développés par des entités privées, permettant de déterminer quelles publications sont mises en avant et ensuite vues par de très nombreuses personnes sur les réseaux sociaux ou quels types de publications enfreignent le règlement de la plateforme. Les plateformes de réseaux

104. Haut-Commissariat des Nations unies aux droits de l'homme, Principes directeurs relatifs aux entreprises et aux droits de l'homme : Mise en œuvre du cadre de référence « protéger, respecter et réparer » des Nations Unies, doc. ONU HR/PUB/11/04 (2011), https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_fr.pdf, principes 11 et commentaire associé.

105. Principes directeurs des Nations unies (déjà cité), principe 19 et commentaire associé.

106. Principes directeurs des Nations unies (déjà cité), principe 1.

107. Comité des droits de l'enfant, Observation générale no 16, 17 avril 2013, Doc. ONU CRC/C/GC/16, § 62.

108. Comité des droits de l'enfant, Observation générale no 16 (déjà cité), § 28.

109. Principes directeurs des Nations unies (déjà cité), commentaire associé au principe 21.

110. Principes directeurs des Nations unies (déjà cité), commentaire associé au principe 15.

111. Comité des droits de l'enfant, Observation générale no 16 (déjà cité), § 65.

sociaux cherchent depuis longtemps à protéger ce qu’elles perçoivent comme le secret commercial de leurs modèles économiques¹¹² et à limiter la capacité des chercheurs et chercheuses indépendants à suivre les résultats de leurs algorithmes « boîte noire¹¹³». À la lumière de la reconnaissance croissante des préjudices dans la vie réelle associés aux plateformes qui favorisent et relaient des contenus extrêmes,¹¹⁴ plusieurs organes de l’ONU et d’autres organismes internationaux ont cherché à élaborer des réponses à ce vide juridique considérable.

En 2021, le Haut-Commissariat des Nations unies aux droits de l’homme a présenté des recommandations aux États et aux entreprises pour traiter les risques en matière de droits humains liés à l’utilisation de l’intelligence artificielle, qui joue un rôle toujours plus important pour perfectionner les systèmes de recommandation de contenu, notamment :

1. « D’exercer systématiquement une diligence raisonnable en matière de droits de l’homme tout au long du cycle de vie des systèmes d’IA qu’elles conçoivent, développent, déploient, vendent, acquièrent ou exploitent. De procéder à des évaluations régulières et complètes des incidences sur les droits de l’homme en tant qu’élément clé de l’exercice de leur diligence raisonnable en matière de droits de l’homme ;
2. D’accroître considérablement la transparence de leur utilisation de l’IA, notamment en informant comme il se doit le public et les personnes concernées et en permettant un audit indépendant et externe des systèmes automatisés. Plus les impacts potentiels ou réels sur les droits de l’homme liés à l’utilisation de l’IA sont probables et graves, plus la transparence est nécessaire.
3. De veiller à ce que toutes les parties prenantes participent aux décisions concernant le développement, le déploiement et l’utilisation de l’IA, en particulier les personnes et groupes concernés ;
4. D’améliorer l’explicabilité des décisions prises à l’aide de l’IA, notamment en finançant et en menant des recherches à cette fin.¹¹⁵ »

À la suite d’appels du Conseil de l’Europe¹¹⁶ et d’autres organes en faveur d’une réglementation étatique de l’utilisation de systèmes algorithmiques, des initiatives ont été lancées telles que la législation sur les services numériques (DSA) de l’UE, et le projet de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l’intelligence artificielle (législation sur l’intelligence artificielle) et modifiant certains actes législatifs de l’Union,¹¹⁷ et visent à renforcer la transparence, à mettre en place des mécanismes de surveillance et à empêcher les utilisations abusives de l’intelligence artificielle.

Le chapitre 7 du présent rapport explique que TikTok ne respecte pas sa responsabilité de respect des droits humains, car l’entreprise ne fait pas le nécessaire pour atténuer les risques pour le droit au respect de la vie privée, le droit à la liberté de pensée et le droit à la santé, dans le cadre de ses fils d’actualité hyper-personnalisés et d’autres éléments de maximisation de l’engagement qui entretiennent le modèle de TikTok fondé sur la surveillance.

112. Alex Hern, “How TikTok’s algorithm made it a success: ‘It pushes the boundaries’” (déjà cité).

113. Mozilla Foundation, Amnesty International and others, “Response to the European Commission’s call for evidence for a Delegated Regulation on data access provided for in the Digital Services Act” (déjà cité) ; Frederick Mostert, “Social media platforms must abandon algorithmic secrecy”, 17 juin 2021, <https://www.ft.com/content/39d69f80-5266-4e22-965f-efbc19d2e776>

114. Amnesty International, *Myanmar. L’atrocité des réseaux sociaux. Meta face au droit à réparations des Rohingyas* – Synthèse (déjà cité) ; *Les géants de la surveillance* (déjà cité).

115. HCDH, Le droit à la vie privée à l’ère du numérique, 15 septembre 2021, doc. ONU A/HRC/48/31.

116. Conseil de l’Europe, Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l’homme, 8 avril 2020.

117. Le projet de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l’intelligence artificielle (législation sur l’intelligence artificielle) et modifiant certains actes législatifs de l’Union vise à réglementer les systèmes d’IA plus généralement, alors que la législation sur les services numériques (DSA) concerne les risques liés aux systèmes algorithmiques des grandes plateformes de réseaux sociaux. Amnesty International, *UE. La législation sur l’IA serait menacée si le Parlement européen légitimait des technologies abusives*, 13 juin 2023, <https://www.amnesty.org/fr/latest/news/2023/06/eu-ai-act-at-risk-as-european-parliament-may-legitimize-abusive-technologies/>

5. UNE CONCEPTION ADDICTIVE PAR NATURE

Contrairement aux autres moyens de communication de masse, les entreprises de réseaux sociaux personnalisent l'expérience de leurs utilisateurs et utilisatrices afin de faire en sorte que ceux-ci passent le plus de temps possible sur la plateforme et d'encourager leur engagement, par des contenus et, l'élément le plus essentiel à leur montée en puissance, des publicités. Le rapport de 2019 *Les géants de la surveillance* d'Amnesty International démontrait que Facebook (aujourd'hui Meta) et Google (dont le propriétaire est Alphabet) ont optimisé l'utilisation des analyses de mégadonnées et des pratiques de conception addictives si efficacement que ces entreprises dominent le marché mondial de la publicité.¹¹⁸ En 2021, les deux entreprises représentaient à elles seules 43 % de toutes les dépenses publicitaires en ligne et hors ligne dans le monde.¹¹⁹ TikTok a intégré ce marché déterminée à attirer rapidement des centaines de millions de jeunes utilisateurs et utilisatrices et à revendiquer les parts de temps d'utilisation et d'attention des utilisateurs et utilisatrices d'autres médias.¹²⁰ Les ressources de marketing de l'entreprise soulignent qu'elle considère non seulement qu'elle est en concurrence avec les autres réseaux sociaux, mais aussi qu'elle vise à s'approprier l'espace alloué également à d'autres activités quotidiennes, clamant que « 41 % des utilisateurs et utilisatrices de TikTok de la génération Z dans le monde indiquent passer moins de temps à écouter des podcasts depuis qu'ils ont commencé à utiliser TikTok.¹²¹»

Les conséquences d'une utilisation prolongée de TikTok signalées aux équipes de recherche par les enfants et les jeunes avec qui Amnesty International s'est entretenue ne se limitent pas au fait de passer moins de temps sur d'autres médias. Plusieurs jeunes avec qui nous sommes entretenus avaient le sentiment qu'en raison de leur utilisation de TikTok, ils n'avaient pas remis des devoirs pour l'école ou l'université, qu'ils passaient moins de temps avec leurs ami-e-s et qu'ils faisaient défiler leur fil d'actualité jusqu'à tard le soir au lieu de dormir, ce qui a de graves conséquences pour leur santé.¹²²

Par exemple, « Mary », qui est inscrite dans une école pour filles dans l'ouest du Kenya, a déclaré :

118. Amnesty International, *Les géants de la surveillance* (déjà cité).

119. *Ebiquity*, "Google, Meta and Amazon are on track to absorb more than 50% of all ad money in 2022", 4 février 2022, <https://ebiquity.com/news-insights/press/google-meta-and-amazon-are-on-track-to-absorb-more-than-50-of-all-ad-money-in-2022/>

120. Yahoo Finance, "Forget Facebook, Snapchat, Twitter: TikTok Is The Breakout COVID-19 Social Media Platform", 24 avril 2020, <https://finance.yahoo.com/news/forget-facebook-snapchat-twitter-tiktok-210717993.html> ; *Business Insider*, "Nearly half of Gen Z is using TikTok and Instagram for search instead of Google, according to Google's own data", 13 juillet 2022, <https://www.businessinsider.com/nearly-half-genz-use-tiktok-instagram-over-google-search-2022-7> ; TikTok, "TikTok Marketing Science Global Time Well Spent Study (Global Results) conducted by Kantar March 2021", <https://tiktok.com/business/en-US/insights> (consulté le 1er June 2023).

121. TikTok, "TikTok Marketing Science Global Time Well Spent Study (Global Results) conducted by Kantar March 2021" (déjà cité).

122. Évoqué lors de plusieurs discussions de groupe à Kisumu, au Kenya, les 13 et 14 mars 2023, ainsi que dans des discussions de groupe dans des écoles à Mombasa, les 15, 16 et 17 mars 2023.

« Quand je me réveille, la première chose que je fais c'est regarder mon téléphone, je ne pense qu'à mon téléphone.¹²³»

« Amy », une autre participante au groupe de discussion, a déclaré :

« Si j'utilise TikTok, j'ai généralement l'intention de regarder cinq minutes, j'essaie de regarder deux (vidéos), mais je finis par continuer de regarder toute la nuit. Je ne veux y passer que cinq minutes, mais les vidéos sont si intéressantes que je finis par regarder [pendant] au moins six heures. »

Interrogés sur l'application de réseaux sociaux la plus difficile à arrêter d'utiliser, les 29 participant-e-s du groupe ont presque unanimement répondu TikTok.¹²⁴

« Joyce », une jeune femme des Philippines, a déclaré :

« Je l'ai supprimée [l'application TikTok] pendant un moment, car j'étais trop dépendante... je passais tellement d'heures sur TikTok, à faire défiler des vidéos, parce qu'on ne peut pas s'empêcher de se demander ce qui va venir après quand on fait défiler.¹²⁵»

Ces témoignages, ainsi que de nombreux autres, d'enfants et de jeunes ont été corroborés par des psychologues spécialistes des adolescent-e-s consultés dans le cadre des recherches, qui ont déclaré avoir constaté une tendance sur TikTok aggravant les problèmes préexistants par des schémas addictifs dans l'utilisation que font les enfants et les jeunes adultes des réseaux sociaux. « Un grand nombre de ces personnes me disent qu'elles ont besoin de prendre une pause dans l'utilisation des réseaux sociaux, qu'elles sont dépendantes aux réseaux sociaux. Tout ce qu'elles veulent faire c'est rester assises à faire défiler le fil d'actualité toute la journée », a déclaré Zeyna Awan, psychologue à Nairobi.¹²⁶ Kevin Gachee, un autre psychologue de Nairobi, a observé un changement générationnel, constatant que les jeunes de la génération Z avec lesquels il travaille semblent souvent être confrontés à une utilisation excessive et dangereuse de TikTok, alors que les adultes des générations précédentes sont plus susceptibles de faire part de problèmes liés à une comparaison sociale néfaste qui serait encouragée par leur utilisation d'Instagram.¹²⁷

Les éléments des études disponibles sont axés sur les données sur l'utilisation des réseaux sociaux par les enfants aux États-Unis et en Europe. Une étude de 2021 a conclu que TikTok est la plateforme sur laquelle les enfants passent le plus de temps aux États-Unis, au Royaume-Uni et en Espagne, les enfants des États-Unis passant en moyenne 99 minutes par jour sur TikTok.¹²⁸ Un récent sondage mené auprès d'adolescent-e-s aux États-Unis a conclu que 45 % des utilisatrices de TikTok « indiquent avoir le sentiment d'avoir développé une "dépendance" à la plateforme ou l'utiliser plus que ce qu'elles n'en avaient l'intention au moins chaque semaine.¹²⁹»

Nos recherches qualitatives au Kenya et aux Philippines semblent pointer vers les mêmes tendances. De toutes les plateformes de réseaux sociaux, TikTok est plus spécifiquement décrite par les jeunes comme étant « addictive » et comme faisant « perdre du temps ». Au moins 25 participant-e-s de 10 des 11 groupes de discussion au Kenya ont fait des déclarations dans ce sens et de nombreux autres ont indiqué être d'accord en hochant la tête. De jeunes utilisateurs et utilisatrices de TikTok ont

123. Discussion de groupe, Kisumu, Kenya, 14 mars 2023.

124. Discussion de groupe, Kisumu, Kenya, 14 mars 2023.

125. Entretien en ligne avec une participante de Manille, 22 mai 2023.

126. Entretien avec une experte, en ligne, le 9 mars 2023.

127. Entretien avec un expert, en ligne, le 9 mars 2023.

128. Qustodio, "Annual Data Report 2021: Living and Learning in a Digital World", chapitre 2, Social Media, <https://www.qustodio.com/en/social-media-qustodio-annual-data-report-2021> (consultée le 20 juillet 2023).

129. Jacqueline Nesi, Supreet Mann & Michael B. Robb, "Teens and mental health: How girls really feel about social media", 2023, https://www.common sense media.org/sites/default/files/research/report/how-girls-really-feel-about-social-media-researchreport_web_final_2.pdf

décrit des schémas d'utilisation extrême pendant la nuit¹³⁰ et une minorité tout de même importante d'enfants allant à l'école au Kenya ont indiqué que, du fait qu'ils partagent leurs téléphones avec leurs parents ou leurs grands frères et sœurs, ils n'y ont accès que tard le soir ou la nuit et utilisent donc souvent les réseaux sociaux sans supervision parentale jusqu'à tard la nuit ou aux premières heures du jour, ce qui les prive de sommeil avant l'école.¹³¹ Ces conclusions pourraient indiquer des risques supplémentaires pour la santé des enfants et adolescent-e-s économiquement désavantagés.

Les nouvelles recherches sur l'utilisation des réseaux sociaux et la santé mentale des adolescent-e-s laissent entrevoir une tendance selon laquelle les enfants souffrant de problèmes de santé mentale seraient plus susceptibles de faire une utilisation excessive des réseaux sociaux. Un rapport de 2021 du Service national de santé (NHS) du Royaume-Uni a conclu que « les enfants de 11 à 16 ans ayant des problèmes de santé mentale étaient particulièrement susceptibles de passer plus de temps sur les réseaux sociaux qu'ils n'en avaient l'intention. Presque deux tiers des enfants de cette catégorie (63,8 %) ont fait des déclarations dans ce sens, contre moins de la moitié des enfants peu susceptibles de souffrir de troubles mentaux¹³² (45,5 %). » Une analyse systématique des études sur l'utilisation des réseaux sociaux et les conséquences associées pour les adolescent-e-s souffrant de troubles mentaux préexistants a conclu que les adolescent-e-s souffrant de dépression ou « ayant signalé des symptômes d'anxiété, de dépression ou des comportements auto-destructeurs » sont plus susceptibles d'avoir des difficultés à réguler leur utilisation d'Internet, y compris des réseaux sociaux.¹³³

5.1 LES STRATÉGIES D'ENCOURAGEMENT DE L'ENGAGEMENT DE TIKTOK

La capacité des personnes à réguler leur utilisation des réseaux sociaux et certains facteurs contextuels peuvent varier, mais, comme d'autres plateformes de réseaux sociaux, TikTok a fait des choix en matière de conception destinés à faire en sorte que les utilisateurs et utilisatrices passent autant de temps que possible sur la plateforme.¹³⁴ En plus d'intégrer des éléments de conception que d'autres plateformes de réseaux sociaux ont employés avec succès pour maintenir l'attention des utilisateurs et utilisatrices sur l'écran,¹³⁵ par exemple la possibilité d'attribuer une mention « J'aime » à une publication, les notifications fréquentes et le fil d'actualité sans fin, TikTok se démarque par son fil « Pour toi », une page hautement personnalisée et longue de contenus recommandés par des algorithmes, sélectionnés pour refléter ce que le système a identifié comme les intérêts de l'utilisateur ou l'utilisatrice en quelques minutes après l'utilisation du premier fil.¹³⁶

130. Discussions de groupe menées à Kisumu, au Kenya, les 13 et 14 mars 2023, à Mombasa, au Kenya, le 16 et 17 mars, et à Machakos, au Kenya, le 22 mars 2023.

131. Au moins neuf jeunes participant-e-s du Kenya ont indiqué, soit dans le cadre de discussions sur quand et comment ils utilisaient les réseaux sociaux, soit lorsqu'ils ont été poussés à réfléchir à la plateforme sur laquelle ils passaient le plus de temps, manquer de sommeil à cause de leur utilisation de TikTok, notamment pendant l'année scolaire. Le nombre d'enfants partageant leur téléphone était variable en fonction des groupes, allant de moins de 10 % à près de 50 % des élèves.

132. NHS, "Mental Health of Children and Young People in England 2021 - Wave 2 follow up to 2017 survey", 30 septembre 2021, <https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2021-follow-up-to-the-2017-survey>.

133. Katarzyna Kostyrka-Allichorne, Mariya Stoilova et autres, "Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and non-suicidal self-injury conditions – a systematic review", février 2023, *Child and Adolescent Mental Health*, Volume 28, numéro 1, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10108198/>.

134. Sophia Petrillo, "What makes TikTok so addictive?: An analysis of the mechanisms underlying the world's latest social media craze", 13 décembre 2021, *Brown Undergraduate Journal of Public Health*, Numéro 2021-2022, <https://sites.brown.edu/publichealthjournal/2021/12/13/tiktok/>.

135. ABC News, "Book excerpt: Jaron Lanier's 'Ten Arguments for Deleting Your Social Media Accounts Right Now'", 19 juin 2018, <https://abcnews.go.com/Technology/book-excerpt-jaron-laniers-ten-arguments-deleting-social/story?id=56009512> ; BBC, "Facebook founding president sounds alarm", 9 novembre 2017, <https://www.bbc.co.uk/news/technology-41936791>

136. TikTok, "How TikTok recommends videos #ForYou", 18 juin 2020, <https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you>.

Là où les plateformes de réseaux sociaux s'appuyaient sur l'indication active de leurs liens sociaux par les utilisateurs et utilisatrices pour prédire leurs intérêts, les recommandations de contenu de TikTok s'appuient très largement sur l'analyse des vidéos que les utilisateurs et utilisatrices regardent ou auxquelles ils réagissent parmi une sélection initiale de vidéos populaires. Ces informations sont alors utilisées pour préciser les intérêts de la personne en recommandant des vidéos, que le système associe au contenu précédemment visualisé.¹³⁷ Cela a pour résultat un fil d'actualité beaucoup plus personnalisé de courtes vidéos, identifiées comme susceptibles de correspondre le mieux à la personne.¹³⁸

En combinant ces éléments, TikTok exploite ce que les psychologues appellent le « schéma de récompense par intermittence de gains ou de pertes sur une machine à sous... profitant d'un conditionnement classique et de processus d'apprentissage fondés sur les récompenses pour permettre la formation de boucles d'habitudes et encourager des utilisations addictives.¹³⁹ En somme, TikTok a exploité au maximum les qualités addictives de différents choix de conception, dont le président fondateur de Facebook avait prévenu en 2017 qu'ils pouvaient avoir des conséquences néfastes involontaires, particulièrement pour la santé et le développement des enfants.¹⁴⁰ Les jeunes avec qui nous sommes entretenus ayant fait part d'une utilisation compulsive de TikTok ont souvent décrit le comportement qui en résulte : un « défilement inattentif¹⁴¹ » par lequel les utilisateurs et utilisatrices « n'ont pas le sentiment d'avoir un lien ou de ressentir quoi que ce soit vis-à-vis de la publication, parce qu'ils n'ont pas le temps de l'analyser lorsqu'ils font simplement défiler.¹⁴²»

Les adolescent-e-s ayant participé à l'étude ont également parlé du rôle perçu de la fonction « J'aime », commune à TikTok et d'autres plateformes de réseaux sociaux, qui exacerbe les incertitudes et garde leur attention fixée sur l'écran.¹⁴³ Ils ont expliqué avoir le sentiment de devoir consulter leur compte pour savoir si leur publication a autant de succès que d'autres. Des élèves d'une zone rurale du Kenya ont même indiqué utiliser des outils d'édition de photos ou de vidéos pour créer de fausses publications sur des expériences que leur famille n'aurait pas les moyens de s'offrir, et ce, afin de gagner des abonnés.¹⁴⁴

Des participant-e-s à l'étude plus âgés, notamment des étudiant-e-s universitaires, ont parlé du rôle perçu de leur profil sur les réseaux sociaux pour définir leur statut social et leur estime de soi : « Lorsque vous arrivez à l'université, vous êtes votre compte sur les réseaux sociaux », a déclaré Nikki, 24 ans, de Manille. Victor, un jeune adulte d'un groupe de discussion de Mombasa a exprimé l'idée que les réseaux sociaux imposent une représentation de soi minutieusement définie, déclarant : « La personne que vous voyez sur les réseaux sociaux n'est pas la personne qui se trouve ici aujourd'hui.¹⁴⁵»

Plusieurs études psychologiques et, comme l'ont révélé les « Facebook Papers », les recherches de l'entreprise elle-même,¹⁴⁶ ont constaté l'effet néfaste de la comparaison sociale fondée sur les

137. Alex Hern, "How TikTok's algorithm made it a success: 'It pushes the boundaries'" (déjà cité).

138. Catherine Wang, "Why TikTok made its user so obsessive? The AI Algorithm that got you hooked", 7 juin 2020, <https://towardsdatascience.com/why-tiktok-made-its-user-so-obsessive-the-ai-algorithm-that-got-you-hooked-7895bb1ab423>.

139. Sophia Petrillo, "What makes TikTok so addictive?: An analysis of the mechanisms underlying the world's latest social media craze", décembre 2021, *Brown Undergraduate Journal of Public Health*, Numéro 2021-2022 (déjà cité).

140. Le président fondateur de Facebook Sean Parker a lui-même déclaré en 2017 : « [Facebook] entrave sans doute la productivité de manières surprenantes. Dieu seul sait ce qui se passe dans la tête des enfants. » BBC, "Facebook founding president sounds alarm" (déjà cité).

141. Entretien en ligne avec Manuel (pseudonyme), 23 ans, étudiant basé à Luzon, 4 mai 2023.

142. Entretien en ligne avec Nikki (pseudonyme), 24 ans, Manille, 8 mai 2023.

143. Discussions de groupe avec des élèves d'établissements d'enseignement secondaire dans le comté de Kisumu, le 14 mars 2023, à Mombasa, le 17 mars, et dans le comté de Machakos, le 22 mars 2023.

144. Discussions de groupe avec des adolescent-e-s dans un établissement d'enseignement secondaire du comté de Machakos, le 22 mars 2023.

145. Discussion de groupe avec des étudiant-e-s universitaires et des jeunes militant-e-s à Mombasa, au Kenya, le 17 mars 2023.

146. *Wall Street Journal*, "Facebook knows Instagram is toxic for teen girls company documents show" (déjà cité).

réseaux sociaux pour la santé mentale des adolescent-e-s.¹⁴⁷ Instagram a mené une expérience de deux ans dans le cadre de laquelle le nombre de mentions « J'aime » attribuées à une publication était masqué pour tous les utilisateurs et utilisatrices, sauf la personne à l'origine de la publication. L'idée a finalement été abandonnée, et il a été décidé de laisser plutôt aux utilisateurs et utilisatrices le choix de décider si les mentions « J'aime » doivent être visibles, expliquant que l'expérience avait reçu un accueil mitigé.¹⁴⁸ L'intérêt de réduire l'accent mis sur la popularité et la comparaison sociale pour stimuler l'estime de soi des jeunes et des interactions plus saines est cependant clair pour des expert-e-s tels que le psychologue Kevin Gachee, basé à Nairobi : « Ces applications sont conçues pour entretenir l'engagement. J'ai félicité Instagram lorsque la plateforme a [temporairement] cessé de montrer le nombre de mentions "J'aime", car cela a considérablement contribué à pousser les gens à se dire : "Je ne ressens pas la pression de publier sans arrêt parce que mes ami-e-s ont obtenu 100 mentions 'J'aime' et j'aimerais en obtenir autant".¹⁴⁹»

Plus inquiétant encore en ce qui concerne les conséquences pour la santé mentale : le fait que les réseaux sociaux poussent les enfants qui s'auto-mutilent à des comportements d'affirmation. Le psychologue Marc Reyes, basé à Manille, a fait part de ses inquiétudes à Amnesty International : « La plupart des personnes qui ont des problèmes de santé mentale se sentent invisibles, mais sur leurs réseaux sociaux, ces personnes sont vues. Les mentions "J'aime" deviennent alors très néfastes : de nombreuses personnes aiment ma publication [sur l'auto-mutilation], donc ce n'est pas grave », ajoutant que « si vous publiez, partagez, relayez [ces contenus], cela peut inciter au suicide ou à l'auto-mutilation.¹⁵⁰»

Malgré ces inquiétudes, TikTok, tout comme Instagram, continue d'afficher les mentions « J'aime » par défaut.

5.2 LES CONSÉQUENCES DES RÉSEAUX SOCIAUX ADDICTIFS POUR LA SANTÉ DES JEUNES

Les technologies numériques ont indéniablement joué un rôle positif dans la réduction des obstacles aux informations et services en matière de santé, par exemple en permettant aux psychologues de s'entretenir avec des patient-e-s à distance pendant les confinements liés à la pandémie de COVID-19, en permettant aux jeunes de rester en contact avec leurs ami-e-s et réseaux de soutien et en donnant aux groupes marginalisés, comme les adolescent-e-s LGBTI, la possibilité de découvrir leur identité et de s'exprimer.¹⁵¹ Mais il est également de plus en plus reconnu qu'en l'absence de réglementation étatique efficace, les plateformes de réseaux sociaux exposent les enfants et les jeunes à des risques systémiques.¹⁵²

Parmi les réponses aux questions de l'enquête menée pour ces recherches, la diversité d'idées et les opportunités qu'offrent les réseaux sociaux ont été saluées, mais de nombreuses personnes ont

147. Elia Abi-Jaoude, Karline Treurnicht Naylor et autres, "Smartphones, social media use and youth mental health", 10 février 2020, Canadian Medical Association Journal, Volume 192, numéro 6, <https://www.cmaj.ca/content/192/6/E136>; Pour obtenir un résumé d'autres recherches sur la comparaison fondée sur les réseaux sociaux et ses conséquences pour la santé mentale, voir Administrateur de la santé publique des États-Unis, "Social Media and Youth Mental Health" (déjà cité).

148. Instagram, "Giving people more control on Instagram and Facebook", 26 mai 2021, <https://about.instagram.com/blog/announcements/giving-people-more-control>.

149. Entretien mené à Nairobi, le 9 mars 2023.

150. Entretien mené à distance le 4 mai 2023.

151. Benjamin Hanckel et Shiva Chandra, "How young LGBTQIA+ people used social media to thrive during COVID lockdowns", The Conversation, 15 mars 2021, <https://theconversation.com/how-young-lgbtqia-people-used-social-media-to-thrive-during-covid-lockdowns-156130>.

152. Voir par exemple le Rapport de la Rapporteuse spéciale sur le droit qu'a toute personne de jouir du meilleur état de santé physique et mentale possible : Innovation et technologies numériques et droit à la santé, 21 avril 2023, Doc. ONU A/HRC/53/65 ; Comité des droits de l'enfant, Observation générale no 25 (déjà cité).

également fait part de préoccupations quant aux conséquences pour leur santé mentale de ce qu'elles ont qualifié de réseaux sociaux « addictifs ». Des jeunes ont indiqué avoir un sentiment d'« anxiété » et de « gêne » vis-à-vis des « images [corporelles] irréalistes » affichées dans le fil d'actualité. Une femme du Kenya a déclaré qu'elle pensait que les recommandations générées par algorithme des plateformes de réseaux sociaux (que les participant-e-s appellent « l'algorithme ») avaient « repéré » ses problèmes de santé mentale. D'autres personnes ont indiqué avoir vu des publications qui, d'après elles, auraient exacerbé leur anorexie autodéclarée, qu'elles ont qualifiées de « bouleversantes », et qui les auraient poussées à l'« auto-mutilation » et leur auraient fait « peur¹⁵³».

Dans des recommandations qu'il a émises en 2023, l'administrateur de la santé publique des États-Unis a clairement mis en garde contre les effets de la conception addictive des réseaux sociaux :

« Un lien a été établi entre l'utilisation excessive et problématique des réseaux sociaux, comme l'utilisation compulsive ou incontrôlable, et les problèmes de sommeil, les troubles de l'attention et les sentiments d'exclusion chez les adolescent-e-s... Des études à petite échelle ont montré que les personnes qui faisaient une utilisation fréquente et problématique des réseaux sociaux pouvaient subir des changements de leur structure cérébrale analogues aux changements constatés chez les personnes souffrant d'une addiction aux stupéfiants ou au jeu.¹⁵⁴»

Les recommandations indiquaient également que la mauvaise qualité de sommeil non seulement avait des conséquences pour la concentration et la réussite scolaires, mais qu'elle était « également liée à un développement neurologique altéré du cerveau des adolescent-e-s, à des symptômes dépressifs et à des pensées ou comportements suicidaires. »

Les recherches sur la manière dont l'utilisation excessive des réseaux sociaux pourrait altérer les structures cérébrales sont encore à un stade précoce, mais des études pointent vers des similarités entre les changements cérébraux observés dans le cadre d'une addiction aux stupéfiants (par exemple la cocaïne) et ceux liés à une utilisation addictive des réseaux sociaux.¹⁵⁵ Les deux comportements entraînent une libération de forts taux de dopamine, ce qui entraîne un sentiment temporaire d'« euphorie », qui donne ensuite à la personne un sentiment de déprime lorsqu'elle cesse d'utiliser la plateforme.¹⁵⁶ Ces conclusions pointent vers l'importance d'une réglementation étatique, similaire à la réglementation d'autres substances addictives, afin de garantir le droit des enfants et des jeunes à la santé.

TikTok doit clairement être au fait des risques de préjudices pour la santé mentale et physique des enfants et des jeunes adultes que représentent l'utilisation compulsive de la plateforme ou les schémas d'utilisation nocifs. Selon les résultats de recherches présentés ici, TikTok pourrait exposer les enfants et les jeunes à des troubles du sommeil et de l'attention et pourrait contribuer à des modifications de la structure cérébrale analogues à celles observées chez les personnes souffrant d'une dépendance à la drogue.

Et pourtant la réponse de l'entreprise n'a pas été suffisante. En mars 2023, TikTok a annoncé l'introduction d'un nouvel outil de gestion du temps passé devant l'écran, qui impose aux personnes de moins de 18 ans d'augmenter manuellement le temps qu'elles peuvent passer sur l'application une fois qu'elles ont atteint la limite quotidienne de 60 minutes.¹⁵⁷ L'efficacité de l'outil n'a pas encore été évaluée de manière indépendante, mais étant donné qu'il place la responsabilité de définir des limites de temps d'utilisation de TikTok sur des adolescent-e-s qui se décrivent largement comme «

153. Réponses au questionnaire en ligne d'Amnesty International, octobre-novembre 2022.

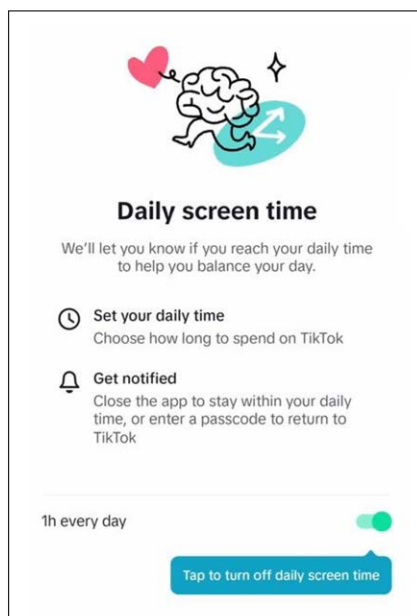
154. Administrateur de la santé publique des États-Unis, "Social Media and Youth Mental Health" (déjà cité).

155. Qinghua He, Ofir Turel et autres, "Brain anatomy alterations associated with Social Networking Site (SNS) addiction", mars 2017, Scientific reports, 7, 45064. <https://www.nature.com/articles/srep45064>.

156. Scope, "Addictive potential of social media, explained", 29 octobre 2021, <https://scopeblog.stanford.edu/2021/10/29/addictive-potential-of-social-media-explained/>.

157. TikTok, "New features for teens and families on TikTok", 1er mars 2023, <https://newsroom.tiktok.com/en-us/new-features-for-teens-and-families-on-tiktok-us>.

dépendants », il est probable qu'elle soit limitée.¹⁵⁸ L'incitation de TikTok est seulement une suggestion de limiter le temps d'utilisation, qui peut facilement être ignorée et qui ne présente pas d'avertissement sur les risques pour la santé.



A single click takes young users from the daily screen time limit notification to a page which lets them permanently disable the feature.

De plus, le changement ne s'applique qu'aux utilisateurs et utilisatrices que le système identifie comme des enfants, alors que l'efficacité du système de vérification de l'âge de TikTok a été remise en question.¹⁵⁹ En effet, l'Office britannique des communications (Ofcom) a constaté que 16 % des enfants britanniques de trois et quatre ans ont accès à TikTok.¹⁶⁰ En avril 2023, l'autorité de réglementation du Royaume-Uni en charge de la protection des données et du respect de la vie privée a condamné TikTok à une amende de 12,7 millions de livres britanniques pour avoir permis à des enfants de moins de 13 ans d'utiliser la plateforme en 2020. Des entretiens menés auprès d'enfants et de jeunes pour le présent rapport suggèrent que de nombreuses personnes de moins de 18 ans mentent sur leur âge pour accéder à la plateforme et contourner les mesures de protection des enfants ou encore accèdent à TikTok par des comptes d'adultes. Aucun de ces enfants ne recevrait alors l'information sur la limite de temps devant l'écran.

Des processus plus efficaces de vérification de l'âge pourraient empêcher les enfants de moins de 13 ans d'accéder à la plateforme et veiller à ce que les enfants de plus de 13 ans bénéficient de mesures de protection. Cependant, les outils de vérification de l'âge actuellement disponibles, qui offrent un plus haut niveau de certitude que la simple déclaration de l'utilisateur ou l'utilisatrice, nécessitent inévitablement une collecte de données plus importante, ce qui soulève des inquiétudes quant au droit au respect de la vie privée des enfants. Ces outils ne règlent en outre pas les problèmes structurels du modèle économique fondé sur la surveillance évoqués dans le présent rapport.

Certaines des personnes interrogées ont déclaré avoir essayé de limiter leur utilisation de TikTok et d'autres plateformes de réseaux sociaux, notamment par des applications de gestion du temps passé devant l'écran, mais avoir trouvé ces outils inefficaces pour faire cesser la tentation. Nikki, une femme de 24 ans des Philippines, a expliqué : « J'ai essayé de faire en sorte d'être prévenue en installant une notification, mais cela n'a pas vraiment fonctionné. Je me retrouve à négocier cinq minutes de plus avec cette application, alors que c'est seulement une notification. » D'autres ont indiqué avoir contourné les efforts déployés par leurs parents pour restreindre leur utilisation des réseaux sociaux. Par exemple, 13 des 15 élèves d'établissements d'enseignement secondaire de Mombasa, au Kenya, ont admis duper leurs parents en prétendant faire leurs devoirs sur leur téléphone alors qu'ils étaient en réalité sur les réseaux sociaux.¹⁶¹

158. Administrateur de la santé publique des États-Unis, "Social Media and Youth Mental Health" (déjà cité), p. 9 ; Jacqueline Nesi, Supreet Mann & Michael B. Robb, "Teens and mental health: How girls really feel about social media", 2023 (déjà cité), p. 6.

159. *New York Times*, "A third of TikTok's U.S. users may be 14 or under, raising safety questions", 14 août 2020, <https://www.nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html> ; BBC, "TikTok fined £12.7m for misusing children's data", 4 avril 2023, <https://bbc.co.uk/news/uk-65175902>.

160. *The Guardian*, "TikTok being used by 16% of British toddlers, Ofcom finds", 29 mars 2022, <https://theguardian.com/technology/2022/mar/29/tiktok-being-used-by-16-of-british-toddlers-ofcom-finds>.

161. Discussion de groupe, Mombasa, Kenya, 17 mars 2023.

L'un des arguments récurrents des discussions de groupe avec des enfants et des jeunes était que le rôle constructif ou destructif que les réseaux sociaux jouent dans la vie d'une personne est une question d'autodiscipline et d'utilisation responsable. La capacité d'action des enfants et des jeunes doit être respectée et des mesures de sensibilisation et d'éducation sont nécessaires pour informer leurs décisions et leur donner les moyens de mieux gérer les environnements numériques et d'identifier les risques et les préjudices. Cependant, il est extrêmement inquiétant que le manque de sécurité dans l'environnement en ligne, qui peut avoir des conséquences pour le droit à la santé, ait été normalisé. On ne peut exiger des enfants et des jeunes qu'ils soient responsables d'éviter les préjudices et d'atténuer les risques d'une plateforme conçue pour assurer leur utilisation continue à tout prix. La responsabilité de créer un environnement sûr, dénué de préjudices, revient aux entreprises de réseaux sociaux qui ont conçu, géré et profité de ces plateformes.

Le chapitre 7 aborde plus en détail les responsabilités de ces entreprises et le non-respect par TikTok de ses responsabilités.

6. DESCENTE EN SPIRALE

« Quand je n'avais pas le moral, je pense que 80 % [des contenus] étaient liés à la santé mentale. C'est une spirale infernale, parce que ça commence avec une seule vidéo. Si une vidéo parvient à capter ton attention, même si tu ne l'aimes pas, elle t'est de nouveau présentée quand tu ouvres TikTok la fois suivante et, parce qu'elle te semble familière, tu la regardes à nouveau et alors, sa fréquence d'apparition dans ton fil augmente exponentiellement ».

Luis, 21 ans, étudiant à Manille¹⁶²

6.1 COMMENT LE FIL « POUR TOI » DE TIKTOK POUSSE À L'EXTRÊME UN SYSTÈME INTRINSÈQUEMENT DANGEREUX

Au cœur de la stratégie d'engagement de TikTok se trouve la page « Pour toi », qui constitue le fil par défaut sur lequel les utilisateurs et utilisatrices de l'application se voient présenter un flux rapide et personnalisé de vidéos courtes. Outre sa nature addictive, cette fonctionnalité se trouve au cœur de préoccupations concernant le rôle que joue la plateforme dans l'exposition de mineur-e-s et de jeunes à des contenus potentiellement nocifs.

Les débats sur une réglementation de la plateforme se sont d'abord concentrés sur les mesures à mettre en place pour obliger les entreprises à supprimer rapidement les publications illégales et injurieuses de la masse de plus en plus importante de contenu généré par les utilisateurs et utilisatrices. Au milieu de ce flux de désinformation et de publications d'incitation qui inonde les plateformes dans des contextes de polarisation politique et de conflits, l'attention du public s'est

162. Luis (pseudonyme), 21 ans, étudiant de premier cycle à Manille, a été interviewé en personne le 6 mai 2023.

finalement orientée vers les méthodes de *diffusion* des contenus activement utilisées par les plateformes de réseaux sociaux.¹⁶³ Comme l'a expliqué une personne anonyme employée par Facebook:

« Nous disposons [...] de preuves irréfutables montrant que les éléments de base du fonctionnement de notre produit, tels que la viralité, les recommandations et l'optimisation de l'engagement, contribuent de façon considérable à la propagation de ce type de discours sur la plateforme. [...] Le fonctionnement de notre plateforme n'est pas neutre.¹⁶⁴»

Des experts des Nations unies, Amnesty International et bien d'autres organisations de la société civile ont recueilli des informations montrant comment les principaux systèmes de recommandation de contenus de plateformes de réseaux sociaux avaient contribué à diffuser et amplifier la « propagande, les contenus extrémistes et la désinformation » lors de récents conflits armés et de contextes de violence politique.¹⁶⁵

Bien que ce rapport porte sur un type différent de contenu amplifié (contenu à caractère dépressif et relatif à l'automutilation) et un ensemble de risques différent (préjudices potentiels pour les mineur-e-s et les jeunes utilisateurs et utilisatrices ayant des problèmes de santé mentale préexistants et qui sont exposés à de tels contenus), la source de ces risques est fondamentalement la même : un système de recommandation algorithmique de plateforme de réseau social, conçu pour créer un fil personnalisé de contenus viraux qui maintiendra l'utilisateur ou l'utilisatrice engagé-e le plus longtemps possible.

En dépit des preuves montrant que les éléments de base du fonctionnement des recommandations personnalisées, de la viralité et de l'« optimisation » conçus pour accroître au maximum l'engagement des utilisateurs et utilisatrices non seulement génèrent des risques de violations des droits humains, mais ont déjà contribué à ce que des préjudices réels se produisent, des plateformes de réseaux sociaux telles que TikTok, Instagram et Facebook continuent de s'en servir pour accomplir leurs objectifs. En effet, le succès de la stratégie d'engagement de TikTok basée sur les recommandations personnalisées a poussé sa concurrente, Meta (société mère d'Instagram et de Facebook) à l'imiter et à investir davantage dans le perfectionnement de ses algorithmes.¹⁶⁶

Pourtant, jusqu'à maintenant, les gouvernements ne sont toujours pas parvenus à proposer des dispositions législatives pour mettre un terme à l'« optimisation » permanente des plateformes en faveur d'une personnalisation des contenus basée sur l'inférence, de la viralité et de la maximisation de l'engagement des utilisateurs et utilisatrices. TikTok fait ouvertement la promotion de sa capacité à créer un fil de plus en plus personnalisé grâce à une analyse des interactions de l'utilisateur ou de l'utilisatrice avec la plateforme, ce qui revient, dans les faits, à un profilage desdits utilisateurs, dès que ceux-ci se mettent à utiliser la plateforme. Selon les propres mots de TikTok :

163. Amnesty International, *Myanmar. L'atrocité des réseaux sociaux* (déjà cité) ; Rapport de la rapporteuse spéciale des Nations unies sur la promotion et la protection du droit à la liberté d'opinion et d'expression, *Désinformation et liberté d'opinion et d'expression*, 13 avril 2021, doc. ONU A/HRC/47/25 ; « *Désinformation et liberté d'opinion et d'expression pendant les conflits armés* », 12 août 2022, doc. ONU A/77/288 ; Election Integrity Partnership, "The Long Fuse: Misinformation and the 2020 Election", 2 mars 2021, <https://www.atlanticcouncil.org/in-depth-research-reports/the-long-fuse-eip-report-read>.

164. « Facebook Papers », "What is Collateral Damage?", 12 août 2019, cité dans Amnesty International, *Myanmar. L'atrocité des réseaux sociaux* (déjà cité).

165. Rapporteuse spéciale des Nations unies sur la promotion et la protection du droit à la liberté d'opinion et d'expression, « *Désinformation et liberté d'opinion et d'expression pendant les conflits armés* », 12 août 2022, doc. ONU A/77/288 ; Amnesty International, *Myanmar. L'atrocité des réseaux sociaux* (déjà cité) ; *The Atlantic*, "Who's behind #IStandWithPutin?", 5 avril 2022, <https://www.theatlantic.com/ideas/archive/2022/04/russian-propaganda-zelensky-information-war/629475> ; Hannah Porter, "A conversation on fighting disinformation in Yemen", Yemen Policy Center, mars 2022, <https://www.yemenpolicy.org/a-conversation-on-fighting-disinformation-in-yemen/> ; Paul Barrett et Justin Hendrix, "A Platform 'Weaponized': How YouTube Spreads Harmful Content – And What Can Be Done About It", NYU Stern Center for Business and Human Rights, juin 2022, <https://www.stern.nyu.edu/experience-stern/faculty-research/platform-weaponized-how-youtube-spreads-harmful-content-and-what-can-be-done-about-it> ; Sara Brown, "In Russia-Ukraine War, Social Media Stokes Ingenuity, Disinformation", MIT Sloan School of Management, 6 avril 2022, <https://mitsloan.mit.edu/ideas-made-to-matter/russia-ukraine-war-social-media-stokes-ingenuity-disinformation>.

166. TechCrunch, "Meta expects recommendation models 'orders of magnitude' bigger than GPT-4. Why?", 29 juin 2023, <https://techcrunch.com/2023/06/29/metabehavior-analysis-model-is-orders-of-magnitude-bigger-than-gpt-4-why/?guccounter=1> ; Platformer, "Meta's Nick Clegg on how AI is reshaping the feed", 30 juin 2023, <https://www.platformer.news/p/metabehavior-analysis-model-is-orders-of-magnitude-bigger-than-gpt-4-why/?guccounter=1>.

« Dès que tu t'embarques dans l'aventure TikTok, tes premières mentions "J'aime", tes premiers commentaires et tes replays vont générer des recommandations, le système commençant à connaître tes goûts en matière de vidéos... Avec le temps, ces recommandations devraient devenir de plus en plus personnalisées. Suivre de nouveaux comptes, explorer des hashtags, des sons, des effets et des sujets populaires : tout cela va contribuer à affiner ton fil.¹⁶⁷»

De précédents rapports d'Amnesty International sont entrés en détail dans les raisons pour lesquelles la collecte massive de données personnelles qui permet ce genre d'« optimisation », fondée sur un profilage personnel plutôt que sur des signaux donnés de manière intentionnelle, constitue une forme de surveillance de masse de la part des entreprises, qui bafoue le droit au respect de la vie privée et risque de contribuer à des préjudices indirects qui vont entraîner des atteintes aux droits humains supplémentaires, ou de causer directement de tels préjudices.¹⁶⁸ Un document Facebook (aujourd'hui Meta) ayant été divulgué à l'insu de l'entreprise a révélé, en 2017, que la compagnie affirmait être capable d'identifier « les moments où les jeunes ont besoin d'un coup de pouce pour retrouver confiance en eux » et qu'elle était apparemment disposée à traiter, réellement, ces moments de tristesse et d'anxiété comme des opportunités commerciales pour les publicitaires.¹⁶⁹ Plus récemment, TikTok a fait la promotion de sa capacité à identifier les moments où les utilisateurs sont « émotionnellement impliqués » dans le contenu publicitaire qui leur est présenté.¹⁷⁰ Les objectifs explicites de Meta et TikTok représentent une attaque envers le droit au respect de la vie privée et à la liberté de penser des mineur.e.s et des jeunes. Des preuves existent, qui montrent que ces deux entreprises ont l'intention d'analyser et de déduire quelles sont les pensées et émotions privées des utilisateurs et des utilisatrices, et de les exploiter à des fins de profit commercial.

Non seulement les adolescents sont inclus dans les stratégies des plateformes visant à augmenter leur nombre d'utilisateurs et d'utilisatrices, mais ils sont particulièrement susceptibles d'être manipulés. Les recherches menées dans le domaine des neurosciences en sont encore à essayer de mieux comprendre comment l'environnement social des adolescents et les technologies modernes affectent le cerveau à ce stade de son développement où il est particulièrement malléable. Il semblerait que la « sensibilité accrue aux récompenses » propre aux adolescents ait le double effet d'encourager les comportements positifs et de rendre les adolescents plus susceptibles d'adopter des comportements à risque.¹⁷¹ La pensée rationnelle, contrairement aux processus affectifs, n'atteint son stade de développement final qu'à l'âge de 25 ans environ.¹⁷² Par conséquent, les adolescents ont bien plus tendance que les adultes à agir de façon impulsive et d'avoir des réactions émotionnelles fortes face à des stimuli extérieurs.¹⁷³ Cela signifie qu'il faut faire preuve de prudence, non seulement à l'égard des effets négatifs de la surveillance des utilisateurs et utilisatrices et de son utilisation pour cibler la publicité, mais aussi, plus largement, à l'égard de l'amplification ciblée des contenus.

Les plateformes de réseaux sociaux jouent un rôle majeur à ce stade clé du développement dans la vie des jeunes et dans leur démarche de découverte de soi. Des documents d'entreprise internes,

167. TikTok, "What is the 'For You' feed?", <https://www.tiktok.com/creators/creator-portal/how-tiktok-works/whats-the-for-you-page-and-how-do-i-get-there> (page consultée en juin 2023).

168. Amnesty International, *Les géants de la surveillance*. (déjà cité) ; Amnesty International, *I feel exposed: Caught in TikTok's Surveillance Web*, (déjà cité).

169. MIT Technology Review, "Is Facebook targeting ads at sad teens?", mai 2017, technologyreview.com/2017/05/01/105987/is-facebook-targeting-ads-at-sad-teens/.

170. Access Now, "Open letter: TikTok's 'Focused View' feature must respect human rights law", 7 février 2023, <https://www.accessnow.org/open-letter-tiktoks-focused-view/>.

171. Zara Abrams, "What neuroscience tells us about the teenage brain", 25 août 2022, Monitor on Psychology, Volume 53, Numéro 5, <https://www.apa.org/monitor/2022/07/feature-neuroscience-teen-brain>.

172. University of Rochester Medical Centre, Health Encyclopaedia, "Understanding the teen brain", 2023, <https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=1&contentid=3051> (consulté le 14 septembre 2023).

173. BBC, "The biggest myths of the teenage brain", 7 septembre 2022, <https://www.bbc.com/future/article/20220823-what-really-goes-on-in-teens-brains>.

en particulier les « Facebook Papers¹⁷⁴», ainsi que des recherches indépendantes montrent que les grandes plateformes exercent une influence énorme sur l'humeur, les perceptions et l'image de soi des enfants et des jeunes par le biais des contenus qu'elles recommandent et des modèles de plateforme qu'elles emploient.¹⁷⁵

Comme l'a déjà affirmé le Comité des droits de l'enfant, la façon dont les plateformes de réseaux sociaux utilisent les déductions tirées de la collecte de données massive à laquelle elles procèdent pour profiler leurs utilisateurs et utilisatrices et « personnaliser » leurs recommandations présente un risque tout à fait réel d'interférence avec les pensées et les émotions de l'enfant, de façon telle à compromettre sa liberté de pensée.¹⁷⁶

De nombreuses études recueillent des informations sur ces risques systémiques, notamment dans le cadre de récents rapports de recherche qui mettent en lumière le rôle que joue TikTok dans l'amplification de contenus problématiques et nocifs. En 2021, Reset Australia a révélé la façon dont le fil « Pour toi » de TikTok promouvait des contenus représentant et encourageant des stéréotypes ethniques et de genre.¹⁷⁷ La même année, le Wall Street Journal a montré comment TikTok happait ses utilisateurs et utilisatrices dans des « spirales » de vidéos sur les troubles de l'alimentation¹⁷⁸ et de contenus à caractère dépressif.¹⁷⁹ En 2022, une étude du Center for Countering Digital Hate sur l'amplification, par TikTok, de contenus liés aux troubles de l'alimentation et à l'automutilation a révélé que les nouveaux comptes dont le nom incluait le mot-clé « loseweight » (perdre du poids) avaient plus de chance de recevoir des recommandations pour ces deux types de contenus préjudiciables.¹⁸⁰ En mars 2023, une autre étude à petite échelle a démontré que le fil « Pour toi » de TikTok n'exigeait qu'un « engagement faible » de la part des jeunes utilisateurs et utilisatrices pour commencer rapidement à recommander des publications extrêmes relatives au suicide et à l'idéologie misogyne et violente des « incels¹⁸¹».

Ces résultats mettent en lumière certains des potentiels préjudices en chaîne liés au système de recommandation de TikTok, qui ajoutent des risques pour la santé des enfants et des jeunes aux risques pour le droit au respect de la vie privée et à la liberté de pensée dont il a précédemment été question.

Des parents endeuillés affirment que la sélection de contenus de TikTok a déjà coûté des vies. Une action en justice intentée aux États-Unis en juin 2022 soutient que TikTok a joué un rôle dans la mort d'au moins sept mineur-e-s ayant participé au « blackout challenge », un défi dans le cadre duquel les participant-e-s partagent des vidéos d'eux en train de s'étouffer, des vidéos qui, d'après l'une des multiples actions intentées, ne sont pas seulement accessibles par le biais d'une recherche, mais mises en avant dans le fil « Pour toi¹⁸²». TikTok conteste l'allégation des parents, ayant déclaré que le défi « exist[ait] déjà avant [la] plateforme et n'a jamais été une tendance TikTok¹⁸³».

174. *Wall Street Journal*, "Facebook knows Instagram is toxic for teen girls company documents show" (déjà cité).

175. Reset Australia, "Surveilling Young People Online: An Investigation Into TikTok's Data Processing Practices", juillet 2021, https://au.reset.tech/uploads/resettechaustralia_policymemo_tiktok_final_online.pdf; Jacopo Pruccoli et autres, "The use of TikTok among children and adolescents with eating disorders: Experience in a third-level public Italian center during the SARS-CoV-2 pandemic", 30 juillet 2022, *Italian Journal of Pediatrics*, Volume 48, <https://doi.org/10.1186/s13052-022-01308-4>; Reset Australia, "Designing for Disorder: Instagram's Pro-Eating Disorder Bubble in Australia", avril 2022, <https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf>; Tech Transparency Project, "'Thinstagram': Instagram's algorithm fuels eating disorder epidemic", 8 décembre 2021, <https://www.techtransparencyproject.org/articles/thinstagram-instagram-algorithm-fuels-eating-disorder-epidemic>.

176. Comité des droits de l'enfant, *Observation générale no 25* (déjà cité), § 62.

177. Reset Australia, "Surveilling young people online" (déjà cité).

178. *Wall Street Journal*, "'The corpse bride diet': How TikTok inundates teens with eating-disorder videos", 17 décembre 2021, (déjà cité).

179. *Wall Street Journal*, "Investigation: How TikTok's algorithm figures out your deepest desires", 21 July 2021, <https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477>.

180. Center for Countering Digital Hate, *Deadly by Design*, 15 décembre 2022, <https://counterhate.com/research/deadly-by-design/>.

181. Ekô, "Suicide, Incels, and Drugs: How TikTok's Deadly Algorithm Harms Kids", mars 2023, https://s3.amazonaws.com/s3.sumofus.org/images/eko_Tiktok-Report_FINAL.pdf.

182. The Verge, "The TikTok 'blackout challenge' has now allegedly killed seven kids", 8 juillet 2022, <https://www.theverge.com/2022/7/7/23199058/tiktok-lawsuits-blackout-challenge-children-death>.

183. The Verge, "The TikTok 'blackout challenge' has now allegedly killed seven kids", 8 juillet 2022 (déjà cité).

Une grande partie des mineur-e-s et des jeunes qui ont participé aux recherches menées dans le cadre de ce rapport étaient tout à fait conscients des risques liés aux pratiques d'engagement des utilisateurs de TikTok par rapport à d'autres grandes plateformes.¹⁸⁴ « Sur toutes les plateformes que j'utilise, il y a un certain degré d'insécurité. Mais je crois que, de toutes les plateformes, c'est sur TikTok que je me sens réellement en danger » a expliqué « Nikki », 24 ans, qui habite à Manille.¹⁸⁵

En réponse aux questions qui leur étaient posées concernant les types de contenus sur lesquels ils étaient tombés dans leurs fils et qu'ils ne souhaitaient pas voir, les jeunes participant-e-s ont fréquemment fait mention de contenus à caractère sexuel ou d'une violence extrême. Cela a été le cas pour « Maria », élève de 17 ans à Manille, qui a expliqué qu'une vidéo qui était apparue dans son fil montrait le meurtre d'une femme, qu'elle pensait être un crime réel.¹⁸⁶ Dans le cadre d'une autre discussion de groupe avec de jeunes adultes à Kisumu, au Kenya, la plupart des participant-e-s ont fait part du fait qu'ils étaient tombés sur la même vidéo virale qui, d'après eux, montrait une femme se faisant déshabiller de force par un groupe d'hommes, puis soumettre à un acte de mutilation génitale.¹⁸⁷

6.2 DES RISQUES ACCRUS POUR LES JEUNES SOUFFRANT DE PROBLÈMES DE SANTÉ MENTALE

De nombreux jeunes déclarant souffrir de problèmes de santé mentale ont raconté aux chercheurs/chercheuses d'Amnesty International comment ils avaient été happés dans des « spirales » de contenus TikTok déclencheurs de réactions négatives et de contenus relatifs à la santé mentale. Ils ont eu l'impression que l'amplification, de la part de TikTok, de contenus liés à la santé mentale de nature problématique contribuait à aggraver leurs symptômes. Voici, par exemple, comment « Luis », un étudiant de 21 ans à Manille souffrant de troubles bipolaires diagnostiqués, a décrit son expérience avec le fil « Pour toi » de TikTok :

« Quand je n'avais pas le moral, je pense que 80 % [des contenus] étaient liés à la santé mentale. C'est une sorte d'engrenage. C'est une spirale infernale, parce que ça commence avec une seule vidéo. Si une vidéo parvient à capter ton attention, même si tu ne l'aimes pas, elle t'est de nouveau présentée quand tu ouvres TikTok la fois suivante et, parce qu'elle te semble familière, tu la regardes à nouveau et alors, sa fréquence d'apparition dans ton fil augmente exponentiellement¹⁸⁸».

Faisant écho aux expériences d'autres personnes, « Luis » a décrit comment des périodes d'engagement délibéré avec des vidéos TikTok reflétant ses propres pensées dépressives ou anxieuses le maintenaient piégé dans des boucles de contenus négatifs, même après que sa phase dépressive soit passée :

« TikTok part du principe que tu veux voir ce genre de contenu pendant un certain temps. Le truc, pour moi, c'est que les contenus que je préfère voir varient énormément. Par exemple, un jour, je suis très angoissé et je veux tel type de contenu. Un autre jour, je me sens très heureux

184. Rien qu'au sein des groupes de discussion kenyans, 40 participant-e-s au moins ont donné des exemples de cas où ils se sont sentis en danger sur TikTok ; par exemple, lorsqu'ils regardaient des vidéos qu'ils considéraient comme violentes (notamment des vidéos qu'ils pensaient être en train de montrer un viol et un meurtre) ou des contenus non adaptés à leur âge (par exemple, des vidéos pornographiques), qui apparaissaient dans leurs fils.

185. Entretien en ligne, Manille, Philippines, 6 mai 2023.

186. Entretien en ligne, Manille, Philippines, 9 mai 2023.

187. Discussion de groupe, Kisumu, Kenya, 13 mars 2023.

188. Luis (pseudonyme), étudiant de premier cycle à Manille âgé d'une vingtaine d'années, a été interrogé en personne le 6 mai 2023.

et je suis ravi par quelque chose. Mais le truc, c'est que je continue de voir le contenu super triste et déprimant, et ça me fiche en l'air. »

« Francis », étudiant dans la province de Batangas, aux Philippines, a lui aussi observé :

« Quand [j]’“aime” une vidéo triste qui [me] parle, tout à coup, toute ma page “Pour toi” est triste. Je me retrouve dans le “TikTok triste”. Ça affecte mon humeur¹⁸⁹. »

Une autre jeune personne participant au groupe de discussion expliquait :

« Comme je suis du genre à trop réfléchir, les contenus que je regarde me donnent [encore] plus tendance à le faire, comme les vidéos dans lesquelles on voit des gens malades ou qui s'autodiagnostiquent. Ça m'affecte mentalement et me donne l'impression que j'ai les mêmes symptômes, ce qui aggrave mon anxiété. Et je ne les cherche même pas, elles apparaissent juste dans mon fil¹⁹⁰. »

Huit jeunes adultes des Philippines ont déclaré que l'application leur avait présenté des blocs de contenus déclencheurs de réactions négatives, qui ont eu des effets néfastes sur leur état de santé et leur bien-être. Dans ce groupe se trouvaient « Cris », un jeune enseignant essayant de gérer son trouble du déficit de l'attention avec hyperactivité (TDAH¹⁹¹), « Daisy », 22 ans, qui souffre de problèmes liés à l'image corporelle,¹⁹² et « Lance », 23 ans, jeune homme transgenre souffrant d'une dysphorie corporelle autodéclarée.¹⁹³ D'autres jeunes personnes, en particulier aux Philippines, ont également fait part de l'impact négatif qu'avait l'amplification de la propagande politique sur leur santé mentale, ainsi que ce qu'ils percevaient comme de la mésinformation, dans le contexte de politique déjà fortement polarisé qui est le leur. Bien qu'il soit difficile d'attribuer les troubles de la santé mentale d'une personne à un seul et unique facteur, ces témoignages montrent qu'il existe un risque que TikTok puisse exacerber des problèmes de santé mentale préexistants chez les mineur-e-s et les jeunes.

6.3 EXPLORATION SYSTÉMATIQUE DU SYSTÈME DE RECOMMANDATION DE TIKTOK

L'enquête d'Amnesty International sur le système de recommandation de TikTok avait pour objectif de contribuer à recueillir de nouveaux éléments de preuve, et ce, de deux façons. Premièrement, elle a intégré à son enquête deux pays (le Kenya et les Philippines), qui ont été sous-représentés dans les études menées à ce jour sur le sujet. Deuxièmement, l'échantillon sur lequel l'enquête a porté était de taille plus importante que cela a souvent été le cas lors de précédentes études menées par des organisations de la société civile dans ce domaine.¹⁹⁴

6.3.1 MODÈLE DE RECHERCHE AUTOMATISÉ (À L'AIDE DE FAUX COMPTES)

Amnesty International et ses partenaires techniques, l'Algorithmic Transparency Institute et AI Forensics, ont mené un audit en deux parties du système de recommandations du fil « Pour toi » de TikTok pour les contenus liés à la santé mentale. Dans un premier temps, une analyse des recommandations faites à de faux comptes en ligne (« sock puppet accounts ») basés au Kenya et aux

189. Discussion de groupe à Tanauan, le 13 mai 2023.

190. Discussion de groupe à Tanauan, le 13 mai 2023.

191. Entretien mené à Manille le 6 mai 2023.

192. Entretien mené à Manille le 6 mai 2023.

193. Entretien mené à Manille le 6 mai 2023.

194. Voir par exemple Center for Countering Digital Hate, *Deadly by Design*, 15 décembre 2022 (déjà cité) ; Reset Australia, « Surveilling Young People Online: An Investigation Into Tiktok's Data Processing Practices », juillet 2021 (déjà cité).

États-Unis a été réalisée.¹⁹⁵ Dans un deuxième temps, une expérience exécutée manuellement a été menée, qui impliquait un compte au Kenya, un aux Philippines et un aux États-Unis.

6.3.2 MODÈLE DE RECHERCHE AUTOMATISÉ (À L'AIDE DE FAUX COMPTES)

Pour l'audit automatisé, les chercheurs/chercheuses ont créé 40 comptes automatisés avec quatre personas prédéfinies imitant chacune différents comportements adoptés par de jeunes personnes sur TikTok. Chaque compte a été paramétré pour fonctionner par sessions uniques de près de 60 minutes par jour pendant 10 jours.¹⁹⁶ Les comptes étaient divisés en sous-groupes correspondant à quatre comportements de défilement des contenus : 20 comptes ont été paramétrés pour imiter les comportements de personnes de 13 ans aux États-Unis, tandis que 20 autres imitaient les comportements de personnes de 13 ans au Kenya (sur lesquels 11 ont été inclus dans l'analyse)¹⁹⁷. L'âge des comptes (13 ans) a été choisi afin d'examiner les recommandations proposées à la plus jeune tranche d'âge autorisée à utiliser TikTok, aux comptes de laquelle TikTok applique des mesures de sécurité propres aux adolescents.¹⁹⁸ Afin de limiter la mesure dans laquelle l'étude pouvait encourager des contenus dangereux pour d'autres utilisateurs, les comptes ainsi créés n'attribuaient pas de mentions « J'aime », ne faisaient pas de commentaires, n'envoyaient pas de messages, n'effectuaient aucune recherche et ne s'abonnaient à aucun autre compte. Ils se contentaient de regarder deux fois les publications recommandées dans le fil « Pour toi » si, dans leurs descriptions, elles incluaient un terme d'une des listes prédéfinies par les chercheurs/chercheuses :

Sous-groupe 1 (comportement dépressif [effet d'entonnoir]) : Les comptes appartenant à ce sous-groupe imitaient le comportement de mineur-e-s s'intéressant à la santé mentale et qui regardent des contenus de plus en plus extrêmes quand ces derniers leur sont présentés. Initialement, ces comptes regardaient plusieurs fois des vidéos associées à une large sélection de 150 termes relatifs à la santé mentale. À mi-chemin de l'expérience, ils étaient ensuite paramétrés pour revoir uniquement les contenus dont les tags figuraient dans une liste plus spécifique de 64 termes, que nous avons sélectionnés avant le début de l'expérience comme étant souvent associés à des contenus à caractère ouvertement dépressif ou relatifs à l'automutilation.

Sous-groupe 2 (intérêt général pour la santé mentale) : Ces comptes visionnaient à plusieurs reprises des vidéos dont les tags figuraient dans une liste de 150 termes liés au domaine de la santé mentale.

Sous-groupe 3 (intérêt général [qualifiés de termes « neutres » dans les graphiques] et intérêt pour la santé mentale) : Les comptes de cette catégorie visionnaient plusieurs fois des vidéos contenant une sélection de 480 termes d'intérêt général, ainsi qu'une liste générale de 150 termes relatifs à la santé mentale.

Sous-groupe 4 (intérêt général [qualifiés de termes « neutres » dans les graphiques, groupe de contrôle]) : Ces comptes regardaient uniquement des vidéos associées à 480 termes d'intérêt général, comme « animalsoftiktok », « comedy » ou « traveltiktok¹⁹⁹».

195. L'audit automatisé n'a pas inclus de comptes aux Philippines en raison des limites du projet. Au lieu de quoi, la recherche s'est concentrée sur une analyse comparative d'un pays sous-représenté dans de précédents travaux dans ce domaine (le Kenya) ainsi que sur un pays et marché clé, où le débat public sur les risques présentés par les réseaux sociaux est actif.

196. Si une session venait à être perturbée par des problèmes techniques, une session supplémentaire était menée, afin que le compte soit actif pendant environ 10 heures en tout. Étant donné que TikTok définit la durée volontaire de connexion à 60 minutes par jour pour les moins de 18 ans, l'expérience des utilisateurs a été testée dans cette même limite de temps quotidienne, bien que celle-ci soit inférieure à la moyenne de temps que passent les adolescents sur TikTok, selon une étude de marché des États-Unis.

197. L'expérience a démarré avec 20 comptes paramétrés pour représenter des utilisateurs adolescents dans chacun des pays, mais seuls 11 comptes kenyans ont fonctionné jusqu'à ce que l'enquête se conclue. Seuls 11 comptes kenyans sur 20 ont mené à bien les dix sessions. Les autres ont été désactivés par TikTok avant d'avoir atteint leur session finale et ont donc été exclus de l'analyse.

198. Voir Annexe 3 : Réponse écrite aux questions de l'équipe de recherche d'Amnesty International.

199. La liste complète des termes figure à l'Annexe 2.

La liste générale de termes relatifs à la santé mentale incluait des termes tels que : « mentalhealth » (santé mentale), « mentalhealthmatters » (la santé mentale est importante), « sadtiktok » (tiktok triste), « schooldraining » (l'école draine mes forces), « struggling » (j'ai du mal), « ventingaccount » (compte pour se défouler), ainsi que des abréviations très courantes, des versions mal orthographiées de termes et des mots codés faisant référence à la dépression et à l'automutilation.²⁰⁰

La liste restreinte de termes relatifs à la santé mentale comprenait un sous-ensemble de termes de la liste générale, qui avaient été identifiés comme étant principalement associés à des contenus problématiques et dangereux lors d'expériences tests.²⁰¹

La principale question à laquelle cette étude cherchait à répondre était : dans quelle mesure des mineur-e-s et des jeunes se voient-ils présenter des vidéos nocives représentant et/ou idéalisant ou encourageant la dépression, l'automutilation et le suicide par le biais de recommandations algorithmiques ? L'hypothèse de travail étant que, si l'effet d'amplification était fort, les chercheurs/chercheuses observeraient une augmentation constante du volume de contenus relatifs à la santé mentale de nature potentiellement nocive présentés aux comptes des sous-groupes 1 et 2 (comportements dépressifs et intérêt général pour la santé mentale) par rapport aux comptes associés à des intérêts plus généraux.

Bien que l'étude ait impliqué un échantillon plus important que les précédents travaux de recherche menée par la société civile dans ce domaine, la taille de l'échantillon, de 40 comptes, n'a toujours pas été suffisamment grande pour mener à bien une quelconque vérification statistique de l'hypothèse. Par conséquent, l'analyse principale a consisté à comparer, de façon descriptive, le volume et la fréquence des contenus relatifs à la santé mentale vus par chaque sous-groupe. L'étude visait à apporter de nouvelles données par rapport aux précédents travaux menés dans ce domaine, en incluant deux groupes qui s'apparentaient à des groupes de contrôle, sous la forme de sous-groupes ne visionnant que des contenus d'intérêt général (qualifiés de « termes neutres » dans les graphiques), ou des contenus d'intérêt général et des contenus relatifs à la santé mentale. L'inclusion de ces deux groupes a permis aux chercheurs/chercheuses d'observer l'ampleur de l'amplification du système de recommandation, en fournissant une mesure de référence du volume et de la fréquence de contenus relatifs à la santé mentale que TikTok présente à un utilisateur « moyen ».

Il est très probable que des facteurs tels que la date, l'heure et la situation géographique soient pris en compte par l'algorithme de recommandation et qu'ils contribuent par conséquent à déterminer les types de contenus que TikTok propose à ses utilisateurs et utilisatrices. Pour veiller à ce que les comparaisons entre les sous-groupes soient aussi solides que possible, les chercheurs/chercheuses ont tenté de les contrôler en mettant en place des comptes automatisés pour accéder à la plateforme à des dates et heures spécifiques depuis le même endroit au Kenya et aux États-Unis. Cette démarche s'est heurtée à certains problèmes techniques, comme la suppression ou des dysfonctionnements de ces comptes automatisés. Les chercheurs/chercheuses ont cependant fait tout leur possible pour que toutes les données soient collectées de façon parallèle dans chaque sous-groupe.

6.3.3 CONCEPTION DE L'EXPÉRIENCE MANUELLE

Pour contrebalancer certaines des limites de l'approche automatisée, comme la nécessité de prédéfinir les termes qui permettraient aux comptes automatisés de reconnaître les contenus pertinents et l'incapacité de ces comptes d'identifier des publications pertinentes en fonction de l'ambiance, de la musique ou des légendes intégrées aux vidéos plutôt qu'en fonction des hashtags, Amnesty

200. La liste complète des termes figure à l'Annexe 2.

201. La liste complète des termes figure à l'Annexe 2.

International et AI Forensics ont également mené une expérience manuelle à petite échelle. Dans le cadre de cette dernière, des heures d'interaction d'un-e chercheur-euse avec les fils « Pour toi » de comptes nouvellement créés imitant eux aussi le comportement de jeunes âgés de 13 ans ont été enregistrées. Trois comptes ont été créés à l'aide de VPN, un pour chacun des pays suivants : Kenya, Philippines et États-Unis. Une fois encore, la personne chargée de la recherche s'est contentée de faire défiler le fil « Pour toi » et de regarder des publications deux ou trois fois si elles avaient un rapport avec la tristesse ou des problèmes de santé mentale.

6.4 MESURER ET CATÉGORISER LES CONTENUS NOCIFS AUX FINS DE L'ÉTUDE

6.4.1 ÉTABLIR UNE MESURE DE L'AMPLIFICATION

Le sens précis de « contenu préjudiciable » est rarement défini dans la littérature actuellement disponible sur les risques et les dommages liés à l'utilisation des réseaux sociaux.²⁰²

Ceci représente un défi de taille, étant donné que le contenu de TikTok est constitué d'un mélange de données vidéos, textuelles et audios et qu'il est, dans ce contexte, collecté à une échelle tellement vaste que toute démarche manuelle de catégorisation en devient impossible à réaliser. En dépit de ces difficultés, il était d'une importance cruciale pour l'enquête de classer les contenus, afin d'obtenir une mesure de l'amplification. Les chercheurs/chercheuses ont donc adopté la démarche suivante :

1. La principale mesure de l'amplification a été le pourcentage de vidéos recommandées qui contiennent un terme de la liste générale de termes relatifs à la santé mentale (soit dans un hashtag, soit dans le texte associé). Les chercheurs/chercheuses ont également analysé et représenté sous forme de graphique les termes qui apparaissent ensemble, afin de comprendre les liens qu'entretiennent les sujets des vidéos entre eux.
2. En s'appuyant sur la recherche qualitative menée avec des mineur-e-s et des jeunes et les informations obtenues lors d'entretiens avec des experts de la santé mentale, les chercheurs/chercheuses ont élaboré un cadre de catégorisation des contenus. Un panel d'examineurs et d'examinatrices a été réuni pour classer manuellement un échantillon de 540 vidéos TikTok conformément à ce cadre, chaque contenu passant indépendamment entre les mains de trois examineurs/examinatrices.
3. Les chercheurs/chercheuses ont ensuite analysé la correspondance entre les catégorisations manuelles et les hashtags et descriptions écrites des contenus. Cette démarche leur a permis de valider notre principale mesure de l'amplification, en évaluant combien de contenus intégrant un terme de la liste générale étaient considérés comme préjudiciables dans le cadre de l'exercice de catégorisation manuelle.

L'une des principales limites de cette mesure de l'amplification, c'est que les hashtags assignés aux vidéos servent à la fois à entraîner les robots de recherche à imiter les comportements de l'utilisateur/utilisatrice et à mesurer le volume de contenus relatifs à la santé mentale qui leur étaient présentés. Avec davantage de ressources, un travail de catégorisation manuelle supplémentaire aurait pu

202. Dans les recommandations qu'il a émises en 2023, l'administrateur de la santé publique des États-Unis les décrit comme suit : « par exemple, les contenus qui encouragent les troubles de l'alimentation, la violence, la consommation de drogues, l'exploitation sexuelle ou le suicide, ou qui évoquent des moyens de se suicider ». Voir : Administrateur de la santé publique des États-Unis, "Social Media and Youth Mental Health" (déjà cité).

permettre d'obtenir un résultat plus précis, voire contribuer à peaufiner les méthodes d'analyse quantitative. Les efforts de recherche de la société civile dans ce domaine demeurent restreints par des contraintes temporelles et budgétaires, auxquelles viennent s'ajouter des obstacles créés par le manque de cadres d'accès à la recherche adaptés et approuvés de manière indépendante au moment de la rédaction du présent rapport (la mise en œuvre du cadre d'accès à la recherche prévu par la Loi sur la sécurité numérique n'étant pas encore d'actualité). Ceci souligne encore une fois la nécessité que des efforts de réglementation soient fournis en dehors de l'Union européenne pour mettre en place des obligations similaires d'accès aux données ainsi que pour créer des obligations exécutoires pour les grandes plateformes de réseaux sociaux, qui les forcent à rendre publiquement compte des évaluations des risques et des actions d'atténuation qu'elles mènent.

6.4.2 CLASSER LES CONTENUS

Bien que l'enquête d'Amnesty International se soit intéressée davantage à l'effet cumulatif de l'amplification de certains types de contenus par TikTok sur une personne mineure sur la durée, plutôt que sur l'effet de défaillances individuelles de la modération des contenus, le codage des données a néanmoins amené les chercheurs/chercheuses à procéder à une catégorisation d'éléments de contenu individuels. Les catégories utilisées dans le présent rapport reflètent les informations fournies par différents experts et clinicien·ne·s ; ils ne peuvent cependant jamais que représenter une image approximative du contenu observé et de sa potentielle capacité à causer des préjudices à la santé mentale d'une personne.²⁰³ Le schéma de catégorisation adopté avait pour objectif de fournir une description transparente des différentes facettes des contenus relatifs à la santé mentale de nature potentiellement dangereuse qui étaient recommandés aux comptes créés dans le cadre de notre étude, plutôt qu'une opposition binaire entre contenus « préjudiciables » et « non préjudiciables ».

En réalité, en dehors de certains cas flagrants de contenus traumatisants, la perception des contenus et les réactions émotionnelles à ces derniers sont éminemment personnelles. Les entretiens menés par Amnesty International avec des mineur·e·s et des jeunes montrent que ce qu'une personne perçoit comme un élément déclencheur pour sa santé mentale, ou préjudiciable à celle-ci, dépend souvent de son expérience et de sa situation personnelles. Cela peut également varier avec le temps, en fonction des changements de l'état émotionnel et mental de la personne.

De plus, concernant les contenus générés par des utilisateurs et utilisatrices, des désaccords existent entre psychologues à propos des avantages et inconvénients du partage d'informations relatives à la santé mentale sur les réseaux sociaux par des personnes qui ne sont pas spécialisées dans le domaine. Comme il a été mentionné plus haut, des plateformes telles que TikTok offrent, pour beaucoup de gens, des possibilités sans précédent de s'exprimer ainsi qu'un sentiment de communauté, notamment pour les personnes qui ont les mêmes problèmes de santé mentale.²⁰⁴ Ces plateformes peuvent également donner accès à des informations touchant à la santé mentale conçues pour être facilement accessibles au plus grand nombre à des personnes qui n'ont autrement pas accès à ce genre de ressources. Cependant, des experts ont fait part de sérieuses inquiétudes concernant les clips vidéo express qui transmettent des informations et de potentielles mésinformations, ainsi que les contenus proposant de (prétendus) conseils médicaux sans avoir la moindre connaissance des personnes qui les regardent et de leurs besoins spécifiques, ou qui encouragent l'autodiagnostic fondé sur des informations parcellaires ou erronées.²⁰⁵

203. Des informations ont été fournies par S. Bryn Austin, Docteur en sciences (STRIPED/Faculté de médecine de Harvard), le Dr Amanda Raffoul, PhD (Faculté de médecine de Harvard), Zeyna Awan, psychologue à Nairobi, Kevin Gachee, psychologue à Nairobi, et Aliya Shah, psychologue à Nairobi.

204. John A. Naslund, Ameya Bondre et autres, "Social media and mental health: Benefits, risks, and opportunities for research and practice" (déjà cité).

205. Joe Martin, "The big issue: Mental health and the TikTok effect", avril 2023, <https://www.bacp.co.uk/bacp-journals/therapy-today/2023/april-2023/the-big-issue/>.

Il est important de noter que ces contenus ne font pas l'objet d'une consommation isolée, mais sont intégrés à un fil au défilement rapide qui, comme le montrent les recherches d'Amnesty International, peut rapidement être inondé de contenus liés à la santé mentale de nature problématique.

Outre les exemples de contenus entrant clairement dans les catégories définies dans le cadre de notre expérience, les chercheurs/chercheuses ont également repéré des publications recommandées difficiles à classer avec certitude, en raison, par exemple, de leur brièveté, qui en rendait le sens et l'intention peu clairs. Afin de compenser de potentiels biais personnels et de valider nos résultats, chaque publication s'est vu assigner séparément une catégorie par trois examinateurs/examinatrices sur la plateforme Junkipedia²⁰⁶ et, lorsque leurs avis différaient, la catégorisation finale a été basée sur une décision de la majorité.

Types de contenus relatifs à la santé mentale pouvant avoir des effets potentiellement nocifs, lorsqu'amplifiés à l'échelle

 <p>Publications relatives à des expériences vécues qui parlent d'anxiété, de dépression et d'automutilation (sans idéalisation)</p>	 <p>Publications montrant des personnes en état de détresse émotionnelle</p>	 <p>Publications décrivant des sentiments de solitude ou d'infériorité comme n'ayant pas de solution et/ou comme étant mérités</p>
 <p>Publications (y compris citations, textes, voix off) visant à représenter l'automutilation, la dépression et les pensées suicidaires comme inéluctables</p>	 <p>Descriptions dramatiques de traumatismes, de souffrances et de suicide</p>	 <p>Images traumatisantes</p>
 <p>Publications qui valorisent, idéalisent ou banalisent la dépression, l'anxiété, l'automutilation et le suicide (y compris par le biais de dessins, de bandes dessinées, de musique ou de légendes écrites)</p>	 <p>Publications communiquant de fausses informations sur la santé mentale, par exemple en dissuadant les utilisateurs/utilisatrices de chercher l'aide de professionnel-le-s ou de prendre des médicaments prescrits par des médecins, ou en encourageant les personnes à s'autodiagnostiquer sur la base d'informations erronées</p>	 <p>Publications qui mentionnent l'intention de s'automutiler ou de se suicider, ou qui expliquent comment s'automutiler ou se suicider, ou qui encouragent de quelque autre façon que ce soit l'automutilation ou le suicide</p>

206. Précisions techniques supplémentaires fournies en Annexe 1.

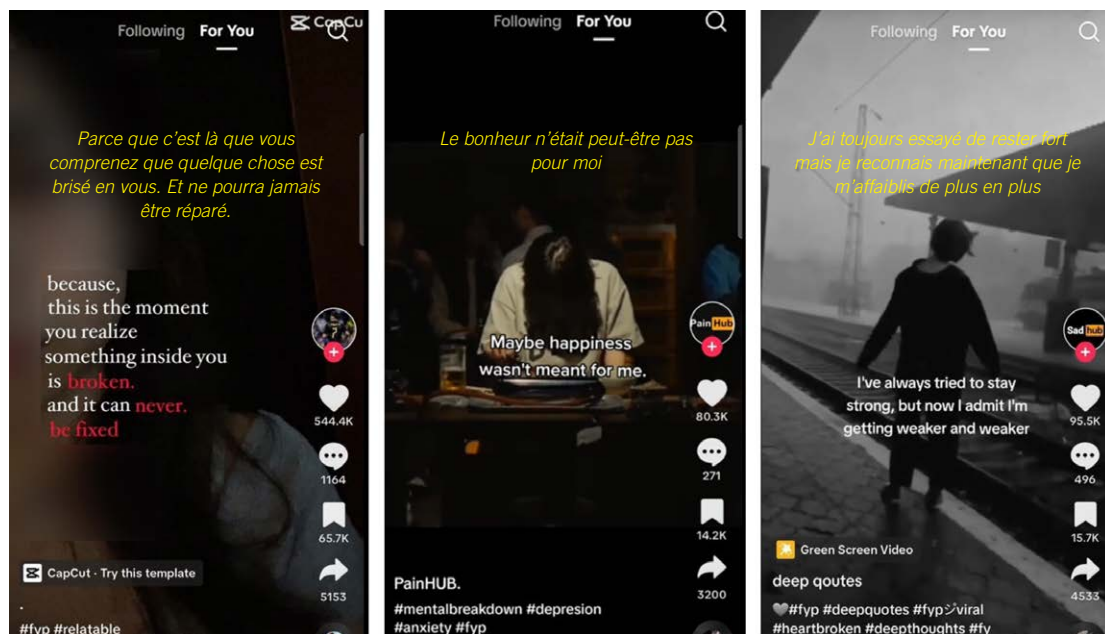
6.5 PRINCIPALES CONCLUSIONS

Les résultats des recherches d'Amnesty International sur TikTok montrent que les mineur-e-s et les jeunes qui regardent des contenus relatifs à la santé mentale sur la page « Pour toi » de la plateforme peuvent facilement tomber dans des « spirales » de contenus dangereux, notamment des vidéos qui idéalisent et encouragent les pensées dépressives, l'automutilation et le suicide.

Cela a été le cas pour les comptes automatisés paramétrés pour représenter des personnes de 13 ans au Kenya et aux États-Unis, certains contenus recommandés étant les mêmes pour les deux pays. Au bout de cinq ou six heures passées sur la plateforme, près d'une vidéo sur deux présentées aux comptes ayant signalé un intérêt pour la santé mentale était relative à ce sujet et potentiellement nocive, soit un volume dix fois plus important que celui présenté aux comptes n'ayant indiqué aucun intérêt pour le sujet.

Ces résultats reflètent ceux de l'analyse manuelle de 540 vidéos recommandées à des comptes au Kenya et aux États-Unis, qui ont montré une progression régulière de 17 % des vidéos présentées le premier jour étant catégorisées comme potentiellement nocives à 44 % des contenus le dixième jour.

L'analyse manuelle menée avec un compte représentant une personne de 13 ans au Kenya, un aux États-Unis et un aux Philippines a révélé un processus de personnalisation encore plus rapide dans chacun des pays, qui réduisait le fil de l'utilisateur/utilisatrice à une « spirale » encore plus prononcée de contenus relatifs à la santé mentale potentiellement nocifs en moins de 21 minutes (comparés à quatre ou cinq heures pour les comptes de l'étude automatisée).



Vidéos TikTok recommandées au cours de sa première heure d'existence à un compte d'adolescent géré manuellement

6.6 L'INTERET POUR LA SANTE MENTALE DECLENCHE L'EFFET DE « SPIRALE »

Lorsque le système de recommandation de la page « Pour toi » de TikTok apprend qu'un utilisateur ou une utilisatrice s'intéresse aux vidéos relatives à la santé mentale, il a de fortes chances de transformer ce signal en une « spirale » vertigineuse de contenus qui traitent de pensées dépressives et d'automutilation et les encouragent.

En raison de nos ressources limitées, l'analyse manuelle n'a permis de passer en revue qu'un échantillon de 540 vidéos sur les milliers qui ont été recommandées (voir section 6.6.4 pour plus de détails), ce qui signifie que la validité des résultats de l'échantillon pour l'ensemble de l'expérience est elle-même limitée. Néanmoins, cet échantillon semble indiquer que les pourcentages suivants, basés sur une analyse quantitative de toutes les recommandations de vidéos associées à la liste générale de termes liés à la santé mentale pour l'expérience globale, sous-estiment probablement le nombre réel de recommandations de vidéos dont le contenu est potentiellement dangereux. Le pourcentage de vidéos associées à un terme figurant dans la liste générale de termes liés à la santé mentale dans l'échantillon était inférieur de 6 % au pourcentage réel de vidéos classifiées manuelles comme étant potentiellement nocives (catégories 1 à 9) (le sujet de l'analyse de l'échantillon sera abordé dans la section 6.6.4).

L'analyse quantitative de l'expérience automatisée a montré que :

- L'effet d'amplification semble puissant. Les chercheurs/chercheuses ont observé une augmentation constante, au cours des cinq à huit premiers jours de l'expérience, du volume de contenus liés à la santé mentale présenté aux comptes ayant indiqué un intérêt pour ce type de contenus (sous-groupe 1 [comportements dépressifs] et 2 [intérêt général pour la santé mentale]), comparés aux comptes dont les intérêts étaient plus généraux (graphiques 1 et 2).
- Le volume de contenus relatifs à la santé mentale a atteint son premier pic entre le cinquième et le huitième jour. Durant cette période, il a représenté entre 40 et 55 % de l'ensemble des contenus présentés aux comptes de nos sous-groupes à comportements dépressifs ou ayant un intérêt général pour la santé mentale aux États-Unis (graphique 1). Les informations recueillies dans le cadre de notre exercice de catégorisation manuelle semblent indiquer que trois vidéos sur cinq identifiées comme étant relatives à la santé mentale en fonction des termes et hashtags utilisés dans leurs descriptions sont potentiellement nocives.
- Le volume de contenus relatifs à la santé mentale présenté à nos sous-groupes ayant indiqué des intérêts plus généraux est resté relativement bas (moins de 10 % de l'ensemble des contenus présentés) et constant sur toute la durée de l'étude, quel que soit le pays concerné (graphiques 1 et 2).
- Pour les comptes à effet d'entonnoir basés aux États-Unis, un déclin des contenus liés à la santé mentale a été observé après le cinquième jour, moment à partir duquel ces comptes étaient programmés pour ne reversionner que les contenus incluant un terme de la liste restreinte (graphique 1). L'analyse semble indiquer que cette diminution provient probablement du fait que la fréquence globale de contenus incluant un terme de la liste restreinte au sein du contenu total recommandé était très basse (< 1 % aussi bien avant qu'après l'effet d'entonnoir) pour quatre des cinq comptes. Par conséquent, les comptes étaient à la recherche de contenus spécifiques, qui ne semblent pas avoir été très disponibles sur la plateforme parmi les recommandations au moment du changement. En l'absence d'un signal efficace, le système de recommandation s'est probablement référé à d'autres facteurs pour façonner les fils « Pour toi » de ces comptes (par exemple, aux sujets et événements populaires du moment).

Figure 1: Percentages (means per subgroup) of recommended videos served to US accounts (whose descriptions include terms from the broad list of mental health-related terms per day²⁰⁷)

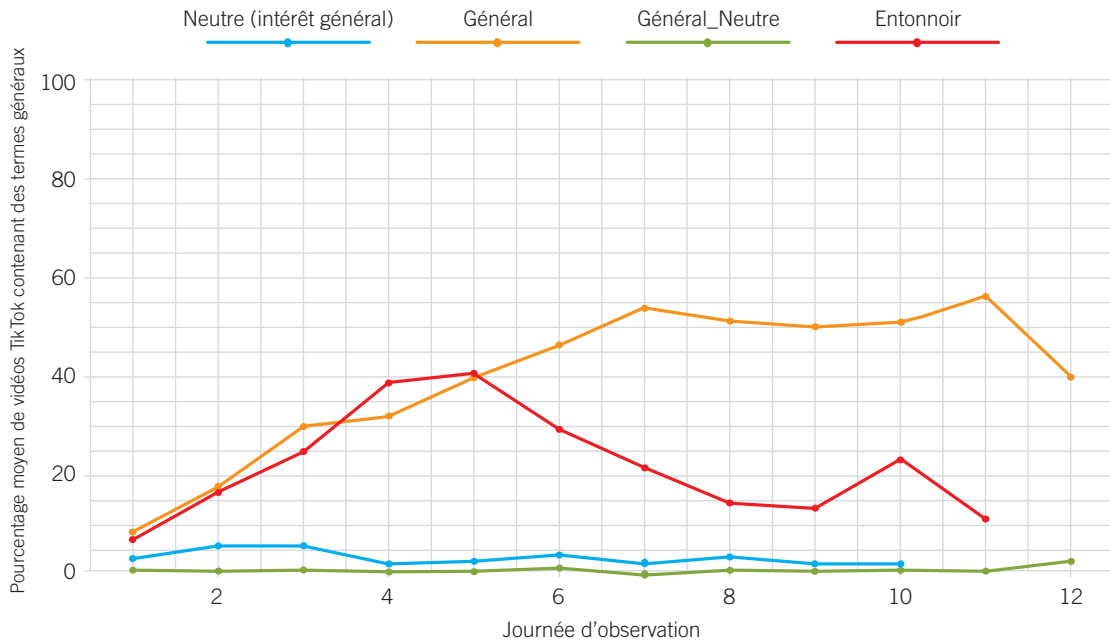
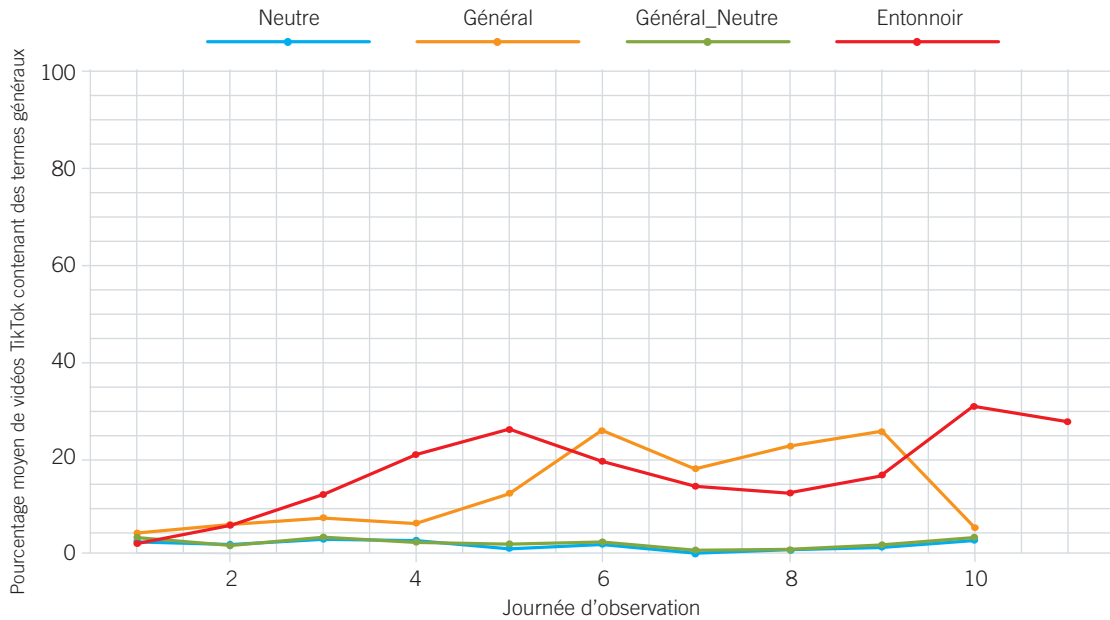


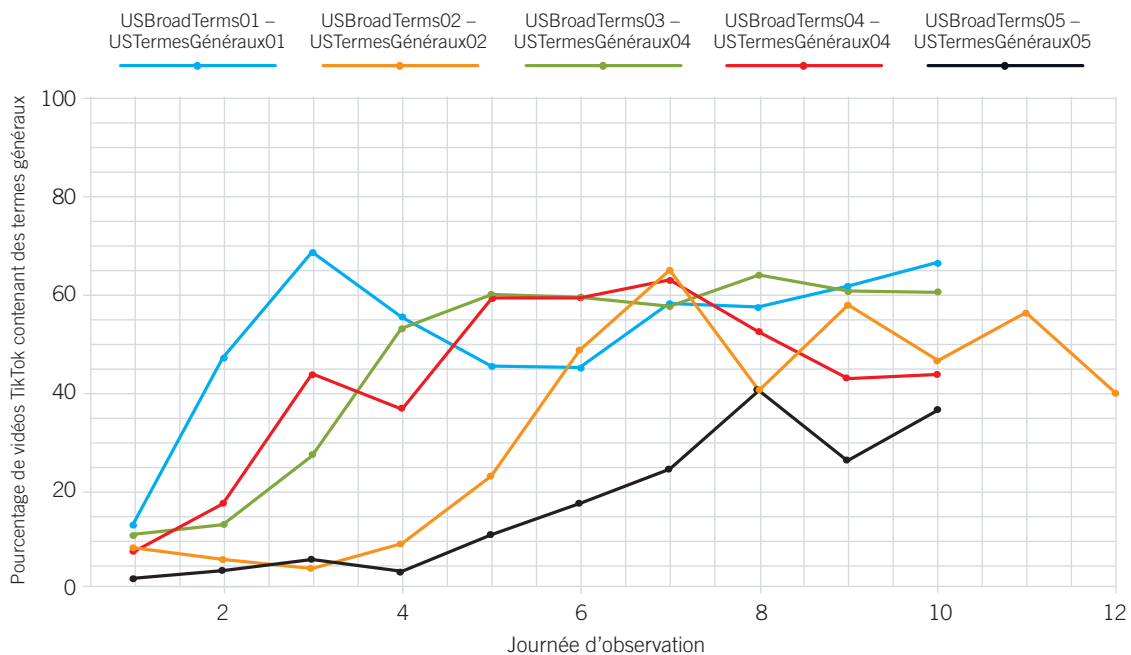
Figure 2: Percentage of recommended videos served to Kenyan accounts (whose descriptions include terms from the broad list of mental health-related terms per day)



207. Si, en raison de problèmes techniques, la durée de fonctionnement d'un compte était insuffisante lors d'une session planifiée d'une heure, cette session incomplète était fusionnée avec une seconde session, enregistrée le même jour. Pour le graphique 2 : Dans un cas, l'un des comptes kenyans n'a pas pu accomplir de session complète et aucune autre session n'a été enregistrée ce jour-là ; la session incomplète a donc été exclue de l'analyse.

- Les comptes kenyans à effet d'entonnoir, en revanche, ont continué de se voir présenter un volume plus constant de contenus relatifs à la santé mentale, avant et après le cinquième jour (graphique 2). L'analyse a montré que ce résultat était dû au petit nombre de comptes kenyans à effet d'entonnoir ayant atteint la fin de la période d'étude. Sur les deux comptes qui n'ont pas été suspendus, un seul a subi un effet de spirale important, qui a probablement incité TikTok à continuer de recommander des contenus liés à la santé mentale après le basculement vers la liste plus restreinte de termes que nous avons identifiés avant le début de l'expérience comme étant souvent associés à des contenus à caractère ouvertement dépressif ou relatifs à l'automutilation.

Figure 3: Percentage of recommended videos served to US accounts (group 2: broad mental health interest) whose descriptions include terms from the broad list of mental health-related terms per day



Quatre comptes sur les cinq ont atteint des pics de plus de 60 % de contenus relatifs à la santé mentale. Sur l'ensemble de leur fil, 20 % au moins de toutes les vidéos recommandées aux comptes basés aux États-Unis dans cette catégorie avaient un rapport avec la santé mentale (comptes associés aux termes de la liste générale).

Pour mettre ces résultats en contexte, « environ 13 pour cent » des vidéos du fil Instagram de l'adolescente britannique Molly Russell au cours des six mois qui ont précédé sa mort auraient été « relatifs au suicide, à la dépression ou à l'automutilation²⁰⁸ ». Dans cette affaire, le médecin légiste a affirmé que ce niveau d'amplification des contenus liés à la santé mentale était suffisant pour avoir contribué significativement à sa mort.

S'il avait été observé sur le compte d'une personne mineure ou jeune réelle, le pourcentage significativement plus élevé de contenus relatifs à la santé mentale, à la dépression et à l'automutilation

208. Politico, "Digital bridge: Platforms on the hook — Transatlantic AI rulebook — Let's talk data transfers", 6 octobre 2022 (déjà cité).

observé dans le cadre des expériences menées par Amnesty International²⁰⁹ sur TikTok, qui n'a résulté que d'un seul des signaux habituellement donnés par les utilisateurs (tels que les visionnages, les mentions « J'aime » et les abonnements), aurait exposé la personne en question à une perspective encore plus élargie. Pour les mineur-e-s et les jeunes en état de vulnérabilité, en particulier, cela risquerait d'aggraver des problèmes de santé mentale préexistants, en y ajoutant un risque supplémentaire de préjudice grave pour leur santé physique. Alors que les chercheurs/chercheuses travaillant sur ce projet n'ont pas aimé de vidéos ni activement cherché de contenus pertinents, et n'ont suivi aucun autre utilisateur ou utilisatrice afin d'éviter de contribuer à l'amplification de contenus potentiellement nocifs, de tels signaux, qui sont habituellement envoyés par la plupart des utilisateurs et utilisatrices, auraient pu entraîner un pourcentage encore plus élevé de contenus connexes.

La section 6.6.2 ci-dessous porte sur l'analyse manuelle de fils de comptes TikTok et examine de plus près les types de contenus qui leur ont été recommandés, en fonction des intérêts indiqués par les comptes pour les vidéos associées à des termes et hashtags liés à la santé mentale. En outre, les méthodes quantitatives apportent des informations relatives au sujet et au ton des recommandations. La carte de réseau ci-dessous montre les termes associés à la santé mentale les plus amplifiés sur l'ensemble des vidéos recommandées, et les liens qu'ils entretiennent avec des hashtags cooccurrents. Les termes les plus courants associés à la liste générale de termes liés à la santé mentale sont tout d'abord des termes génériques, tels que « mentalhealth » (santé mentale), « mentalhealthmatters » (la santé mentale est importante) et « mentalhealthawareness » (connaissances sur la santé mentale) (15 à 20 % de ces vidéos intégrant aussi « vent » [se défouler] dans leurs autres hashtags), suivis de « vent », « anxiety » (anxiété) et « sad » (triste).

209. Il faut émettre une réserve ici, du fait que les chercheurs/chercheuses d'Amnesty International n'ont pas pu examiner les catégories appliquées au fil de Molly Russell et que celles-ci pourraient ne pas être directement comparables.

À la différence des résultats observés pour les comptes indiquant uniquement un intérêt pour la santé mentale, les chercheurs/chercheuses n'ont pas vu d'effet de « spirale » se produire pour les comptes associés à la fois à la liste de termes liés à la santé mentale et à la liste de termes neutres : 0,4 à 1,4 % des fils kenyans incluaient des vidéos en lien avec la liste générale de termes relatifs à la santé mentale, et moins de 1 % de tous les fils basés aux États-Unis classés dans ce groupe. Ceci s'explique vraisemblablement par l'importance qui a été donnée aux centres d'intérêt sans lien avec la santé mentale (avec trois fois plus de termes d'intérêt général que de termes relatifs à la santé mentale) et le fait que les contenus relatifs à des centres d'intérêt neutres sont probablement beaucoup plus courants sur l'ensemble des vidéos pouvant être recommandées à un utilisateur.

De manière quelque peu contradictoire avec ce qui précède, Amnesty International a observé que, bien qu'ils n'aient montré aucun intérêt actif pour des vidéos associées à de quelconques termes relatifs à la santé mentale, trois comptes basés aux États-Unis sur cinq, qui étaient programmés pour ne réviser que des contenus en lien avec les animaux domestiques, les voyages et d'autres vidéos d'intérêt général similaires (termes « neutres ») ont vu 2 à 4 % de leurs fils être associés à des termes et des hashtags relatifs à la santé mentale (liste générale de termes relatifs à la santé mentale), avec des pics allant jusqu'à 12 % (sur 50 vidéos consécutives). Les comptes kenyans appartenant à ce même sous-groupe ont vu 0,6 à 0,7 % de publications liées à la santé mentale. Des analyses supplémentaires seraient nécessaires pour mettre ces résultats en contexte. Le phénomène pourrait cependant être dû, par exemple, à l'un des termes d'intérêt général apparaissant dans la description de certaines de ces vidéos, ou aux taux d'engagement élevés de ces vidéos, qui auraient conduit à leur amplification dans les fils des comptes.

6.5.2 RÉSULTATS DES EXPÉRIENCES MANUELLES : LES EFFETS DE « SPIRALE » PEUVENT SE DÉCLENCHER EN 20 MINUTES

Amnesty International a observé un effet de « spirale » encore plus rapide dans le cadre des expériences manuelles et un taux encore plus élevé de contenus potentiellement préjudiciables qu'avec les comptes automatisés ayant indiqué un intérêt pour la santé mentale. Au cours de l'observation des recommandations proposées à un compte enregistré par un prétendu adolescent de 13 ans aux Philippines, la première vidéo accompagnée du tag « #depressionanxiety » (sic) (#dépressionanxiété) et montrant un garçon en état de détresse a été suggérée à l'utilisateur au bout de 67 secondes de défilement du contenu recommandé sur la page « Pour toi ». À partir de 12 minutes de défilement, 58 % des publications recommandées avaient un rapport avec l'anxiété, la dépression, l'automutilation et/ou le suicide, et entraient dans l'une des catégories susmentionnées de contenus relatifs à la santé mentale pouvant avoir des effets néfastes sur les mineur·e·s et les jeunes souffrant déjà de problèmes de santé mentale.

Dans le cadre de l'expérience manuelle menée aux États-Unis, la quatrième vidéo présentée était taguée « #paintok » (« toksouffrance »). Elle contenait un texte qui disait : « quand tu te rends compte qu'on ne t'a jamais fait passer avant quoi que ce soit d'autre de toute ta vie, mais que tu es juste cette personne qui comble le vide dans les vies des autres jusqu'à ce qu'ils n'aient plus besoin de toi ». À partir de la vingtième vidéo (au bout de moins de trois minutes de connexion), 57 % des vidéos avaient un rapport avec des problèmes de santé mentale, avec au moins neuf publications qui idéalisait, banalisaient ou encourageaient le suicide en une heure uniquement.

Dans le cadre de ces expériences manuelles, c'est le compte kenyan qui a vu la progression la plus lente vers un fil rempli de contenus relatifs à la santé mentale. Néanmoins, une fois ce stade atteint (20 minutes après le début de l'expérience), 72 % des vidéos recommandées au cours des 40 minutes suivantes étaient en lien avec des troubles de la santé mentale, avec au moins cinq références aux pensées suicidaires ou au désir de mort de la personne ayant créé le contenu. La première vidéo faisant référence aux problèmes de santé mentale d'un créateur/d'une créatrice de contenu a été

recommandée trois minutes après le début de l'expérience (25e vidéo recommandée). Absolument aucune des vidéos relatives à la santé mentale n'avait été produite par un-e professionnel-le des soins de santé mentale ou une organisation de santé mentale reconnue.

6.6.2 ETATS DE NOTRE CATÉGORISATION MANUELLE : UN EXAMEN PLUS APPROFONDI DES TYPES DE CONTENUS RELATIFS À LA SANTÉ MENTALE RECOMMANDÉS PAR LA PAGE « POUR TOI » DE TIKTOK

Pour mieux comprendre quels types de vidéos TikTok recommandaient aux comptes, les chercheurs/chercheuses ont procédé à une analyse manuelle d'échantillons de 20 minutes visionnés par un compte basé au Kenya et un aux États-Unis, appartement chacun aux deux sous-groupes de comptes ayant expérimenté un effet de « spirale » important (effet d'entonnoir et intérêt général pour la santé mentale), lors du premier, du sixième et du dixième jour. Les chercheurs/chercheuses ont manuellement codé 540 éléments de contenu. Chaque élément a fait l'objet d'un codage séparé par trois examinateurs/examinatrices et catégorisé en fonction d'une décision à la majorité de leur part.

Sur ces éléments, 21 % contenaient un terme de la liste générale et 27 % étaient classés comme potentiellement dangereux (catégories 1 à 9, tableau 1). Ceci est conforme aux résultats susmentionnés de l'analyse basée sur les termes, les chercheurs/chercheuses ayant uniquement catégorisé des contenus tirés de comptes appartenant à deux sous-groupes, celui à effet d'entonnoir et celui associé à la liste générale de termes relatifs à la santé mentale. Les publications relatives à des expériences vécues abordant les sujets de l'anxiété, de la dépression et de l'automutilation constituent la catégorie de vidéos la plus courante (13 % de l'ensemble du contenu). Amplifiées à cette échelle et ne présentant qu'une perspective très restreinte de ce type d'expérience vécue, elles peuvent risquer d'exacerber des problèmes de santé mentale existants chez les mineur-e-s et les jeunes.²¹⁰ La catégorie suivante est constituée de publications montrant des personnes qui pleurent ou sont en état de détresse émotionnelle (4,3 % de toutes les vidéos) ainsi que de descriptions dramatiques de traumatismes, de souffrances et de suicide (4 %). Quatre pour cent supplémentaires de toutes les vidéos idéalisent ou banalisent la dépression, l'anxiété, l'automutilation et le suicide, ou parlent de projets de mort par suicide, expliquent comment s'automutiler ou encouragent ce type de comportement ou le suicide.

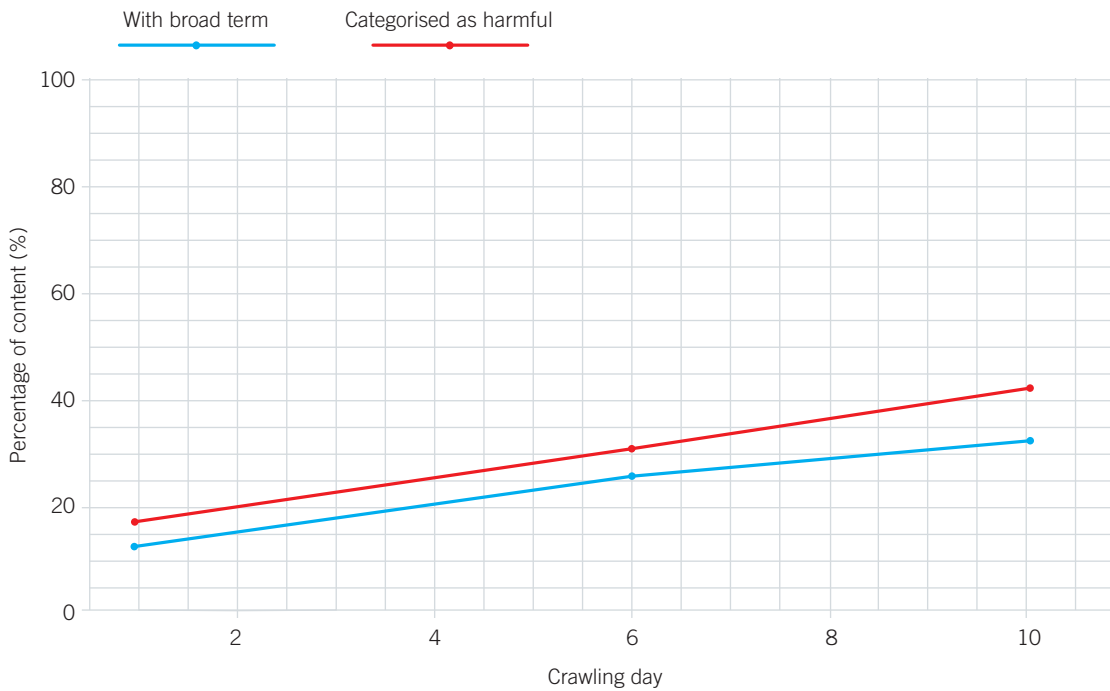
210. Katarzyna Kostyrka-Allchorne, Mariya Stoilova et autres, "Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review", février 2023, *Child and Adolescent Mental Health*, Volume 28, numéro 1, (déjà cité) ; North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 octobre 2022 (déjà cité).

Tableau 1 : Catégorisation manuelle d'échantillons tirés de l'expérience automatisée

Code	Catégorie	Nombre de contenus catégorisés manuellement	Pourcentage de contenus incluant un terme de la liste générale
0	Pas pertinent	393 (73 %)	11,2
1	Publications relatives à des expériences vécues qui parlent d'anxiété, de dépression et d'automutilation (sans idéalisation)	68 (13 %)	44,1
2	Publications montrant des personnes en état de détresse émotionnelle	23 (4,3 %)	39,1
3	Publications décrivant des sentiments de solitude ou d'infériorité comme n'ayant pas de solution et/ou comme étant mérités	8 (1,5 %)	37,5
4	Publications visant à représenter l'automutilation, la dépression et les pensées suicidaires comme inévitables	4 (0,7 %)	75
5	Descriptions dramatiques de traumatismes, de souffrances et de suicide	22 (4 %)	59,1
6	Images traumatisantes	0	-
7	Publications qui valorisent, idéalisent ou banalisent la dépression, l'anxiété, l'automutilation et le suicide (y compris par le biais de dessins, de bandes dessinées, de musique ou de légendes écrites)	15 (3 %)	60
8	Publications qui communiquent de fausses informations au sujet de la santé mentale , par exemple en dissuadant les utilisateurs de chercher l'aide de professionnel-le-s ou de prendre des médicaments prescrits par des médecins, ou en encourageant les personnes à s'autodiagnostiquer sur la base d'informations erronées	2 (0,4 %)	50
9	Publications qui mentionnent l'intention de s'automutiler ou de se suicider , ou qui expliquent comment s'automutiler ou se suicider , ou qui encouragent de quelque autre façon que ce soit l'automutilation ou le suicide	5 (1 %)	60

Le graphique 5 représente l'augmentation constante sur le temps du partage de contenus potentiellement nocifs relatifs à la santé mentale. Une comparaison entre le pourcentage de contenus catégorisés manuellement et le pourcentage de vidéos contenant un terme de la liste générale de termes relatifs à la santé mentale montre que l'analyse quantitative, qui se fonde uniquement sur les termes, sous-estime le volume réel de contenus potentiellement dangereux, qui pourraient présenter des risques pour les mineur-e-s et les jeunes souffrant déjà de problèmes de santé mentale.

Graphique 5 : Progression du contenu catégorisé manuellement le premier, le sixième et le dixième jour de nos expériences. Nous représentons à la fois la progression des contenus classés comme nocifs (selon le cadre que nous avons établi) et celle des contenus dont la description inclut des termes figurant de la liste générale de termes relatifs à la santé mentale.



METTRE EN CONTEXTE L'ANALYSE BASÉE SUR LES TERMES GRÂCE À L'EXERCICE DE CATÉGORISATION MANUELLE

Pour replacer dans leur contexte les résultats de l'analyse principale basée sur les termes et les hashtags, les chercheurs/chercheuses ont voulu savoir dans quelle catégorie entraient les contenus intégrant des termes de la liste générale de termes relatifs à la santé mentale au sein du cadre de catégorisation manuelle. Ceci permet de comprendre simultanément quel volume du contenu amplifié par le système de recommandation est potentiellement nocif et quel volume de contenu potentiellement nocif n'est pas décelé par les comptes automatisés et n'est pas pris en compte dans l'analyse basée sur les termes, parce que les descriptions n'incluent pas de hashtags ou de termes de la liste générale.

Soixante-deux pourcent des contenus intégrant des termes de la liste générale étaient classés comme potentiellement préjudiciables. Dix-huit pourcent des vidéos ne contenant aucun des termes de la liste étaient classés comme potentiellement préjudiciables. Ensemble, ces données semblent indiquer que la liste générale de termes relatifs à la santé mentale permet de prédire de façon relativement fiable quel contenu est potentiellement nocif ou non. Néanmoins, il reste à la fois : (1) un volume important de contenus (38 %) contenant un terme de la liste générale qui ne sont pas catégorisés comme potentiellement nocifs et (2) un volume important de contenus potentiellement nocifs, mais ne contiennent aucun des termes de la liste générale.

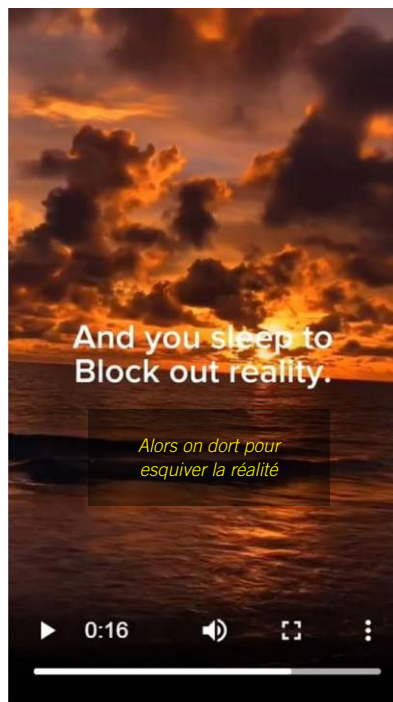
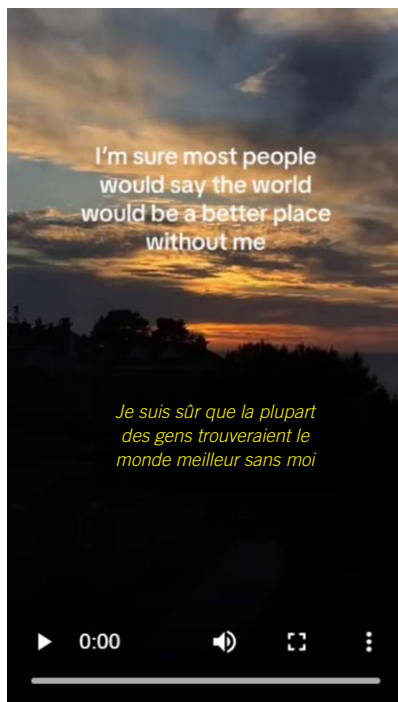
6.7 EXEMPLES DE PUBLICATIONS RECOMMANDÉES



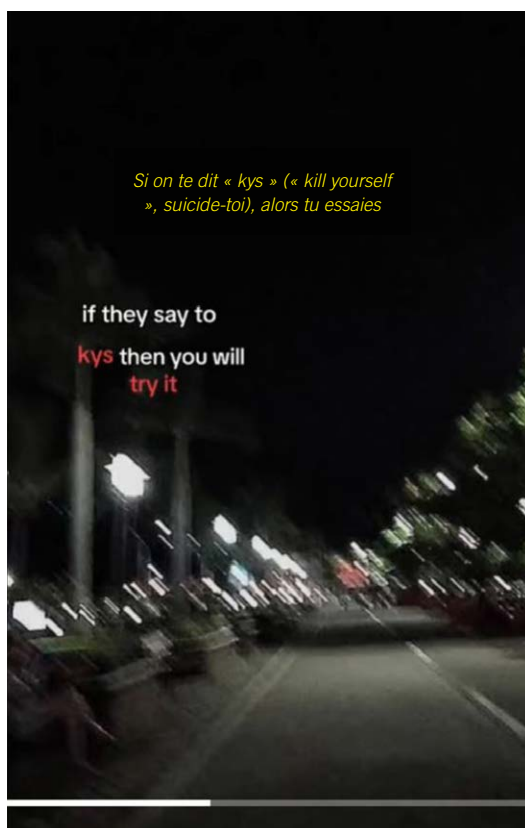
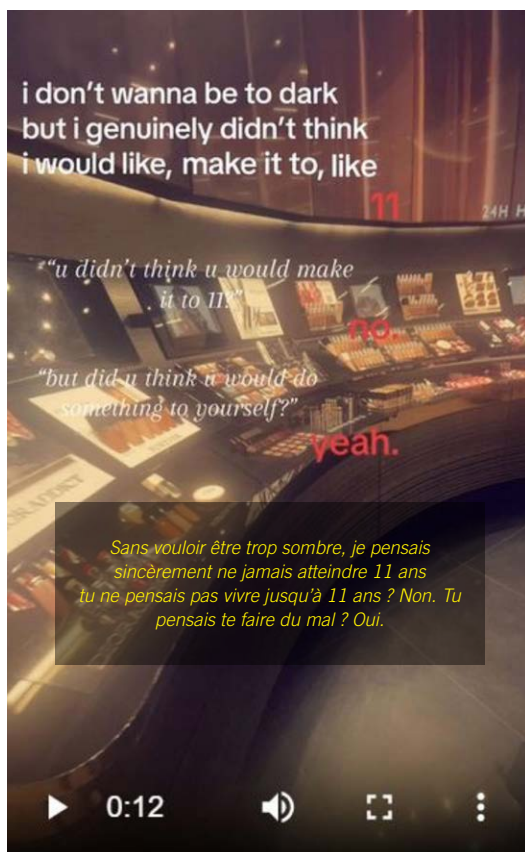
Cette publication décrivant la souffrance d'une jeune fille, accompagnée des tags #broken (brisée) #vent (se défouler) #ventacc (compte pour se défouler) #overthinking (je réfléchis trop) #emotionallydrained (émotionnellement vidée) #ventaccount (compte pour se défouler) #ventmentalhealth (se défouler santé mentale) #yoursafeplace (ton endroit sûr) #sadtiktok (tiktok triste) et d'autres, a été recommandée à trois comptes basés aux États-Unis et un au Kenya.

Une voix off générée par l'intelligence artificielle dit :

« Tu ne comprends pas à quel point elle se sent vidée et pourtant, elle va à l'école tous les jours. Pourtant, elle a toujours un sourire aux lèvres, mais comment sais-tu qu'elle n'est pas en galère. Elle est tellement fatiguée et avec tout ce que vous, les garçons, lui dites, ça ne fait qu'empirer. Elle rentre à la maison en pleurant silencieusement, pour que personne ne l'entende et ne lui demande ce qui ne va pas. Elle ne peut plus continuer comme ça. Tu ne vois pas la souffrance qu'elle subit et tout le mal qu'elle se donne pour rester en vie. Tu aggraves les choses avec toute la merde que tu lui dis, elle n'arrivera pas à faire face encore longtemps. Elle fait vraiment tout ce qu'elle peut. »



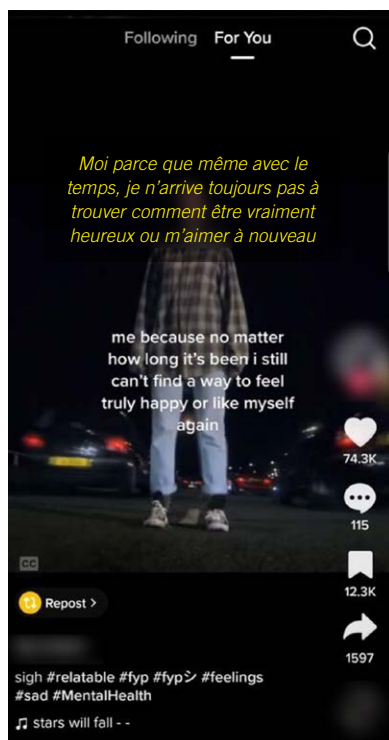
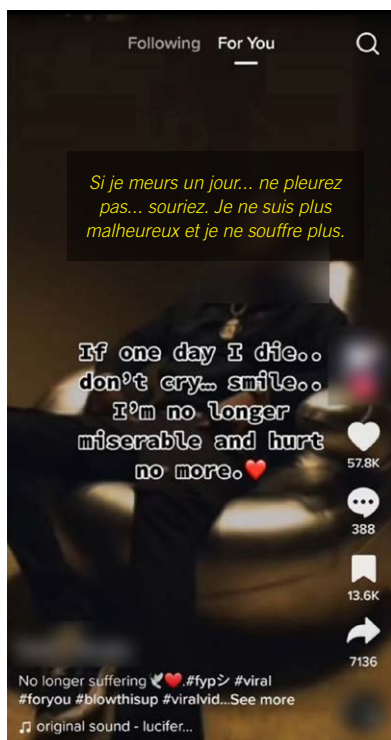
Ces publications à caractère suicidaire et dépressif sur fond d'images romantiques ont été recommandées à un compte basé aux États-Unis classé dans le sous-groupe à comportement dépressif/effet d'entournoir.



Exemples de publications parlant de suicide, notamment par le biais d'un langage codé et d'abréviations probablement utilisées pour éviter d'être détectées par le système de modération de contenus automatisé (« kys » = « kill yourself » = suicide-toi). Un grand nombre de publications relatives à la dépression et au suicide que nous avons suivies avaient pour toile de fond des routes sombres et des jeux d'arcade.

Citation de la vidéo en bas à gauche (taguée « #MentalHealthMatters », « la santé mentale est importante ») :

« Avez-vous déjà regardé une plaquette de comprimés en pensant à tout avaler ? Avez-vous déjà simplement souhaité en finir ? Avez-vous déjà simplement souhaité mourir ? Avez-vous déjà pensé au soulagement que vous ressentiriez, à toute la douleur qui disparaîtrait ? Mais comme je le disais, c'est juste une idée »



Les publications ci-dessus, recommandées à des comptes d'adolescents et adolescentes, expriment des pensées suicidaires d'une façon qui idéalise et/ou encourage l'acte lui-même. La vidéo de droite est une scène de film, sur laquelle la personne qui a créé le contenu a superposé son commentaire. Elle montre une jeune fille en détresse, qui marche face à des voitures roulant sur une autoroute, en criant « Allez ! ». L'extrait est coupé juste avant que la collision attendue se produise. Cette publication a été recommandée au compte géré manuellement aux États-Unis.



Publications parlant d'autotatation en réaction à un sentiment de colère ou de déception face à une situation d'exclusion sociale. La vidéo de droite semble avoir été filmée à Londres. Elle a été recommandée à un compte kenyan entre des vidéos d'un créateur de contenu professionnel de la santé mentale au Royaume-Uni. Les comptes kenyans se sont souvent vu présenter des publications de créateurs et créatrices de contenus basés aux États-Unis et au Royaume-Uni.



Publication banalisant l'automutilation. La vidéo réutilise un son couramment utilisé, celui de deux voix chantant « la la lalalalala », coupé dans la seconde image. La vidéo a été recommandée à un compte basé aux États-Unis (sous-groupe à comportement dépressif/effet d'entonnoir) avec les tags suivants : #ventaccount (compte pour se défouler) #mentalhealthmatters (la santé mentale est importante) #ventacc (compte pour se défouler) #vent (se défouler) #deepvibes (ambiance intense) #sadtok (tok triste) #sadtiktok (tiktok triste) #sadvibes (ambiance triste) #MentalHealth (santé mentale) #nightvibes (ambiance nocturne) #chilltiktok (tok détente) #feelinglonely (je me sens seul-e) #trustissues (problèmes de confiance) #chilltiktok (tiktok détente) #iwanttobealone (je veux être seul-e) #help (au secours). Lorsque nous avons procédé à une vérification quelques semaines après qu'elle a été recommandée au compte, elle n'était plus disponible, ayant été supprimée soit par la personne qui l'avait créée, soit par TikTok.



« Je me suis mis à penser à quel point la vie de tout le monde serait mieux sans moi. »



POUSSÉ·E·S VERS LES TÉNÉBRES

COMMENT LE FIL « POUR TOI » ENCOURAGE L'AUTOMUTILATION ET LES IDÉES SUICIDAIRES

Amnesty International

Ces exemples sont représentatifs des types de publications auxquels des mineur-e-s et des jeunes vont probablement être exposés s'ils indiquent s'intéresser aux contenus relatifs à la santé mentale. Les publications qui idéalisent les pensées suicidaires ou dépressives, ou qui banalisent l'automutilation, de même que les publications et commentaires mentionnés ci-dessus, qui semblent encourager l'automutilation et le suicide, donnent une idée des risques systémiques qui ont été identifiés dans le cadre de cette étude. Un grand nombre de ces publications ont été amplifiées par le système de recommandation de TikTok, au point d'avoir obtenu des centaines de milliers de vues.

Ensemble, ces résultats montrent non seulement que TikTok ne modère pas correctement les éléments individuels de contenu sur sa plateforme, mais que son système de recommandation amplifie activement de tels contenus à grande échelle, au risque d'aggraver les problèmes de santé mentale dont souffrent certains mineur-e-s et jeunes.

6.8 LES LIMITES DU PROJET

L'une des principales limites qui se sont posées à l'enquête technique a été l'exclusion de tous les contenus non anglophones. Bien que de nombreux entretiens menés auprès de mineur-e-s et de jeunes adultes au Kenya et aux Philippines indiquent que l'anglais est la principale langue d'interaction avec la plateforme, de nombreux jeunes publient, cherchent et visionnent aussi des contenus dans d'autres langues parlées localement. Le discours sur la santé mentale, qui est lié à un langage plus médical, semble être majoritairement partagé en anglais dans les deux pays en question,²¹¹ mais les publications des utilisateurs concernant leurs expériences vécues peuvent également être produites et partagées dans d'autres langues. De précédents rapports ont signalé le fait que TikTok et Meta consacrent bien moins de ressources à la modération des contenus non anglophones.²¹² Il est par conséquent possible que les chercheurs/chercheuses aient pu observer des taux plus élevés d'amplification de contenus problématiques et nocifs dans des langues autres que l'anglais.

LES LIMITES DES AUDITS AUTOMATISÉS

Comme nous l'avons déjà noté dans ce rapport, TikTok et d'autres plateformes de réseaux sociaux ont mis en place des obstacles à la recherche indépendante, par exemple en interdisant la collecte automatique de données (« moissonnage ») dans leurs Conditions générales d'utilisation. Au cours de l'enquête technique menée par Amnesty International, beaucoup des comptes kenyans ont été suspendus avant d'avoir atteint la fin de notre période d'observation, peut-être parce qu'un VPN avait été utilisé pour simuler leur situation géographique. Ils ont donc dû être exclus de l'analyse.

En dehors de ces difficultés techniques, le principal problème que pose l'utilisation de comptes automatisés pour mener à bien une expérience à l'échelle sans exposer les chercheurs/chercheuses à la charge que représente le fait d'être en contact avec ce genre de contenus perturbants sur des périodes prolongées, c'est que la capacité des comptes automatisés à imiter les comportements humains est limitée. L'expérience manuelle à petite échelle a souligné ces limites. Alors qu'un compte automatisé ne peut réagir qu'à des aspects prédéfinis des vidéos pertinentes (limités, dans ce cas, aux termes et hashtags figurant dans les descriptions des vidéos), un-e chercheur/chercheuse ou un-e

211. Ceci semble être lié aussi bien à des facteurs socio-économiques que culturels. Par exemple, le tagalog, l'une des langues les plus parlées aux Philippines, ne dispose pas de traduction exacte pour le mot « dépression ». Le mot tagalog qui s'en approche le plus est celui de « tristesse ». Les thérapeutes et leurs patient-e-s souvent plutôt aisés sont plus susceptibles d'utiliser l'anglais dans le cadre de conversations sur la santé mentale. Les organisations de santé mentale non lucratives à qui nous avons parlé dans le cadre de ce projet et qui sont présentes sur TikTok publient également en anglais.

212. Mozilla Foundation, "From Dance App to Political Mercenary: How Disinformation on Tiktok Gaslights Political Tensions in Kenya", 8 juin 2022, <https://foundation.mozilla.org/en/blog/new-research-disinformation-on-tiktok-gaslights-political-tensions-ahead-of-kenyas-2022-elections> ; Amnesty International, *Myanmar. L'atrocité des réseaux sociaux* (déjà cité).

utilisateur/utilisatrice serait capable d'identifier des contenus pertinents par l'association de différents éléments, comme les descriptions, le contenu parlé, les visuels et les textes affichés, la musique, les sons et le nom du profil. Il est donc probable que les comptes automatisés n'ont pas manifesté d'intérêt pour de nombreuses publications pertinentes, notamment parce que l'analyse manuelle et de précédentes explorations manuelles ont montré que certains des contenus les plus extrêmes liés à l'automutilation et au suicide n'incluent pas de descriptions ou de hashtags, le lien avec d'autres contenus pertinents semblant plutôt se faire par le biais d'éléments sonores communs.

7. LES ÉCHECS DES ENTREPRISES

7.1 MANQUE DE DILIGENCE REQUISE

Le cas de Molly Russell, l'adolescente britannique de 14 ans, décédée des suites d'un acte d'automutilation après avoir visionné des contenus à caractère dépressif sur Instagram, montre de la manière la plus tragique qui soit qu'exposer de jeunes utilisateurs et utilisatrices présentant des symptômes dépressifs à un fil constitué d'une quantité importante de publications qui banalisent, voire idéalisent les pensées dépressives, l'automutilation et le suicide peut aggraver leurs problèmes de santé mentale préexistants et encourager ainsi des actes dangereux et même dévastateurs dans la vie réelle.²¹³ Comme il a précédemment été évoqué, les systèmes de recommandation algorithmiques qu'utilisent TikTok et d'autres entreprises de réseaux sociaux pour accroître au maximum l'engagement des utilisateurs et utilisatrices ont le potentiel de faire des inférences concernant les pensées et les émotions des utilisateurs et des utilisatrices. Le système de recommandation répond à ces inférences concernant l'état mental des personnes en leur montrant des contenus plus personnalisés et présentant un attrait individuel qui, selon ses prédictions, vont engager l'utilisateur ou l'utilisatrice, avec pour effet potentiel de perpétuer l'état émotionnel ou mental dont le système de recommandation a fait l'inférence et de piéger les utilisateurs et utilisatrices dans une boucle rétroactive. De cette manière, le système de recommandation peut influencer les pensées d'une personne mineure ou jeune, interférer avec celles-ci, voire les manipuler. Ceci représente une forme de pouvoir bien plus ciblée et personnelle sur les mineur-e-s et les jeunes que celle précédemment exercée par la diffusion de contenus à un large public par les médias traditionnels.

Une telle interférence ou manipulation constitue une atteinte au droit à la liberté de pensée et peut aggraver les problèmes de santé mentale préexistants chez de jeunes utilisateurs et utilisatrices, posant de sérieux risques pour leur santé physique, voire pour leur vie.²¹⁴ Compte tenu de la fragilité émotionnelle des enfants et des jeunes adultes,²¹⁵ qui a fait l'objet de nombreuses études, de cas tels que celui de Molly Russell et des nombreux éléments rapportés par la société civile et les médias, sans oublier les révélations des « Facebook Papers », TikTok doit savoir et s'être rendu compte que son

213. BBC News, "Molly Russell: Coroner's report urges social media changes" (déjà cité).

214. Katarzyna Kostyrka-Allchorne, Mariya Stoilova et autres, "Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and nonsuicidal self-injury conditions – a systematic review", février 2023, *Child and Adolescent Mental Health*, Volume 28, numéro 1, (déjà cité) ; North London Coroner's Service, "Regulation 28 report to prevent future deaths", 13 octobre 2022 (déjà cité).

215. University of Rochester Medical Centre, Health Encyclopaedia, "Understanding the teen brain", 2023, (déjà cité) ; Administrateur de la santé publique des États-Unis, "Social Media and Youth Mental Health" (déjà cité).

système algorithmique de recommandation risque de faire entrer les jeunes utilisateurs/utilisatrices dans une spirale de publications potentiellement dangereuses pouvant aggraver des problèmes de santé mentale préexistants.

Comme mentionné à la section 4.4, pour remplir ses obligations aux termes des Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme (Principes directeurs des Nations unies), TikTok doit mettre en place, comme il se doit, une procédure de diligence raisonnable en matière de droits humains qui lui permette d'identifier les préjudices réels et potentiels qu'elle cause, de les prévenir et d'en atténuer les effets, ainsi que de rendre compte de la manière dont elle y remédie.

Dans la réponse écrite qu'elle a fournie aux questions d'Amnesty International le 12 juillet 2023, TikTok n'a pas révélé quels risques spécifiques, notamment à l'égard des droits des enfants, elle avait identifiés au cours des phases de conception et d'élaboration de l'application, ni la procédure pour identifier les risques qu'elle posait et les stratégies d'atténuation qu'elle avait mises en place avant son lancement. À la suite de celui-ci, TikTok aurait dû et devrait toujours continuer de faire preuve de toute la diligence requise.

Dans sa lettre à Amnesty International, l'entreprise prétend que son approche en matière de sécurité des jeunes s'appuie sur son engagement en faveur des droits humains, et en particulier sur les Principes directeurs des Nations unies « et leur appel à mener à bien une diligence raisonnable en matière de droits humains ». Cet engagement figure sur son site Internet, mais TikTok ne dispose d'aucune politique relative aux droits humains accessible au public, mis à part cet engagement de haut niveau en faveur des droits humains sur son site Internet, ce qui constitue un élément essentiel de toute démarche de diligence raisonnable en matière de droits humains, comme le soulignent les Principes directeurs des Nations unies et le Guide OCDE sur le devoir de diligence.²¹⁶

Amnesty International a poursuivi en demandant par écrit à TikTok si l'entreprise disposait d'une politique en la matière couvrant un vaste éventail de droits humains, notamment les droits au respect de la vie privée, à la liberté de pensée et à la santé, et si cette politique était accessible au public. Les chercheurs et chercheuses ont également demandé qui, parmi les membres de la direction de l'entreprise, est responsable de sa mise en œuvre. TikTok a répondu en renvoyant Amnesty International à l'engagement en faveur des droits humains affiché sur le site Internet de l'entreprise, qui, d'après l'entreprise, est fondé sur plusieurs cadres internationaux des droits humains, notamment les Principes directeurs des Nations unies, que l'entreprise s'est engagée à respecter. Cependant, TikTok n'a pas communiqué de politique en matière de droits humains comme l'a demandé Amnesty International, ni les informations sur les personnes, au sein de la direction, responsables de la mise en œuvre de cette politique ou sur la manière dont elle est intégrée aux plus hauts niveaux de l'entreprise dans l'ensemble des fonctions.²¹⁷

Dans sa réponse écrite du 12 juillet 2023, TikTok a affirmé qu'elle disposait d'une équipe centralisée chargée de son travail en faveur des droits humains, ainsi que d'autres membres du personnel au sein de l'entreprise qui, selon elle, sont « habilités par la direction à traiter tout risque identifié²¹⁸ ». Amnesty International a demandé à TikTok combien de membres composent cette équipe et à qui, au sein de la direction, cette équipe doit rendre des comptes. Amnesty International a également demandé quels moyens étaient donnés à ces membres du personnel afin de leur permettre d'identifier les risques, quelles procédures étaient mises en place lorsque des risques étaient identifiés et quelles mesures les équipes pouvaient prendre. TikTok n'a pas répondu à cette question.²¹⁹

216. OCDE, Guide sur le devoir de diligence pour une conduite responsable des entreprises, <https://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-Business-Conduct.pdf> (consulté le 18 juillet 2023).

217. Voir Annexe 4 : Réponse écrite de TikTok, 29 octobre 2023.

218. Voir Annexe 3 du présent rapport : Réponse écrite de TikTok, 12 juillet 2023.

219. Voir Annexe 4 : Réponse écrite de TikTok, 29 octobre 2023.

Les principes communautaires de TikTok incluent une sous-section portant sur les politiques de TikTok relatives à la sécurité et au bien-être des jeunes, qui sont supervisées par une branche de l'équipe Safety Product Policy. En outre, d'après TikTok, son « équipe Platform Fairness est spécialisée dans les questions relatives aux droits humains, à l'équité et à l'inclusion et applique une démarche multidimensionnelle à la révision de ses politiques, des fonctionnalités de son produit et de ses systèmes algorithmiques ».

Néanmoins, dans ses lettres, l'entreprise n'a donné aucune précision sur ce qu'englobe cette « démarche multidimensionnelle » ni sur les risques spécifiques qu'elle recherche ou a identifiés, notamment en ce qui concerne son système de recommandation. Elle a indiqué que la question des droits humains avait été intégrée à l'ensemble de ses principes communautaires et qu'elle dispose de conseils consultatifs globaux, qui incluent des experts en droits de l'enfant, et qu'elle est membre de WeProtect Global Alliance et de la Tech Coalition,²²⁰ organisations au sein desquelles elle échange avec ses pairs à propos des risques actuels et futurs. En juin 2023, TikTok a également annoncé la création d'un Conseil des jeunes²²¹ qui, selon ce qu'elle affirme, fournira « une occasion plus structurée et plus régulière pour les jeunes de partager leur point de vue » sur la création de politiques, de produits et de programmes.

Ces initiatives présentent, pour TikTok, la possibilité de s'engager davantage et de façon suivie avec les détenteurs et détentrices de droits, les experts, les organisations de la société civile et les autres parties intéressées. Il reste cependant difficile de savoir quelle influence aura cet engagement sur les processus de diligence requise et d'évaluation des risques de TikTok, et à quelle fréquence les parties prenantes seront consultées.

Dans sa réponse à Amnesty, TikTok n'a pas fait mention des risques spécifiques que posent ses pratiques de collecte des données et de profilage et son système algorithmique de recommandation pour les droits au respect de la vie privée et à la liberté de pensée ni du risque posé par les fils de contenus hautement personnalisés pour le droit à la santé. L'entreprise a déclaré, dans sa lettre du 12 juillet, qu'« à l'avenir [...] [elle] procédera à des évaluations périodiques de l'impact, en partenariat avec des parties tierces, qui pourront prendre la forme d'évaluations à l'échelle de l'entreprise et à l'échelle des équipes, ainsi que d'évaluations spécifiques de produits ou de branches d'activité ». Elle n'a fourni aucun détail supplémentaire, notamment concernant les moments où démarreront les évaluations, leur régularité ou l'éventuelle publication des résultats. Ceci va à l'encontre des recommandations des Nations unies, qui appellent à procéder à des « évaluations régulières et complètes des incidences sur les droits de l'homme » en tant qu'« élément clé de [l'exercice de] leur diligence raisonnable en matière de droits de l'homme²²²».

Amnesty International a demandé davantage d'informations à TikTok sur les politiques et pratiques de diligence requise en matière de droits humains de l'entreprise. TikTok a répondu que l'entreprise « consulte plusieurs parties prenantes afin d'informer [sa] diligence requise en matière de droits humains et [est] en train de mettre en œuvre un certain nombre de recommandations pour les opérations en matière de confiance et de sécurité formulées à l'issue de notre travail avec les consultants d'Article One sur les droits humains. Ces recommandations sont mises en œuvre par l'équipe Platform Fairness, en partenariat avec un groupe de travail sur les droits humains composé de collègues d'autres équipes de l'ensemble de l'entreprise. L'évaluation recommandait que TikTok mène une évaluation de l'impact sur les droits des enfants, qui va être lancée en partenariat avec Article One.²²³»

220. WeProtect Global Alliance et la Tech Coalition sont deux initiatives au sein desquelles s'impliquent de grandes plateformes technologiques. Elles ont pour mission de protéger les mineur-e-s contre les abus et l'exploitation sexuels en ligne.

221. Salle de presse de TikTok, « Mise à jour du Mode Connexion Famille et mise en place d'un Conseil des jeunes », 27 juillet 2023, <https://newsroom.tiktok.com/fr-fr/mise-a-jour-du-mode-connexion-famille-et-conseil-des-jeunes>.

222. Rapport du Haut-Commissaire des Nations Unies aux droits de l'homme, « Le droit à la vie privée à l'ère du numérique », 15 septembre 2021, (déjà cité).

223. Voir Annexe 4 : Réponse écrite de TikTok, 29 octobre 2023.

Bien qu'il soit positif que l'entreprise ait mis en place un groupe de travail sur les droits humains composé de membres du personnel d'équipes de l'ensemble de l'entreprise et que TikTok prévoi.e de mener une évaluation de l'impact pour les droits des enfants, le fait que l'entreprise n'ait pas mené d'évaluations de l'impact de ce type jusque-là, compte tenu de l'immense popularité de la plateforme auprès des personnes de moins de 18 ans, est une négligence considérable. Des doutes subsistent quant au fait que TikTok ait été capable d'évaluer correctement les risques que sa plateforme pose pour les enfants en l'absence de telles évaluations de l'impact, ce qui reflète son incapacité à mener correctement les procédures de diligence requise.

En effet, TikTok indique, dans sa lettre du 29 octobre 2023 à Amnesty International, que l'une des autres recommandations est de « créer un processus de diligence requise à l'échelle de l'ensemble de l'entreprise qui comprendra des évaluations périodiques de l'impact en matière de droits humains », et que l'entreprise est en train d'élaborer ces processus, « qui devront proposer des éléments déclencheurs indiquant le moment où nous devons mener une évaluation. » Le fait que TikTok n'ait actuellement pas de procédure de diligence requise à l'échelle de l'ensemble de l'entreprise constitue un grave manquement à sa responsabilité de respecter les droits humains. Ce que l'entreprise entend par des « éléments déclencheurs indiquant le moment où nous devons mener une évaluation » reste également à définir. Compte tenu des graves risques que la plateforme représente pour les droits humains des enfants, du grand nombre d'enfants qui utilisent la plateforme dans le monde et de l'importance qu'elle a prise dans la vie de nombreux enfants, il est indispensable que l'entreprise mène des évaluations de l'impact en matière de droits humains de manière régulière dans le cadre de sa diligence requise, et chaque fois qu'elle met en œuvre de nouvelles fonctionnalités de conception.

Dans sa réponse écrite, TikTok n'a pas inclus de liste des risques spécifiques identifiés. Étant donné que TikTok commence seulement maintenant à élaborer une procédure de diligence requise en matière de droits humains qui, d'après l'entreprise, sera conforme aux normes internationales relatives aux droits humains, notamment aux Principes directeurs des Nations unies, il n'est pas surprenant que l'entreprise n'ait pas transmis de liste des risques spécifiques qu'elle a identifiés ou des mesures qu'elle a mises en place pour les prévenir et les atténuer, puisque, jusque-là, l'entreprise n'a pas honoré sa responsabilité de respecter les droits humains en menant une diligence requise en matière de droits humains adaptée.

Cette absence de procédure de diligence requise adaptée est contraire aux Principes directeurs des Nations unies qui imposent aux entreprises d'avoir « en place des politiques et des procédures par lesquelles elles peuvent à la fois connaître les droits de l'homme et montrer qu'elles les respectent dans la pratique.²²⁴ » Les Principes directeurs des Nations unies précisent aussi que « montrer dit communiquer, en assurant un certain degré de transparence et de responsabilité aux individus ou aux groupes susceptibles d'être touchés²²⁵ ». Un autre élément essentiel de la diligence raisonnable, tel que le souligne le Guide OCDE sur le devoir de diligence, implique de communiquer sur la façon dont les incidences ont été traitées.²²⁶ Dans ses lettres, TikTok a évoqué plusieurs mesures destinées à promouvoir la sécurité, la santé et le bien-être des enfants, mais certaines de ces mesures ne traitent pas les questions abordées dans le présent rapport. De plus, comme nous le montrerons ci-après, un grand nombre des mesures que l'entreprise a mises en œuvre et qui pourraient être destinées à répondre à tout risque identifié d'entraîner des mineur.e-s et des jeunes dans des « spirales » de contenus nocifs relatifs à la santé mentale ne sont pas adéquates.

224. Principes directeurs des Nations unies, (déjà cités), principe 21 et commentaire associé.

225. Principes directeurs des Nations unies, (déjà cités), principe 21 et commentaire associé.

226. OCDE, Guide sur le devoir de diligence pour une conduite responsable des entreprises, (déjà cité).

7.2 DES RÉPONSES INADÉQUATES À DES RISQUES GRAVES

Pour répondre aux questions spécifiques concernant les mesures qu'elle prend pour prévenir les risques pour la santé mentale de jeunes, TikTok, dans sa lettre à Amnesty International, s'est référée à ses principes communautaires, qui définissent quels types de contenus sont interdits et, s'ils font l'objet d'un signalement ou sont décelés de quelque autre manière que ce soit, sont supprimés de la plateforme.²²⁷ Ses principes communautaires interdisent notamment les contenus « montrant, promouvant ou fournissant des instructions sur le suicide et l'automutilation, et les défis, jeux et pactes qui s'y rapportent » ainsi que les contenus « montrant ou promouvant de fausses informations relatives au suicide et à l'automutilation » et « partageant des projets de suicide et d'automutilation²²⁸».

Pourtant, l'analyse manuelle d'échantillons de publications des pages « Pour toi » de comptes ayant indiqué un intérêt pour la santé mentale menée par Amnesty International a révélé qu'environ une vidéo recommandée sur 100 encourageait l'automutilation et que 3 % supplémentaires valorisaient, idéalisaient ou banalisaient la dépression, l'automutilation ou le suicide. Ces vidéos auraient été non seulement accessibles aux utilisateurs et utilisatrices cherchant des contenus de ce type, mais aussi, comme le montrent nos travaux de recherche, activement amplifiées par le système de recommandation de TikTok.

Bien que l'entreprise n'ait jamais admis publiquement les risques pour les droits humains liés à la création d'un fil de contenus hautement personnalisés, elle a toutefois intégré à sa plateforme des outils dont la conception semble avoir été fondée sur la possibilité que de tels risques existent. En 2021, elle a annoncé être en train d'œuvrer pour varier les recommandations de contenus, afin d'éviter de générer des fils aux perspectives trop étroites, déclarant que : « Chez TikTok, nous reconnaissons que voir trop d'une chose, que ce soit des animaux, des conseils de remise en forme ou des parcours personnels de recherche du bien-être, ne correspond pas à l'expérience de découverte diversifiée que nous visons à créer²²⁹».

En juillet 2022, on a appris que TikTok s'apprêtait à déployer un outil ayant pour objectif d'aller dans le sens de cette diversification des contenus.²³⁰ Dans sa lettre de juillet à Amnesty International, TikTok a également écrit qu'elle « interrompt[ait] les schémas de répétition, afin que les contenus qui peuvent ne posent pas de problème s'ils sont visionnés de manière occasionnelle, comme les contenus relatifs aux pratiques extrêmes de fitness ou de régime, ne soient pas présentés en paquets pouvant devenir problématiques²³¹».

L'enquête d'Amnesty International n'a pourtant trouvé aucune preuve que les « spirales » dans lesquelles les comptes étaient entraînés étaient interrompues par des contenus sans rapport à un degré tel qu'ils pourraient en atténuer les risques. Au contraire, les chercheurs/chercheuses ont observé, au fil du temps, des pics répétés dans l'effet de « spirale » (voir par exemple le graphique 3 en section 6.6.1).

En 2022, TikTok a lancé un autre outil visant à permettre aux utilisateurs d'exclure de leurs fils des vidéos taguées avec certains mots ou hashtags. Il vient s'ajouter au bouton « Pas intéressé(e) », qui existait déjà, et grâce auquel les utilisateurs et utilisatrices peuvent indiquer leur souhait de ne pas

227. Voir l'Appendice du présent rapport : Réponse écrite de TikTok, 12 juillet 2023.

228. TikTok, « Principes communautaires », mis à jour en mars 2023, <https://www.tiktok.com/community-guidelines/fr-fr/mental-behavioral-health/>

229. TikTok, "An update on our work to safeguard and diversify recommendations", 16 décembre 2021, <https://newsroom.tiktok.com/en-us/an-update-on-our-work-to-safeguard-and-diversify-recommendations>.

230. The Verge, "TikTok is giving you new ways to control your For You page", 13 juillet 2022, <https://www.theverge.com/2022/7/13/23205795/tiktok-algorithm-hashtag-filters-safety-content-levels>.

231. Voir l'Appendice du présent rapport : Réponse écrite de TikTok, 12 juillet 2023.

232. Mashable, "TikTok doesn't seem to care if I'm 'not interested'", 28 avril 2022, <https://mashable.com/article/tiktok-not-interested>.

233. Entretien en ligne avec Nikki (pseudonyme), 24 ans, Manille, 8 mai 2023.

se voir recommander de vidéos similaires. Des utilisateurs/utilisatrices ont mis en doute l'efficacité du bouton « Pas intéressé(e)²³²», un constat partagé par plusieurs des jeunes ayant participé à nos travaux de recherche. Ils ont déclaré l'avoir utilisé dans l'espoir d'éviter de continuer de voir des contenus qu'ils estimaient avoir un effet négatif sur leur santé mentale, mais ont constaté son effet limité sur les recommandations qui ont suivi. Comme l'a expliqué « Nikki », la militante manillaise de 24 ans : « J'ai laissé tomber. Je ne clique plus dessus [sur divers outils de contrôle utilisateur de la plateforme], parce que même quand je clique, que je signale [des contenus] et que je les masque, ils finissent quand même par réapparaître dans mon fil²³³».

L'enquête d'Amnesty International n'a pas testé l'efficacité du nouvel outil de filtrage de mots-clés, mais l'immense variété de hashtags et de termes utilisés par la « communauté de créateurs » de TikTok pour les contenus relatifs à l'anxiété, la dépression, les différentes formes d'automutilation et le suicide, qui incluent notamment des fautes d'orthographe intentionnelles et des mots codés, en limite probablement l'efficacité.

La lettre de réponse à nos conclusions envoyée par l'entreprise en octobre 2023 évoque également le mode Connexion Famille, qui permet aux parents de personnaliser la limite de temps quotidienne pour leurs enfants et d'utiliser l'outil de filtrage de mots-clés pour empêcher leurs enfants de voir des contenus associés à certains mots spécifiques dans leur fil « Pour toi » et leur fil « Suivi²³⁴». Ces contrôles parentaux peuvent ajouter une couche de protection, particulièrement si les enfants ont le sentiment de ne pas pouvoir prendre eux-mêmes de mesures de protection, même si cette protection reste limitée par les éléments évoqués précédemment concernant le fait de filtrer tous les termes liés à un sujet ou un type de contenu. Les contrôles parentaux créent toutefois aussi de nouvelles complexités et peuvent poser des risques pour la réalisation des droits des enfants dans des situations dans lesquelles l'intérêt supérieur de l'enfant, par exemple concernant l'accès à certaines informations sur un sujet spécifique, ne correspond pas aux croyances ou opinions des parents.

Dans la réponse écrite qu'elle a fournie à Amnesty International,²³⁵ TikTok mettait aussi en avant sa fonctionnalité « rafraîchir », qui permet aux utilisateurs et utilisatrices de réinitialiser leur fil et de réformer le système de recommandation.²³⁶ Une fois de plus, cependant, les résultats de recherche d'Amnesty International semblent suggérer que cette fonctionnalité serait d'une utilité limitée pour une personne mineure ou jeune ayant indiqué un intérêt pour la santé mentale, étant donné que le système de recommandation serait réformé pour amplifier de nouveau les contenus potentiellement nocifs en un bref espace de temps, à moins que ces personnes passent d'elles-mêmes tous les contenus relatifs à la santé mentale. De surcroît, même si elles passaient ces contenus, les résultats de l'enquête technique menée dans le cadre de nos recherches semblent indiquer que les utilisateurs et utilisatrices qui ne montrent aucun intérêt actif pour ce type de contenus reçoivent quand même des recommandations de contenus connexes, bien que dans une proportion moindre que les personnes qui signalent un intérêt actif pour ceux-ci.

Enfin, les chercheurs/chercheuses ont également testé l'utilité de l'outil de transparence « Pourquoi cette vidéo » qui, d'après TikTok, est censé « apporter davantage de contexte au contenu recommandé dans le fil « Pour Toi » des utilisateurs²³⁷». Mais les bulles d'information que les chercheurs/chercheuses ont pu voir s'afficher dans le cadre de l'expérience manuelle incluaient uniquement des explications simplistes et imprécises, telles que « cette vidéo est populaire dans votre pays », ou «

234. Voir Annexe 4 du présent rapport : Réponse écrite de TikTok, 29 octobre 2023, et TikTok, "User safety: What is Family Pairing?", consulté le 30 octobre 2023, <https://support.tiktok.com/en/safety-hc/account-and-user-safety/user-safety#4>

235. Voir Annexe 3 du présent rapport : Réponse écrite de TikTok, 12 juillet 2023.

236. TikTok, « De nouveaux moyens pour ré-actualiser votre fil "Pour Toi" sur TikTok », 16 mars 2023, <https://newsroom.tiktok.com/fr-fr/de-nouveaux-moyens-pour-reactualiser-votre-fil-pour-toi-sur-tiktok>.

237. TikTok, « Découvrez pourquoi une vidéo est recommandée dans le fil « Pour Toi » », 20 décembre 2022, <https://newsroom.tiktok.com/fr-fr/decouvrez-pourquoi-une-vidéo-est-recommandee-dans-le-fil-pour-toi>.

cette vidéo est longue, vous semblez aimer les vidéos longues », qui n'expliquent en rien la raison pour laquelle une vidéo spécifique a été présentée au compte, ou comment l'effet de personnalisation s'est appliqué au compte en question, qui s'est vu plonger dans une « spirale » abrupte.

Il est clair que des outils de contrôle utilisateur plus efficaces sont nécessaires. Cependant, même si ces outils étaient efficaces, ils ne constituent que des mesures de protection isolées, qui font porter la responsabilité de créer un espace de navigation sécurisé aux utilisateurs et utilisatrices, notamment à des personnes jeunes et vulnérables, qui peuvent ne pas être en mesure de prendre d'elles-mêmes leurs distances par rapport à des contenus préjudiciables ou susceptibles de déclencher chez elles des réactions négatives. De ce fait, ces outils ne constituent pas une réponse adaptée aux risques systémiques associés au système de recommandation de TikTok.

Les mesures que prend TikTok ne sont pas à la hauteur de ce qui est attendu pour que l'entreprise remplisse son devoir de diligence raisonnable en matière de droits humains et, comme le montre l'enquête d'Amnesty International, elle ne prévient pas les risques graves que présente sa plateforme pour la santé physique et mentale des jeunes utilisateurs et utilisatrices souffrant déjà de problèmes de santé mentale.

8 CONCLUSION ET RECOMMANDATIONS

CONCLUSION

La page « Pour toi » de TikTok et le système de recommandations par algorithme sur lequel elle se fonde ont contribué à rendre la plateforme omniprésente dans la vie des enfants et des jeunes. Pendant que les spécialistes du marketing s'émerveillent de sa capacité à conserver l'attention des utilisateurs et utilisatrices et que les concurrents de TikTok s'empressent de s'inspirer des caractéristiques qui ont fait son succès, les enfants et les jeunes sont exposés à un système qui fait de leurs failles psychologiques un moyen de maximiser leur « participation ».

En personnalisant les fils de manière fluide et en un temps record, TikTok a créé une plateforme extrêmement addictive, qui comporte de graves risques pour la santé des utilisateurs et utilisatrices. Pour les enfants et les jeunes du monde entier qui souffrent de dépression, d'anxiété ou d'autres problèmes de santé mentale, TikTok est particulièrement dangereux car, comme le montre ce rapport, le fil « Pour toi » peut facilement entraîner celles et ceux qui signalent un intérêt pour la santé mentale dans une « spirale » de contenu dépressif, notamment des vidéos qui idéalisent, banalisent voire encouragent l'automutilation et le suicide.

Bien que les systèmes algorithmiques des plateformes de réseaux sociaux soient surveillés de plus près depuis la publication des « Facebook Papers » et l'enquête des services médico-légaux britanniques sur la mort de Molly Russell, TikTok et d'autres plateformes poursuivent leur quête effrénée de la participation des utilisateurs et utilisatrices, au mépris des dégâts que cela peut engendrer sur ces personnes. Pire encore, ces entreprises placent la diffusion virale des contenus et le nombre d'heures passées sur la plateforme au-dessus de la sécurité de leurs systèmes. Ce faisant, TikTok et les autres géants des réseaux sociaux ont normalisé la surveillance et la manipulation des utilisateurs et utilisatrices dans le monde entier, portant ainsi atteinte à leurs droits au respect de la vie privée et à la liberté de pensée.

Ces atteintes peuvent sembler abstraites et intangibles, mais elles ont de réelles conséquences sur la vie des enfants et des jeunes du monde entier. En faisant d'une surveillance à visée commerciale un moyen de maintenir le regard sur les écrans, parfois au moyen d'un flot de contenu dépressif et de contenu interdit en lien avec la santé mentale, TikTok met en danger le droit de ses jeunes utilisateurs et utilisatrices à la santé mentale et physique.

Ces risques n'ont pas encore été pleinement cernés, en particulier parce que les résultats de recherche et l'attention des entreprises sont injustement centrés sur l'Europe et l'Amérique du Nord. Néanmoins, TikTok continue à croître : chaque jour, des milliers d'enfants et de jeunes s'inscrivent sur une plateforme susceptible de comporter des dangers pour eux.

Il faut que TikTok revienne de toute urgence ses processus de collecte et d'amplification des données et applique une procédure exhaustive de diligence raisonnable en matière de droits humains. Quoi qu'il en soit, l'action isolée d'une seule entreprise ne saurait suffire à contenir les effets d'un modèle économique fondamentalement incompatible avec les droits humains. Par conséquent, il faut que les États réglementent efficacement les activités des « géants du numérique » comme TikTok, conformément au droit international relatif aux droits humains et aux normes connexes, de façon à protéger les droits des enfants et des jeunes.

Cela nécessite que les États, qui se sont concentrés pendant trop longtemps sur le fait d'imposer aux plateformes de supprimer le contenu illégal ou, dans le cas des régimes plus répressifs, le contenu indésirable, modifient de toute urgence leur ligne de mire. En effet, ils doivent s'attaquer aux causes profondes de la prolifération de contenu dangereux en ligne, qui correspond à un modèle économique voué à maximiser la participation aux dépens des droits humains des utilisateurs et utilisatrices. Alors que les débats mondiaux sur les risques de l'intelligence artificielle se poursuivent, il est crucial que tous les États adoptent et appliquent des réglementations qui préviennent et atténuent les dangers liés à certains des systèmes algorithmiques les plus omniprésents, qui dominent déjà nos espaces publics numériques.

RECOMMANDATIONS

RECOMMANDATIONS À L'INTENTION DE TIKTOK

- TikTok et les autres entreprises technologiques dont le fonctionnement dépend d'activités intrusives fondées sur les données qui s'apparentent à une surveillance commerciale à grande échelle doivent se tourner vers un modèle plus respectueux des droits. Pour commencer, ils doivent veiller à ce que leurs politiques et processus de diligence nécessaire en matière de droits humains prennent en compte les risques systémiques et les incidences généralisées sur les droits humains qui sont liés à leur modèle d'activité, en particulier en ce qui concerne le droit au respect de la vie privée, le droit à la liberté d'opinion et de pensée et le droit à la santé. Ils doivent se montrer transparents au sujet des risques qu'ils ont décelés, notamment pour les droits humains, et de la manière dont ces risques ont été gérés.
- Il faut que TikTok cesse de maximiser « la participation des utilisateurs et utilisatrices » aux dépens de la santé et des autres droits humains de ceux-ci, compte tenu des éléments attestant les impacts négatifs de l'utilisation compulsive de la plateforme, en particulier sur la santé et le bien-être des jeunes utilisateurs et utilisatrices. Dans le cadre de son processus de diligence raisonnable en matière de droits humains, TikTok doit déceler, en coopération avec les utilisateurs et utilisatrices, y compris les enfants et les jeunes, ainsi que des spécialistes indépendants, les éléments de conception qui encouragent une utilisation addictive de la plateforme et des comparaisons sociales, et les remplacer par une expérience axée sur la « sécurité par la conception » et l'intérêt supérieur de l'enfant.
- TikTok doit faire preuve de la diligence nécessaire en matière de droits humains de façon volontariste et continue, tout au long du cycle de vie des technologies algorithmiques, avant et après le déploiement et la mise en œuvre de nouveaux systèmes et fonctionnalités de conception, afin que les risques puissent être identifiés dès la phase de développement et que les atteintes aux droits humains et les autres dangers puissent être gérés dès le lancement de ces technologies.
- TikTok doit faire participer des enfants et des jeunes, des universitaires, des spécialistes de la

société civile et d'autres parties prenantes pertinentes à ses processus de diligence requise en matière de droits humains. Les enfants et les jeunes doivent aussi jouer un rôle central dans la mise en œuvre de la « sécurité par la conception » en étant associés au processus d'élaboration des outils et des caractéristiques des plateformes de réseaux sociaux.

- Des évaluations de l'impact sur les droits humains doivent être publiées régulièrement et comprendre des informations détaillées sur les risques et les mesures d'atténuation prises en fonction des pays (en particulier là où les systèmes peuvent avoir un impact accru du fait de conflits politiques ou de situations d'urgence humanitaires), de certaines catégories d'utilisateurs et utilisatrices (les enfants et les jeunes, notamment) et des changements apportés aux produits.
- Pour respecter la vie privée des utilisateurs et utilisatrices et leur donner une réelle capacité de choix et de contrôle, un écosystème de réseaux sociaux sans profilage devrait être la norme et non pas une option parmi d'autres. Les algorithmes utilisés par TikTok et les autres plateformes en ligne pour organiser le contenu ne devraient donc pas être basés par défaut sur le profilage (par exemple, en fonction du temps de visionnage ou de la participation), et doivent reposer sur une démarche volontaire d'inscription plutôt que de désinscription, le consentement pour l'inscription devant être spécifique et donné en connaissance de cause (notamment grâce à l'utilisation d'un langage accessible aux enfants), librement et sans ambiguïté.
- Il faut que TikTok cesse de collecter des données personnelles sensibles et d'opérer des déductions quant aux intérêts, à l'état émotionnel ou au bien-être des utilisateurs et utilisatrices en fonction du temps de visionnage et de la participation en vue de « personnaliser » les recommandations de contenu et de cibler les publicités. Au lieu de se livrer à une surveillance généralisée pour adapter les fils aux intérêts des utilisateurs et utilisatrices, TikTok doit permettre à ceux-ci de faire part de leurs intérêts de manière délibérée (par exemple, il pourrait être demandé aux utilisateurs et utilisatrices de communiquer des intérêts en particulier s'ils/elles souhaitent obtenir des recommandations personnalisées) et seulement sur la base d'un consentement libre, spécifique et donné en connaissance de cause.
- TikTok doit prendre des mesures supplémentaires pour empêcher que les utilisateurs et utilisatrices à risque ne tombent dans des comportements compulsifs et des « spirales » de contenus potentiellement nocifs. Il pourrait s'agir notamment d'instaurer un nombre maximal quotidien de recommandations personnalisées proposées aux enfants et une liste régulièrement mise à jour de termes liés à du contenu potentiellement problématique en matière de santé mentale, termes considérés comme acceptables en tant que mots-clés de recherche mais pas à des fins d'amplification dans le fil « Pour toi ».
- Introduire des mesures de « friction » comme stratégie d'atténuation. Dans le cadre de ses processus de diligence raisonnable, TikTok doit investir dans la recherche afin de définir et d'intégrer des mesures destinées à limiter l'amplification algorithmique rapide et souvent disproportionnée des contenus potentiellement problématiques.
- À titre provisoire, améliorer de toute urgence l'efficacité des mesures visant à diversifier les recommandations de contenu dans le fil « Pour toi », notamment en introduisant des outils efficaces de contrôle pilotés par les utilisateurs et utilisatrices. Ces outils doivent être faciles à trouver et à comprendre et fournir aux utilisateurs et utilisatrices un moyen efficace d'empêcher les futures recommandations de contenu liées à un sujet, un hashtag ou un utilisateur/une utilisatrice en particulier.
- Améliorer nettement la transparence en ce qui concerne l'utilisation des algorithmes d'organisation et de modération du contenu, en veillant à ce que leurs mécanismes soient accessibles au

public et expliqués de manière intégrée, dans des termes clairs, compréhensibles et adaptés aux enfants, dans toutes les langues pertinentes.

- Veiller à la cohérence des décisions prises en matière de modération de contenu, assurer une supervision humaine de la modération automatisée et prévoir des investissements suffisants dans les ressources allouées à la modération, dans toutes les langues.
- TikTok doit créer des campagnes publiques et des messages de sensibilisation sur sa plateforme au sujet des différentes fonctionnalités de sécurité que les utilisateurs et utilisatrices peuvent activer. Ces campagnes pourraient être portées à la connaissance des utilisateurs et utilisatrices par divers canaux, comme des publications de promotion sur les fils et des notifications internes à l'application encourageant à s'informer pour pouvoir utiliser les divers outils de sécurité en toute confiance.
- Permettre à des chercheurs et chercheuses indépendants d'accéder à tous les systèmes algorithmiques pertinents et de les étudier, et, si nécessaire, d'obtenir des données pour réaliser des études indépendantes sur les risques systémiques et les dangers pour les droits humains. Le Règlement sur les services numériques de l'UE a établi un cadre concernant l'accès aux fins de la recherche qui doit servir de modèle à des cadres similaires visant TikTok et les autres grandes plateformes de réseaux sociaux ailleurs dans le monde. Là où la réglementation existante ne fournit pas de directives claires sur les critères d'accès et les processus de validation, TikTok et les autres entreprises de réseaux sociaux doivent s'employer à créer des organes consultatifs indépendants pour veiller à ce qu'une procédure de validation indépendante et efficace soit en place.

RECOMMANDATIONS À L'INTENTION DES ÉTATS

Recommandations aux fins d'une véritable protection des données et d'une réglementation efficace des plateformes

Les États doivent :

- Faire le nécessaire pour que l'accès aux infrastructures et services numériques essentiels (comme TikTok et d'autres plateformes de réseaux sociaux) ainsi que l'utilisation de ceux-ci ne soient pas subordonnés à une surveillance ubiquitaire des utilisateurs et utilisatrices, qu'il s'agisse d'enfants, de jeunes ou d'adultes. Cela nécessitera de promulguer et/ou de faire appliquer des lois exhaustives de protection des données qui soient conformes au droit international relatif aux droits humains et aux normes connexes afin d'interdire les publicités ciblées sur la base de pratiques de suivi intrusives. Ces lois doivent restreindre la quantité et le champ des données personnelles qui peuvent être recueillies, limiter strictement la finalité du traitement de ces données par les entreprises, et garantir la protection des déductions qui sont faites au sujet des personnes à partir du recueil et du traitement de données personnelles. Elles doivent également exiger des entreprises qu'elles fournissent des informations claires à leurs utilisateurs et utilisatrices sur la finalité de la collecte de leurs données personnelles, ce dès le début, et qu'elles ne traitent pas ces données d'une manière incompatible avec cette finalité ou avec leur obligation de respecter les droits humains.
- En premier lieu, empêcher les entreprises de conditionner l'accès à leurs services au fait d'« accepter » la collecte, le traitement ou le partage des données personnelles à des fins de ciblage de contenu, de marketing et de publicité.
- Réglementer les entreprises de réseaux sociaux afin de veiller à ce que les algorithmes utilisés par les plateformes en ligne pour organiser le contenu ne soient pas fondés par défaut sur le

profilage, et que ces plateformes reposent sur une démarche volontaire d'inscription plutôt que de désinscription, le consentement pour l'inscription devant être spécifique et donné librement, en connaissance de cause et sans ambiguïté. La collecte et l'utilisation de données personnelles sensibles obtenues par déduction (par exemple, le fait de baser les recommandations sur le temps de visionnage et les mentions « Like », qui permettent de déduire des informations sensibles) pour personnaliser les publicités et les recommandations de contenu doivent être interdites. Au contraire, les utilisateurs et utilisatrices doivent avoir le contrôle des signaux ou des intérêts déclarés qu'ils souhaitent que la plateforme prenne en compte dans la construction de leur fil. Celles et ceux qui préfèrent un fil basé sur des recommandations personnalisées doivent avoir la possibilité de communiquer leurs intérêts personnels à la plateforme en donnant spécifiquement leur consentement en toute liberté et en connaissance de cause et sur la base d'invites formulées en des termes compréhensibles par les enfants.

- Les processus réglementaires doivent comprendre une véritable consultation des groupes concernés, y compris les enfants et les jeunes, ainsi que des spécialistes indépendants et des organisations de la société civile.
- Veiller à ce que des législateurs nationaux indépendants chargés de la protection des données soient en place, que leur indépendance soit garantie par la loi et qu'ils disposent de ressources, de compétences et de pouvoirs suffisants pour enquêter véritablement sur les violations de la réglementation commises par les entreprises de réseaux sociaux et les sanctionner, conformément au droit et aux normes internationaux relatifs aux droits humains. Ils doivent être en mesure d'assurer un contrôle indépendant et efficace de la conception des plateformes, ainsi que de la conception, du développement et du déploiement des systèmes algorithmiques, et de veiller à ce que les entreprises soient juridiquement responsables de déceler, de prévenir et d'atténuer les atteintes aux droits humains associées à ces systèmes.
- Promulguer ou faire appliquer des cadres réglementaires garantissant l'exercice, dans la pratique, du droit de choisir des solutions respectueuses de la vie privée à la place des modèles économiques fondés sur la surveillance. Il s'agit notamment de garantir l'interopérabilité (capacité de communiquer avec des contacts existants en utilisant une autre plateforme compatible) plutôt que la simple portabilité des données pour que les utilisateurs et utilisatrices puissent passer d'un service à un autre sans que leur vie sociale en pâtisse, ainsi que d'atténuer les effets de réseau.
- Exiger de TikTok et des autres entreprises de réseaux sociaux qu'ils expliquent, en des termes adaptés à l'âge, leurs conditions générales aux enfants, ou aux parents et aux autres personnes responsables dans le cas de très jeunes enfants. Ces éléments doivent être formulés dans un langage clair et simple, fournir des informations transparentes aux utilisateurs et utilisatrices tout au long du processus (pas uniquement au début) et contenir des renseignements clairs sur les choix offerts en matière de contrôle et les paramètres par défaut. Les entreprises de réseaux sociaux doivent également être tenues de faciliter la compréhension de leurs conditions générales au moyen d'éléments non textuels, comme des images, des vidéos et des animations, et être facilement joignables pour répondre à d'éventuelles questions.
- Faire de l'intérêt supérieur de l'enfant le principal élément à prendre en compte dans la réglementation de la publicité et du marketing adressé et accessible aux enfants. Le parrainage, le placement de produit et toutes les autres formes de contenu à visée commerciale doivent être clairement distingués de tous les autres contenus et ne doivent pas perpétuer des stéréotypes discriminatoires (comme ceux fondés sur le genre, la race, l'âge ou le handicap).

Recommandations relatives à la diligence raisonnable en matière de droits humains

Les États doivent :

- Exiger juridiquement des entreprises technologiques qu'elles exercent une diligence raisonnable de façon volontariste et constante pour cerner les incidences et les risques de leurs activités mondiales en ce qui concerne les droits humains, y compris ceux liés à leurs algorithmes ou découlant de leur modèle économique dans sa globalité, et pour y remédier. Lorsque des entreprises ciblent des enfants ou ont pour utilisateurs et utilisatrices des enfants, elles doivent être tenues d'intégrer les droits des enfants dans leurs processus de diligence nécessaire, en particulier de réaliser et de rendre publiques des évaluations de l'impact sur les droits des enfants, en accordant une attention particulière aux effets spécifiques et parfois graves de l'environnement numérique sur les enfants. Les États doivent prendre des mesures adéquates pour prévenir et surveiller les atteintes aux droits des enfants commises par des entreprises, enquêter sur ces atteintes et les sanctionner le cas échéant.

Recommandations relatives aux recours effectifs

Les États doivent :

- Investir dans la mise en œuvre de programmes d'éducation efficaces sur le numérique, l'encourager et la promouvoir pour veiller à ce que les personnes comprennent leurs droits, notamment le droit à un recours effectif en cas d'atteintes aux droits humains concernant la protection des données, le respect de la vie privée ou autre, dans le cadre de l'utilisation de services numériques.
- Garantir l'accès à un recours utile en cas d'atteintes aux droits humains liées aux incidences des entreprises technologiques, quel que soit le lieu où se produisent ces atteintes, notamment si elles résultent des activités de leurs filiales (à l'étranger ou sur le territoire national). Des mécanismes de réparation facilement accessibles et compréhensibles doivent être en place pour que les personnes concernées puissent déposer des plaintes quand leurs droits ont été bafoués.

ANNEXE

1. CONFIGURATION TECHNIQUE DE L'ÉTUDE

Aux fins de cette enquête, nous avons utilisé une version relativement récente de l'application mobile TikTok pour Android. Nos comptes étaient gérés de manière automatisée et nous interceptons les communications entre l'application et les serveurs de TikTok pour collecter des données client. Nous avons ensuite envoyé ces données aux serveurs de Junkipedia, le système d'audit élaboré par notre partenaire technique, l'Algorithmic Transparency Institute, où nous avons enrichi notre collection de données par des métadonnées supplémentaires, dont des transcriptions écrites de la bande-son de vidéos TikTok, des données de participation et des données sur les médias.

Nous avons créé les comptes américains directement sur des appareils Android se trouvant aux États-Unis. Pour les comptes kenyans, nous avons créé des comptes dans un pays différent, que nous avons ensuite utilisés au moyen d'un réseau virtuel privé (VPN) kenyan. Après la création des comptes et le lancement des expériences, neuf des 20 comptes kenyans ont été désactivés par TikTok avant la fin de l'expérimentation. Nous les avons donc exclus de notre analyse.

Le processus de création de compte a parfois nécessité la résolution d'un CAPTCHA et, occasionnellement, la vérification de l'adresse électronique utilisée lors de la création du compte. Nous avons eu pour pratique de vérifier les adresses électroniques de chaque compte créé. Toutes les dates de naissance des utilisateurs ont été fixées aléatoirement en mai 2010. Ainsi, TikTok a rendu tous les comptes privés. Occasionnellement, TikTok nous envoyait un message pour nous signaler que notre compte avait été classé comme privé et, dans de nombreux cas, il y avait aussi un pop-up intitulé « Privacy Highlights for Teens » (principales informations de confidentialité pour les adolescent-e-s), assorti d'un ensemble de vidéos sur la collecte et l'utilisation des données par la plateforme.

2. LISTS OF TERMS USED IN THE AUTOMATED EXPERIMENT

Broad Mental Health Interest:

alwayshereforyou❤️❤️
anxiety
bipolar
blade
blowup
bpd
broken
canileave
cantfightanylonger
childhoodtrama
cry
cryforhelp

crying
darkvibes
deppresed
depressiøn
depression
depressed
depressedtiktok
depressing
depressingslideshow
depression
depressionanxiety
depressionawareness

depressions
dontknowwhattodo
drained
drainedmentally
emotionallydrained
endpoint
harm
heartbroken
hurting
icanttakethis anymore
idk anymore
ihatemylife

imgoinginsane	readytoenditall	tombofthemaskgameplay
imnotok	relatablevent	tombofthemaskvent
insecurities	sad	tompofthemask
ithurts	sadhub	totm
ithurtssobad	sadness	trauma
itnevergetsbetter	sadquotes	twvent
itsalwaysmyfault	sadrelatabletiktok	ursafeplace
iwannacry	sadslideshow	vent
jarofinsecurities	sadsound	vent12677
justsadtbh	sadstory	ventacc
justventing	sadstuff	ventaccount
lifehurts	sadtiktok	ventart
lifeisunfair	sadtimes	ventartist
lifesucks	sadtok	ventdrawing
løsergrp	sadvibes	ventedit
mentalhealthawareness	safeplace	venting
mental	safeplaceforeveryone	ventingaccount
mentalbreakdown	safespace	ventingfeelings
mentalhealth	scars	ventingpost
mentalhealthawareness	schizoslideshows	ventingthoughts
mentalhealthawarness	schooldraining	ventmentalhealth
mentalhealthmatters	sh	ventout
mentalhealthtiktoks	sharpener	ventpost
mentalillness	shawareness	ventpost
mentallydrained	shizopost	ventposting
mentallyill	shizoposting	ventpoststocrypto
mentallyunstable	shscars	vents
newsreporter voice	shvent	venttiktok
newsreporter voicesad	socialanxiety	venttok
nobodycares	so tired	venttome
notok	stressedout	venttomeifneeded♥
notokay	struggling	venttttt
overthinking	suffering	ventvideo
pain	suicide	ventviral
painhub	tired	ventwithme
paintok	tiredoflife	webothinalotofpain
panicattack	tombofmask	whendoesitend
ptsd	tombofthemask	yoursafeplace

Narrow Mental Health Interest:

blade	scars	ventingaccount
canileave	schizoslideshows	ventingfeelings
cantfightanylonger	schooldraining	ventingpost
crying	sh	ventingthoughts
depressingslideshow	sharpener	ventmentalhealth
endpoint	shizopost	ventout
icanttakethis anymore	shizoposting	ventpost
idkanymore	shvent	ventpost
ihatemylife	so tired	ventposting
imgoinginsane	suffering	ventpoststocrypto
insecurities	suicide	vents
itnevergetsbetter	tombofthemaskvent	venttiktok
itsalwaysmyfault	twvent	venttok
jarofinsecurities	vent	venttome
justventing	vent12677	venttomeifneeded♥
løsergrp	ventacc	ventttttt
mentallillness	ventaccount	ventvideo
mentallyill	ventart	ventviral
mentallyunstable	ventartist	ventwithme
readytoenditall	ventdrawing	webothinalotofpain
relatablevent	ventedit	
sadslideshow	venting	

General Interest:

adorable	art	babyfever
aesthetic	ascendingarts	babygoat
affirmations	australia	babygoats
agriculture	australianshepherd	babygoatsoftiktok
alaska	awesome_earthpics	babylove
alberta	babiesoftiktok	babytok
americanview	babyanimals	backpacking
animal	babychicks	backyardfarm
animallover	babycow	backyardflock
animalsoftiktok	babycowsoftiktok	backyardgarden
arizona	babydog	backyardpoultry

beach	chickencoop	cutecow
beautifuldestinations	chickenkeeping	cutecows
beauty	chickenmath	cutedog
beginnergardener	chickenmom	cutedogs
bestfriend	chickensoftiktok	cutepet
bird	chicks	cutepets
birds	chicktok	cutepuppy
birdsoftiktok	chihuahua	dachshund
birdtok	chihuahuastiktok	dachshundpuppy
bonsai	chocolate	dachshundsoftiktok
bordercollie	christmas	dairy
bottlebaby	college	dancewithturbotax
boymom	colorado	deer
brain	comedy	doggo
braintraining	consciousness	doglife
bull	conservation	doglove
bunnies	construction	doglover
bunniesoftiktok	containergarden	doglovers
bunny	containergardening	dogmom
bunnylove	cooking	dogmomlife
bushcraft	cottage	dogoftheday
cactus	cottagecoreaesthetic	dogsofinstagram
calf	cottagegarden	dogsoftiktok
california	countryboy	dogtraining
calvesoftiktok	countrylife	dressage
calvingseason	countryliving	duck
camping	countrymusic	ducks
canada	countrymusicug	egg
catlife	countryside	eggs
catlover	cow	energyhealing
catlovers	cowboy	england
catmom	cowgirl	englishcountryside
cattle	cowtok	enlightenment
cattledog	crazychickenlady	entrepreneur
cattok	crystals	equestrianlife
catvideo	cuteanimals	fairycore
chasingwaterfalls	cutebaby	family
chicken	cutecat	fantasticearth

faranimals	funnydogs	growfood
farmer	funnydogvideos	growth
farmerlife	funnypets	growyourown
farmgirl	funnyvideo	growyourownfood
farmher	furbaby	handmade
farmhouse	gardendesign	harvest
farming	gardener	hawaii
farmkid	gardenersoftiktok	health
farmlifeisthebestlife	gardening101	healthylifestyle
fish	gardeningforbeginners	healthyrecipes
fishing	gardeningtiktok	hen
fitness	gardeningtips	hens
florida	gardeninspiration	herbalism
flower	gardenlife	herbs
flowerfarm	gardenproject	highlysensitiveperson
flowerfarmer	gardentips	hike
flowergarden	germanshepherdpuppy	hikersoftiktok
fluffy	germanshepherdssoftiktok	hikertok
fluffycow	girlswhohike	hiketok
fluffycows	glacier	hikingszn
fluffycowsoftiktok	glaciernationalpark	hikingtiktok
foodie	goat	hilarious
foodies	goats	history
foodlover	goatsoftiktok	hobbyfarm
foodtiktok	goatok	homedecor
foodtok	godisgood	homegrown
football	goldenretrieverlife	homesteader
ford	goldenretrieverpuppy	homesteadinglife
forest	goodmorning	homesteadlife
foryoupageofficiall	goodthing	horror
freerangechickens	gooutside	horsegirl
frenchbulldog	granola	horselover
frenchie	granolagirl	horseriding
fresheggs	granolatok	horseshow
fruit	grasspuppies	horsesontiktok
funnyanimals	grasspuppy	horsetok
funnycat	greenhouse	horsetraining
funnydog	greenscreenvideo	housecow

houseplants	minifarm	northwestisbest
houseplantsoftiktok	minifluffycows	ny
hungry	minihighland	ocean
hunting	minimoo	oklahoma
husky	minipig	oregon
innerchild	minnesota	organic
innerchildhealing	mom	organicgardening
instagood	momlife	outdoorlife
ireland	momtok	outdoorsy
italy	monkey	outdoorsygirls
johndeere	montana	pacificnorthwest
joke	montanagram	parenting
jokes	montanalife	parents
kentucky	montanamoment	parentsoftiktok
kids	moo	partswork
kidsoftiktok	motherhood	peaceful
kitten	motocross	peoplepleaser
kittens	mountain	permaculture
kittensoftiktok	mountains	petchicken
kitty	mountainsarecalling	petchickens
labrador	music	petcow
lake	nationalpark	petlover
landscape	nationalparkgeek	pig
landscape_lovers	nationalparks	piggy
lawofattraction	naturallandscape	piglet
lifecoach	naturally	piglets
lifehack	naturelove	pigs
lightworker	naturelover	pigsoftiktok
livestock	naturelovers	pitbullsoftiktok
loveyourself	naturephotography	plant
makeup	natures_hub	plantas
meow	naturetiktok	plantlover
mexico	naturetok	plantmom
microcow	naturevibes	plantsoftiktok
miniaturedachshund	newborn	planttiktok
minicows	newengland	planttok
minicowsoftiktok	newyork	pnwonderland
minidachshund	nigeriandwarfgoats	pony

pots	seniordog	universe
poultry	sheep	upstateny
pregnant	showcattle	urbangarden
puppies	silkiechicken	utah
puppiesoftiktok	silkies	vacation
puppydog	silly	vanlife
puppylife	slowliving	vegetablegarden
puppytiktok	smallbusiness	veggiegarden
puppytok	smallfarm	visitengland
pygmygoats	snow	visitmontana
quakerpregrain	snowstorm	walk
rabbits	springgarden	wanderlust
rabbitsoftiktok	springtime	washington
rain	springvibes	washingtonstate
ranch	stallion	water
ranching	storytime	waterfall
ranchlife	summer	waterfallhike
reactivedog	sunrise	waterfalls
reasonforbooking	sunset	weather
relateable	switzerland	westcoast
relationshipadvice	teachersoftiktok	wild
rescue	tennessee	wildanimals
rescueanimals	texas	wildflowers
rescuedog	thruhike	wildlife
river	tiktoktravel	wildlifephotography
roadtrip	toddlersoftiktok	winter
rodeo	tomato	wisdom
rooster	tractor	womenwhohike
roostersoftiktok	travelbucketlist	wood
rural	traveltiktok	woods
sausagedog	traveltok	woodworking
scotland	tree	work
scottishhighland	trees	workingdog
seeds	treework	wyoming
seedstarting	truck	yoga
selfsufficient	uk	yummy

3. TIKTOK'S WRITTEN RESPONSE OF 12 JULY 2023

TikTok response page 1



July 12th, 2023

Dear Ms. Abdul Rahim,

Thank you once again for the outreach and for continuing the conversation started earlier this spring with the team at Amnesty Tech. We appreciate the time you've taken to develop this questionnaire related to children's rights in the digital environment and have collected and consolidated responses to the items highlighted in your letter. To that end, we would like to share more about TikTok's efforts across each of the areas highlighted in the questionnaire:

DUE DILIGENCE

Our approach to youth safety is informed by our larger [commitment](#) to human rights, including the United Nations Guiding Principles on Business and Human Rights and its call to conduct human rights due diligence. Our teams, which are composed of youth safety policy experts, proactively assess human rights risks related to young people. We also work to embed a human rights-based approach across all our [Community Guidelines](#). We have [Global Advisory Councils](#), which include experts in children's rights and we are members of the [WeProtect Global Alliance](#) and the [Tech Coalition](#) where we engage with our peers on current and upcoming risks. Furthermore, our [recently announced](#) Youth Council will serve as an additional sounding board to create policies, products, and programs in tune with the needs of our users. We also work with industry experts, non-governmental organizations, and industry associations around the world in our commitment to building a safe platform for our community. We collaborate with organizations in different regions to share best practices, create programs, and exchange ideas on safety-related topics. We regularly consult with external stakeholders and partners in this area and will continue to update you on our work in future.

We believe that human rights are integral to the work of all teams at TikTok. At TikTok, thousands of people are focused on helping to make our platform safe for our community to explore entertaining content and share their creativity. We have a centralized team that manages our efforts across the company, and we have champions across the organization, empowered by leadership to address any identified risks. TikTok's policies on youth safety and well-being, as articulated in this [sub-section in the Community Guidelines](#), are developed by our Trust & Safety Product Policy team, and are overseen by a sub-team that specializes in youth safety and well-being issues. This sub-team closely collaborates with the Product Policy team responsible for account-level enforcement. The Platform Fairness team specializes in issues of human rights, fairness, and inclusion and applies a multifaceted approach to the review and revision of policies, product features, and algorithmic systems. In the future, we will conduct periodic impact assessments in partnership with third parties, which may take the form of company-wide assessments, team assessments, and product or business line specific assessments. In our human rights risk assessment process, gender is one of many considerations that has been identified to

have an impact on the risk level, and we design our mitigation measures to address such considerations.

TikTok embeds safety-by-design principles, which means that we implement a safety-first approach in the design of the platform and that user safety is embedded as a priority throughout the product and feature development decision-making processes. TikTok's goal is to provide young people with an experience that is developmentally appropriate and helps to ensure a safe space for self-exploration. TikTok is only for those aged at least 13, or 14 in certain jurisdictions. In the US, we offer a curated, view-only experience for those under age 13 that includes additional safeguards and privacy protections. We work to design tools and policies that promote a safe and age-appropriate experience for teens 13-17. In relation to product design, we take several steps including: (1) limiting access to [certain product features](#), (2) utilizing [Content Levels](#) that sort content by levels of thematic comfort, (3) using restrictive [default privacy settings](#), and (4) making content uploaded by accounts registered to users under 16 ineligible for the For You Feed (FYF).

Our policies prohibit content that may put young people at risk of exploitation, or psychological, physical, or developmental harm. We have many policies to promote youth safety on the platform. Below are the risks we have identified, and what is appropriate/inappropriate for our teen users.

NOT allowed

- Sexual exploitation of young people, including child sexual abuse material (CSAM), grooming, solicitation, and pedophilia
- Physical abuse, neglect, endangerment, and psychological abuse of young people
- Trafficking of young people, promotion or facilitation of underage marriage, and recruitment of child soldiers
- Sexual activity of young people
- Nudity or significant body exposure of young people
- Allusions to sexual activity by young people
- Seductive performances by young people
- Consumption of alcohol, tobacco products, and drugs by young people

Age-restricted (18 years and older)

- Cosmetic surgery that does not include risk warnings, including before-and-after images, videos of surgical procedures, and messages discussing elective cosmetic surgery
- Activities that are likely to be imitated and may lead to any physical harm
- Significant body exposure of adults
- Seductive performances by adults
- Sexualized posing by adults
- Allusions to sexual activity by adults
- Blood of humans and animals
- Consumption of excessive amounts of alcohol by adults
- Consumption of tobacco products by adults

For You Feed ineligible

- Any content created by an under-16 account
- Moderate body exposure of young people
- Intimate kissing or sexualized posing by young people

If we become aware of youth exploitation on our platform, we will ban the account, as well as any other accounts belonging to the person, and make reports to law enforcement/NCMEC as necessary.

We also offer several tools and controls to support our community's safety and well-being, such as the guides within our [Safety Center](#), which focus on our approach to safety, privacy and security on TikTok. To specifically support our younger community on TikTok, we have developed a [Youth Portal](#), which offers both in-app tools and educational content for our younger users to enjoy their best possible experience.

DATA COLLECTION

Our [privacy policies](#) set out, among other things, what data we collected as well as how we use or share such data from TikTok users. Please also refer to [this article](#) in our help center for information about our recommendation system. TikTok does not use "sensitive personal data," as the term is defined under GDPR, to personalize content. Nor does TikTok use data collected from users and machine learning to draw inferences about protected characteristics beyond gender and age-range. TikTok does not sell user's personal information or share user's personal information with third parties for purposes of cross-context behavioral advertising where restricted by applicable law.

At TikTok, we take special care when crafting the experiences teens have on the platform, including the ads they see. Currently, some teens may be shown ads based on their activities on and off TikTok, such as the accounts they follow, the videos they like, and their profile information. On June 28, 2023, we [announced](#) that we are restricting the types of data that can be used to show ads to teens by region. This means that people in the United States aged 13 to 15 will no longer see personalized ads on TikTok based on their activities off TikTok and people in the European Economic Area, United Kingdom, and Switzerland aged 13 to 17 will no longer see personalized ads on TikTok based on their activities on or off TikTok. We're continuing to work toward providing all people on TikTok with transparency and controls so they can choose the experience that's right for them.

TikTok's [Anti-Discrimination Ad Policy](#) prohibits advertisers from using our ads products to discriminate against people unlawfully. Accordingly, advertisers may not include any unlawfully discriminatory or harassing content in their advertising or any content that encourages unlawful discrimination or harassment. In addition, advertisers may not use audience selection tools to: (a) wrongfully target specific groups of people for advertising in a way that breaches applicable laws or regulations; or (b) wrongfully exclude specific groups of people from seeing their ads, in breach of applicable laws or regulations.

In order to prioritize our users and ensure a positive experience on our platform, we do not allow advertisers to unlawfully target or exclude users based on the following categories, including without limitation by using Custom Audiences or the tools and audiences made available on our platform:

Categories advertisers may not use for discriminatory ad targeting where unlawful:

- Legally protected classes based on the local laws of the region, such as race, ethnicity, age, familial status, and sexual orientation
- National identity, country of citizenship, country of origin, or veteran status or identity or beliefs in regards to political groups, religion or union affiliations
- Personal, financial, or legal hardships
- Individual health statuses or disabilities, including mental, physical, genetic, or emotional health and conditions

When advertisers use TikTok's advertiser tools (e.g., to display advertisements on their own websites and applications), TikTok prohibits those advertisers from transmitting certain types of sensitive information back to TikTok. Section 2.8 of TikTok's Business Products (Data) Terms prohibits advertisers from sharing with, or enabling TikTok to collect, "Business Products Data that you know or ought reasonably to know is from or about children or that includes health or financial information, or other categories of sensitive information (including any information defined as sensitive or special category data under applicable laws, regulations and applicable industry guidelines. TikTok restricts advertisers' use of lookalike audiences based on its [Anti-Discrimination Ad Policy](#) (referenced above) and we have [policies](#) that restrict specific ad categories to 18+.

AGE GATE AND MINIMUM AGE APPEALS

TikTok has a 12+ rating in the App Store, which lets parents use device-level controls to block people under the age of 12 from downloading the app. To help keep people from using TikTok if they're not yet old enough to do so, we've designed a neutral, industry-standard age gate that requires people to fill in their complete birthdate to discourage people from simply clicking a pre-populated minimum age. For accounts banned or restricted because we believe the account holder is under a particular minimum age, the account holder can appeal. See [here](#) and [here](#) for more information. TikTok continues to investigate industry standards and best practices when considering other options for users to submit age information.

ALGORITHMIC RECOMMENDER SYSTEMS, HELP FEATURES AND ACCESS TO MENTAL HEALTH-RELATED CONTENT

As a platform used by millions of people in the US and more than 1 billion people around the world, we're committed to protecting our community every day. TikTok removes content that violates our [Community Guidelines](#) and offers features that help people explore TikTok safely, including:

- Redirecting searches linked to terms like #eatingdisorders or #suicide to prompt people to view support resources, such as helplines along with information on how they can seek assistance.
- Enabling people to [refresh](#) their feed if they feel what they are seeing is no longer relevant to them
- Using keywords to tailor their feeds to avoid potentially [triggering content](#)

TikTok cares deeply about the well-being of our community members and wants to be a source of happiness, enrichment, and belonging. We welcome people coming together to find connections, participate in shared experiences, and feel part of a broader community. We work to make sure this occurs in a supportive space that does not negatively impact people's physical or psychological health. To accomplish this, we work with our internal experts, and external partners, such as Digital Wellness Lab, Crisis Text Line, Butterfly Foundation, and the International Association for Suicide Prevention, to shape our approach to mental health content. Research shows that content related to mental health impacts different people in different ways. Nonetheless, TikTok puts limits on certain types of mental health content that can appear on the platform

We want TikTok to be a place where people can discuss emotionally complex topics in a supportive way without increasing the risk of harm and we also want to ensure that TikTok encourages self-esteem and does not promote negative social comparisons. Our Community Guidelines have a range of policies that are designed to ensure that emotionally complex topics can be discussed in a supportive way without increasing the risk of harm. For example, we do not allow showing, promoting or sharing plans for suicide, nor do we allow showing or promoting of disordered eating or any dangerous weight loss behaviors. We also take steps to age restrict content related to certain topics, like cosmetic surgery, that may promote negative social comparison in younger users. We also interrupt repetitive patterns, so that content that may be fine if seen occasionally, like extreme fitness or dieting content, is not being viewed in clusters that may be more problematic.

TikTok uses a combination of machine and human moderation to identify suitable candidates for effective signposting of resources and evaluates these search terms in accordance with TikTok's policy frameworks. We also actively solicit feedback from external organizations, such as the International Association for Suicide Prevention, Samaritans, Comenzar de Nuevo, and others to identify emerging terms that could benefit from resource signposting.

As part of TikTok's commitment to safety, education, and uplifting our community and partners, we launched a Mental Health Media Education Fund and donated over \$2 million in ad credits to organizations working on supporting mental well-being, including:

- **Alliance for Eating Disorders** ([@allianceford](#)) - National Alliance for Eating Disorders is a nonprofit organization providing education, referrals, and support
- **American Foundation for Suicide Prevention** ([@afspnational](#)) - American Foundation for Suicide Prevention, Saving lives + bringing hope

- **Crisis Text Line** ([@crisistextline](#)) - Crisis Text Line provides free, 24/7 mental health support. Text TIKTOK to 741741
- **Made of Millions** ([@madeofmillions](#)) - Made of Millions is a global advocacy nonprofit on a mission to change how the world perceives mental health
- **National Alliance on Mental Illness** ([@nami](#)) - National Alliance on Mental Illness helps Americans affected by mental illness
- **National Eating Disorders Association** ([@neda](#)) - NEDA supports those affected by eating disorders, and serves as a catalyst for prevention, cures and access to quality care
- **Peer Health Exchange** ([@peerhealthexchange](#)) - Peer Health Exchange provides youth with support, resources, and education to make healthy decisions

As part of this initiative, we're also hosting a series of TikTok training sessions to equip our partners with the tools they need to share information with their communities during critical moments, such as World Mental Health Day in October or back-to-school season. This collaboration represents just one part of our continued efforts to advocate for positive mental health and reach people in need of support, and we're grateful that nonprofits and advocacy groups choose TikTok as a platform to share their knowledge and to reach a wide audience.

Encouraging Supportive Conversations

To accompany our Media Education Fund, we also launched a [#MentalHealthAwareness hub](#) for our community to easily learn about well-being topics, connect with advocates, and support organizations that provide important resources.

#MentalHealthAwareness Creator Spotlight

TikTok is a vibrant and welcoming place, enabling creators and the wider community to share their personal stories. Whether they're advocating for more open discussion about depression and anxiety or sharing tips on how people can manage body or self-esteem issues, creators in the #MentalHealthAwareness community help foster open, honest, and authentic conversations. During Mental Health Awareness Month, we spotlighted 10 creators who use TikTok to educate the community on #MentalHealthAwareness and have made a significant impact both on and off the platform over the past year.

In closing, we want to reiterate TikTok's commitment to protecting all members of our community, especially our younger users. Our continued dialogue with Amnesty International is critical as TikTok works to build trust and improve our overall approach to providing young people with an experience that is developmentally appropriate and helps to ensure a safe space for self-exploration

We thank you for your questions and appreciate the opportunity to provide additional details as needed.

With warm regards,
TikTok Trust & Safety

4. TIKTOK'S WRITTEN RESPONSE OF 29 OCTOBER 2023

TikTok response page 1



October 29, 2023

Lauren Dean Armistead, Head of the Children's Digital Rights Team (Interim)
Rasha Abdul Rahim, Director of Amnesty Tech
Michael Kleinman, Director of Silicon Valley Initiative
Amnesty International
1 Easton Street
London, WC1X, ODW
United Kingdom

Dear Ms. Armistead, Ms. Abdul Rahim and Mr. Kleinman,

Thank you for letter dated October 12, 2023, in which you invite TikTok to respond to Amnesty International's research reports regarding TikTok's corporate responsibility to respect human rights in relation to children and young people's use of the TikTok platform. We appreciate the opportunity to address this important topic and reaffirm our deep commitment to protecting the human rights, safety and well-being of people under the age of 18 on the platform. To that end, we would like to share more about TikTok's efforts across the themes highlighted in your report findings:

Privacy and Advertising Policies

At TikTok, the privacy and security of our users is among our highest priorities. We take our responsibility to safeguard people's privacy and data security seriously. In line with industry practices, we collect information that users choose to provide to us and share with the broader TikTok community, as well as information that helps the app function, operate securely, and improve the user experience. We detail the information we collect in [our privacy policies](#). There are a number of inaccuracies about our practices described in the reports that we would like to clarify.

Report 1's assertions centered on TikTok's data collection practices do not accurately describe TikTok's privacy practices, nor the platform's capabilities. The TikTok app has its own built-in search engine functionality and does not directly collect what people search for outside of the app or through other search engines. TikTok does not collect precise geolocation in the US, European Economic Area, UK and many other regions. In regions where we do collect precise geolocation, we obtain consent prior to collection and people can revoke this consent at any time.

In addition to the inaccuracies described above, the assertions about LGBTQ+ content made by the *The Wall Street Journal* and referenced in the report are incorrect, as we said at the time. TikTok does not identify individuals or infer sensitive information such as sexual orientation or race based on what they watch. Additionally, as we explained to the *The Wall Street Journal*, watching a video is not necessarily a sign of someone's

identity. There are many reasons someone may engage with content; there are allies who engage with LGBTQ+ content but may not identify as LGBTQ+ themselves. There are people who enjoy baking content but aren't bakers. There are people who watch sports content and aren't athletes. People come to TikTok to discover new, entertaining content.

When we build products and features for our platform, we do so by keeping privacy in mind and building in privacy principles throughout the product development lifecycle. We also believe it's important to ensure strong protections to help keep minors and young people safe, which is why we've introduced privacy features and tools to support age-appropriate experiences on our platform. In our previous letter and above, we provided an overview of the global default protections, tools and guidance we have implemented to protect teen privacy and keep them safer on the platform. We will continue this important work.

Report 1 also inaccurately describes our ads practices. First, TikTok prohibits advertisers from using our ads products to discriminate against people unlawfully. We provided our [Anti-Discrimination Ads Policy](#) in our prior letter, which describes our stance on discriminatory ads. The report inaccurately characterizes our enforcement of this policy. All advertisements on TikTok are subject to our Community Guidelines and Advertising Policies. As a result, advertisements are not permitted if they violate TikTok's policies. Second, we implemented restrictions regarding the types of data that can be used to show ads to teens by region in an announcement in [July of 2023](#). This policy has been implemented and we will continue to move toward providing our community with transparency and controls so they can choose the experience that's right for them.

Human Rights

TikTok is committed to respecting the human rights of all people, especially community members between the ages of 13-17. Our commitment to human rights, available on our [website](#), is informed by several international human rights frameworks which we have pledged to uphold. These include the International Bill of Human Rights (which includes the Universal Declaration of Human Rights [UDHR], the International Covenant on Civil and Political Rights [ICCPR], and the International Covenant on Economic, Social and Cultural Rights [ICESCR]), (2) the International Labour Organization [ILO] Declaration on Fundamental Principles and Rights at Work, (3) the Convention on the Rights of the Child [CRC], and (4) the United Nations [Guiding Principles on Business and Human Rights](#) [UNGPs].

TikTok consults with a range of stakeholders to inform our human rights due diligence. For instance, we are implementing a number of recommendations to our trust and safety operations that have resulted from our engagement with [Article One Advisors](#) on human rights. These recommendations are implemented by our platform fairness team in partnership with a human rights working group of colleagues on teams across the company. The assessment recommended that TikTok to conduct a child rights impact

assessment, which we will be launching in partnership with Article One. Another recommendation was to develop a company-wide human rights due diligence process which will include conducting periodic human rights impact assessments. In partnership with [Business for Social Responsibility](#) (BSR), we are developing this human rights due diligence toolkit which will propose the triggers around when we need to conduct an assessment. This toolkit and corresponding processes are aligned with international human rights standards, most notably the UNGPs.

Finally, as your report correctly indicates, we have embedded a human rights approach across our Community Guidelines and have advisory councils around the globe, which include experts in children's rights. We are members of the WeProtect Global Alliance and the Tech Coalition. TikTok recently [announced](#) the formation of our Youth Council, which will enable us to listen to the experiences of those who directly use our platform and be better positioned to make changes to create the safest possible experience for our community. We've been working to build the council with youth representing a diversity of backgrounds and geographies and will keep Amnesty Tech informed as work progresses.

Teen Safety and Mental Health

TikTok is committed to ensuring the safety and well-being of our teenage community members. We strive to navigate the complexity of supporting our community's well-being on our platform with nuance. We take a four-pronged approach that involves removing harmful content, age-restricting or dispersing content that may not be suitable for younger members of the community, and empowering people by providing them with tools and connecting them to resources. Our [Community Guidelines](#) do not allow content that shows, promotes, or shares plans for suicide or self-harm or content that shows or promotes disordered eating or any dangerous weight loss behavior. We also age-restrict certain topics that may pose unique risks to children, such as videos showing or promoting cosmetic surgery that do not include risk warnings, including before-and-after images, videos of surgical procedures, and messages discussing elective cosmetic surgery.

As the report references, TikTok has developed and implemented systems that limit content related to certain topics that may be fine if seen occasionally, but potentially problematic when presented in aggregate. These systems include coverage for topics like misery, hopelessness, sadness, and diet and fitness. We continue to work on expanding and implementing these systems, including adding more mental health topics.

We regularly consult with health experts, remove content that violates our policies, and provide access to supportive resources for anyone in need, including children. We are mindful that triggering content is unique to each individual and remain focused on fostering a safe and comfortable space for everyone, including people who choose to share their recovery journeys or educate others on these important topics. We have

published a [guide](#) for creators with suggestions on how to talk about mental health while keeping themselves safe and being respectful to other community members.

TikTok also offers tools to help parents and younger members of our community manage their screen time. As referenced in your report findings, TikTok automatically sets a 60-minute screen time limit for every account belonging to a user below age 18. Before adopting this feature and choosing this limit, we consulted academic research and experts from the [Digital Wellness Lab](#) at Boston Children's Hospital. As your report mentions, teens are able to continue watching after the 60-minute limit is reached by entering a passcode. We have found that enacting more restrictive screen time controls may increase the risk of teens lying about their age, while the passcode feature requires teens to make an active decision to extend their screen time. The interruption introduces friction into the experience, which gives people an opportunity to pause and reflect on whether they wish to continue watching. Additionally, our Family Pairing features allow parents to customize the daily screen time limit for their teens and implement stricter standards if they feel it is needed. We are monitoring the efficacy of the time limit default and are continuing to innovate to make it more effective.

In addition to Family Pairing, TikTok offers a wealth of resources for parents and guardians to help safeguard their teen's safety, privacy, and well-being on the platform, which can be found directly in [TikTok's Safety Center](#). Additionally, TikTok's Youth Portal offers both in-app tools and educational content that empower young users to keep their account secure and limit their online footprint to the degree they feel comfortable. The [Youth Portal](#) includes a "You're in Control" video series which provides safety and security tips for all users.

TikTok is committed to upholding human rights and maintaining a safe platform for all members of our community, especially our younger users. We appreciate the opportunity to respond to these reports and welcome a continued dialogue on these important issues.

Sincerely,



Lisa Hayes
Head of Safety Public Policy & Senior Counsel, Americas, TikTok



**AMNESTY INTERNATIONAL
IS A GLOBAL MOVEMENT
FOR HUMAN RIGHTS.
WHEN INJUSTICE HAPPENS
TO ONE PERSON, IT
MATTERS TO US ALL.**

CONTACT US



info@amnesty.org



+44 (0)20 7413 5500

JOIN THE CONVERSATION



www.facebook.com/AmnestyGlobal



[@amnesty](https://twitter.com/amnesty)

POUSSÉ·E·S VERS LES TÉNÈBRES

COMMENT LE FIL « POUR TOI » ENCOURAGE L'AUTOMUTILATION ET LES IDÉES SUICIDAIRES

Lors de la pandémie de COVID-19, TikTok est devenu une plateforme mondiale, attirant des centaines de millions d'enfants et de jeunes, en grande partie grâce à sa page « Pour toi », un fil déroulable à l'infini contenant des suggestions personnalisées de vidéos, et au système de recommandation qui le sous-tend.

Par cette personnalisation fluide poussée à l'extrême, TikTok a créé une plateforme addictive, en dépit des éléments de plus en plus nombreux qui attestent les graves risques pour la santé qui découlent d'une utilisation compulsive des réseaux sociaux par les enfants. En examinant de plus près les risques du ciblage de contenu sur TikTok, les recherches d'Amnesty International montrent que la page « Pour toi » peut facilement faire tomber les enfants et les jeunes qui signalent un intérêt pour la santé mentale dans des « spirales » de contenus potentiellement dangereux, notamment des vidéos qui idéalisent et encouragent les pensées dépressives, l'automutilation et le suicide. Les risques liés à TikTok renforcent les difficultés des enfants et des jeunes aux prises avec la dépression, l'anxiété et l'automutilation en mettant leur santé mentale et physique en danger.

Il faut que TikTok revienne de toute urgence ses processus de collecte et d'amplification des données et applique une procédure exhaustive de diligence raisonnable en matière de droits humains. Quoi qu'il en soit, l'action isolée d'une seule entreprise ne saurait suffire à contenir les effets d'un modèle économique fondamentalement incompatible avec les droits humains. Il faut que les États réglementent les « géants de la haute technologie » conformément au droit international relatif aux droits humains et aux normes connexes, qui imposent de protéger et de concrétiser les droits des enfants et des jeunes.