OCT. 2023

L'EVOLUTION DES CONDITIONS D'UTILISATION DES RESEAUX SOCIAUX ET LEUR IMPACT SUR LES DROITS DE L'HOMME

Etude exploratoire 2022-2023

Programme de recherche 'Gouvernance et régulation des réseaux sociaux'

UBO Université de Bretagne Occidentale
IUF institut universitaire de France
Lab-LEX Laboratoire de recherche en droit UR 7480
GEODE

L'évolution des conditions d'utilisation des réseaux sociaux et leur impact sur les droits de l'homme

Etude exploratoire 2022-2023

Réalisée dans le cadre du programme de recherche

'Gouvernance et régulation des réseaux sociaux'

Octobre 2023

Université de Bretagne Occidentale

institut universitaire de France

Lab-LEX Laboratoire de recherche en droit UR 7480

GEODE

TABLE DES MATIERES

CONTRIBUTEURS

Valère NDIOR

Valère Ndior est professeur de droit public à l'Université de Brest (laboratoire Lab-LEX) et membre junior de l'Institut universitaire de France. Il est également membre de GEODE - Géopolitique de la Datasphère, de l'OBVIA - Observatoire sur les impacts sociétaux de l'IA et du numérique et Affiliated Fellow au Yale Information Society Project. Il conduit, au sein du Lab-LEX, le programme de recherche *Gouvernance et régulation des réseaux sociaux* (GRS). Il a été membre du groupe de travail 'Legal Frameworks' au Global Internet Forum to Counter terrorism en 2022 et a participé aux travaux du Christchurch Call Advisory Network en 2022-2023.

Martin ARCHIMBAUD

Martin Archimbaud est doctorant contractuel en droit public à l'Université de Brest (laboratoire Lab-LEX) et membre associé au laboratoire GEODE – Géopolitique de la Datasphère. Il est titulaire d'une allocation de recherche de la région Bretagne (projet IntBlock). Sa thèse de doctorat porte sur les blocages et coupures d'internet à l'initiative des gouvernements qu'il étudie sous l'angle du droit international public. Il a été chargé de recherche pour le programme Gouvernance et régulation des réseaux sociaux en 2023.

Elisabeth BOULVARD-CHOLLET

Elisabeth Boulvard-Chollet est doctorante à l'Université de Brest (laboratoire Lab-LEX). En 2023, elle a réalisé un stage au Lab-LEX durant lequel elle a mené, pour le programme GRS, une recherche sur le traitement, par les conditions d'utilisation des réseaux sociaux, des violences sexuelles et basées sur le genre.

LE LAB-LEX

Le Lab-LEX est un laboratoire de recherche en droit, commun à l'Université de Bretagne Occidentale (Brest) et à l'Université Bretagne Sud. Le Lab-LEX a pour principal objectif la recherche fondamentale et appliquée en droit privé, sciences criminelles et en droit public. L'équipe est forte d'une cinquantaine d'enseignants-chercheurs et d'une quarantaine de doctorants. Elle est répartie entre Brest, Quimper, Vannes et Lorient.

REMERCIEMENTS

Les auteurs adressent leurs remerciements à l'équipe d'Open Terms Archive dont le travail a servi de base à la réalisation de cette étude. Ils remercient également la direction du laboratoire Lab-LEX, l'Institut universitaire de France, le centre GEODE – Géopolitique de la Datasphère, l'IREDIES – Institut de recherche en droit international et européen de la Sorbonne et le Yale Information Society Project. Les auteurs ont vu leurs réflexions enrichies par les commentaires des participants à un séminaire de recherche organisé le 15 septembre 2023 à l'IREDIES, particulièrement Hippolyte Bernard, Liane Huttner, Thomas Leclerc, Anne-Thida Norodom, Michel Séjean, Pauline Trouillard.

La réalisation de la présente étude a été permise grâce à des financements du laboratoire Lab-LEX, de l'Institut universitaire de France et du centre GEODE. L'équipe d'Open Terms Archive s'est limitée à la fourniture d'éclairages techniques sur le fonctionnement de sa base de données.

Aucune de ces entités n'a émis d'instruction ou de directive à l'égard des porteurs de projet.

PRESENTATION DU RAPPORT

Le présent rapport a été rédigé par des chercheurs en droit de l'université de Brest (laboratoire Lab-LEX) et met en lumière les modifications de conditions d'utilisation des réseaux sociaux ayant un impact sur les droits de l'homme. L'analyse a été menée sous un angle juridique à partir des données du programme Open Terms Archive, une base de données qui « *enregistre publiquement chaque version des conditions d'utilisation des services en ligne pour en permettre le contrôle démocratique* ».

> « [Le pouvoir des plateformes numériques] est façonné par les règles énoncées dans des documents complexes et en constante évolution qui définissent le fonctionnement de ces plateformes : conditions d'utilisation, politiques de confidentialité, règles de la communauté… Ces conditions offrent souvent des droits et des opportunités inégales selon les juridictions et constituent de plus en plus des normes conçues de manière unilatérale, avec peu voire pas de contrôle démocratique. (…) Open Terms Archive enregistre publiquement ces termes dans différentes langues et pays plusieurs fois par jour, en améliorant leur lisibilité et en mettant en évidence leurs changements ».
>
> Extrait du modèle d'impact OTA, https://opentermsarchive.org/fr/impact

L'étude, expérimentale et exploratoire, concerne une période s'étalant de janvier 2023 à juin 2023. Elle concerne une sélection de sites ayant connu de nombreuses modifications de gouvernance et controverses au cours des années écoulées et rassemblant un volume important d'utilisateurs : **YouTube, Facebook, Instagram, TikTok, Twitter** (désormais X[1]).

Le choix de se limiter à un **échantillon de plateformes** répond à la nécessaire adéquation entre les moyens que peut mobiliser le programme de recherche à ce stade et le volume important de données disponibles. L'ambition du projet pourra être revue à la hausse à l'issue de cette phase exploratoire, concernant les plateformes étudiées, en cas de reconduite du travail de veille.

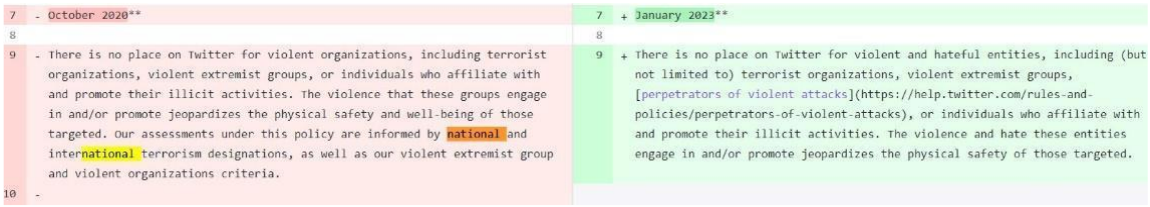
En outre, seules les modifications ayant un **impact sur les droits de l'homme** ont été étudiées. Sont visés particulièrement les droits de l'homme dans les volets civils, politiques, sociaux et culturels, via un suivi des standards relatifs aux données personnelles, au discours de haine, à la désinformation ou aux contenus terroristes et extrémistes. Cette étude exploratoire vise à déterminer l'opportunité de poursuivre l'évaluation annuelle des modifications de conditions générales d'utilisation.

METHODOLOGIE

Après avoir été initiés à l'utilisation de la base de données OTA, les chercheurs ont de façon hebdomadaire consulté les documents enregistrés dans la collection « PGA » et la collection « Contrib » d'Open Terms Archive de janvier 2023 jusqu'à la fin juin 2023[2], pour les **plateformes YouTube, Facebook, Instagram, TikTok, Twitter**. Les modifications observées ont été répertoriées dans un dossier collaboratif, comportant des relevés hebdomadaires datés, incluant le cas échéant des captures d'écran issues de la base de données OTA. Dix relevés ont ainsi servi de documents de travail entre **le 30 janvier 2023 et le 8 juin 2023**, étant entendu que les chercheurs n'ont pas créé de relevé les semaines durant lesquelles aucune modification n'a été constatée.

[1] Twitter a été renommée X fin juillet 2023, soit après la fin de la présente étude. La dénomination Twitter est conservée dans le rapport compte tenu de la période durant laquelle l'observation a été menée.

[2] https://github.com/OpenTermsArchive/pga-versions ; https://github.com/OpenTermsArchive/contrib-versions/



Capture d'écran issue de la base de données OTA

Le travail a, dans la majorité des cas, été mené à partir des conditions d'utilisation extraites des documents en anglais couvrant la juridiction européenne, afin de ne pas occulter des modifications récentes n'ayant pas été traduites en français. Généralement, sauf mention contraire, c'est la version anglaise de ces standards qui fait foi à l'échelle globale du point de vue des entreprises. Meta indique par exemple sur son site que « *la version anglaise (États-Unis) des Standards de la communauté constitue l'ensemble de politiques le plus à jour et qu'elle doit être utilisée comme document principal* ». La version française est toutefois citée dans l'étude lorsqu'elle est disponible et correspond au contenu anglophone.

À l'issue de la période d'observation, les chercheurs ont étudié leurs relevés et mené des discussions sur le risque que les modifications opérées puissent affecter les droits des utilisateurs, à la lumière des instruments internationaux pertinents : **Pacte international relatif aux droits civils et politiques, Pacte international relatif aux droits économiques, sociaux et culturels, Convention de sauvegarde des droits de l'homme et des libertés fondamentales, Charte des droits fondamentaux de l'UE, etc**. Les chercheurs ont volontairement écarté un certain nombre de modifications ne présentant pas un caractère substantiel, tels le remplacement d'un terme par un autre considéré comme équivalent ou des modifications de lien hypertexte.

Le recours aux bases de données d'OTA s'est avéré décisif pour distinguer plus aisément les différences entre anciennes versions et nouvelles versions des CGU. Dans certains cas, c'est uniquement par le biais d'OTA qu'ont été identifiées des modifications qui n'étaient pas explicitement signalées par les réseaux sociaux.

Les conclusions dégagées ont été présentées à des chercheurs et praticiens extérieurs au projet, lors d'un séminaire fermé qui s'est tenu le 15 septembre 2023 à Paris, pour que ceux-ci puissent présenter leurs observations.

Le changelog de Meta

Meta dispose d'un « changelog » qui permet de comparer, en suivi des modifications, différentes versions de ses standards de la communauté et de remonter, dans la plupart des cas, jusqu'au 25 mai 2018 (date d'entrée en application du Règlement général sur la protection des données personnelles). Si la mise à disposition de cet outil de transparence est appréciable, il présente l'inconvénient de comprendre des textes alternant entre le français et l'anglais, ce qui rend ardue la comparaison rigoureuse des textes. Cette configuration suscite deux difficultés. La première porte sur le degré de correspondance entre les termes juridiques utilisés dans les versions anglaise et française (lorsque cette dernière existe). La deuxième concerne la question de l'accessibilité linguistique des standards du point de vue des utilisateurs qui ne seraient pas en mesure de comprendre les règles applicables à leurs activités.

CONTEXTE

Les réseaux sociaux sont devenus incontournables : plus de trois milliards d'utilisateurs emploient quotidiennement les services offerts par Meta, Twitter ou TikTok. Toutefois, les publications qu'ils effectuent à flux constant engendrent de nombreuses dérives que seuls les mécanismes de modération permettent de contenir par la suppression de contenus, le marquage de ceux-ci

(étiquettes, labels, avertissements) ou l'éviction d'utilisateurs. Les propos discriminatoires, appels à la haine ou à la violence, dénigrements, diffamations, atteintes à la vie privée ou opérations de désinformation font partie des comportements nocifs que les plateformes doivent réguler (lorsqu'elles n'y ont pas contribué). L'enjeu pour les entreprises administrant les réseaux sociaux est à la fois de protéger leur modèle économique en conservant l'essentiel de leur audience sur les plateformes et d'éviter de s'exposer aux sanctions que pourraient leur infliger les autorités en cas de défaillance.

Les réseaux sociaux produisent donc des conditions d'utilisation, standards de la communauté, principes, « valeurs » qu'ils opposent à leurs utilisateurs à l'échelle globale. Ces référentiels sont régulièrement amendés par les dirigeants des réseaux sociaux en réponse aux controverses ou aux réformes législatives les visant. Notons que les entreprises de réseaux sociaux comportent parfois des départements dédiés aux droits de l'homme, à la lutte contre les discriminations, à la sécurité ou à l'intégrité des contenus pour définir les modalités d'utilisation des plateformes selon des standards définis en interne.

Outre le fait que les conditions d'utilisation de Facebook, YouTube, TikTok ou Twitter soient structurées autour d'un certain nombre de « valeurs » et de « principes », ces conditions ont la particularité de mobiliser des concepts et des notions qui évoquent le droit international des droits de l'homme. Elles sont parfois modelées de façon à ressembler à des textes « fondateurs », voire à des instruments d'apparence constitutionnelle. Or, elles demeurent de nature privée.

Les comportements des utilisateurs sont ainsi encadrés par une variété de règles dont la complexité et les modalités de mise en œuvre les rendent parfois difficiles à jauger. En cas de violation de ces règles (susceptibles d'être plus strictes que les règles d'origine étatique), des sanctions graduées peuvent être infligées aux utilisateurs allant du simple avertissement à la mise au ban de la communauté. Au-delà de leur impact sur les droits et libertés des utilisateurs, ces sanctions peuvent avoir des conséquences pécuniaires pour ceux d'entre eux qui tirent des revenus des activités menées en ligne.

Le Digital Services Act

Le Règlement sur les services numériques (ou Digital services Act – DSA) a été adopté à l'échelle de l'Union européenne le 19 octobre 2022. Il impose aux plateformes en ligne (réseaux sociaux, moteurs de recherche…) des obligations en matière de modération des contenus illicites, audits algorithmiques, transparence des procédures ou de protection des utilisateurs mineurs. Le règlement exige également des plateformes qu'elles protègent les droits fondamentaux des utilisateurs :

Articler 1er, 1) : « *Le présent règlement a pour objectif de contribuer au bon fonctionnement du marché intérieur des services intermédiaires en établissant des règles harmonisées pour un environnement en ligne sûr, prévisible et fiable qui facilite l'innovation et dans lequel les droits fondamentaux consacrés par la Charte, y compris le principe de protection des consommateurs, sont efficacement protégés* ».

L'article 14 est consacré spécifiquement aux conditions générales d'utilisation et aux exigences auxquelles doivent se conformer les plateformes, notamment en matière de modération de contenus, de notifications de modifications importantes ou d'intelligibilité.

Le règlement est entré en application le 25 août 2023 pour la catégorie des « Très grandes plateformes en ligne » (des plateformes et moteurs de recherche comptant au moins 45 millions d'utilisateurs actifs par mois, dont font partie les services étudiés dans ce rapport). En cas de violation de leurs obligations par les plateformes, la Commission européenne peut imposer des amendes pouvant aller jusqu'à 6 % du chiffre d'affaires mondial des entreprises visées.

La présente étude tente d'aborder des problématiques fréquemment débattues concernant les réseaux sociaux :

Défaut de transparence de la part des plateformes

Déterminer si des communiqués sont publiés par les entreprises lorsque des modifications substantielles des conditions d'utilisation sont constatées. Sont considérées comme « substantielles » des modifications qui altèrent la nature des droits et obligations mis à la charge des utilisateurs ou qui affectent la capacité des autorités publiques à évaluer leur adéquation.

> Ex. : suppression de pans significatifs des instruments pertinents sans notification aux utilisateurs ; modifications de conditions d'utilisation sans modification de la date de dernière actualisation de la page concernée ; variation des conditions d'utilisation en fonction de la zone géographique dans laquelle se trouve l'utilisateur alors que les règles sont présentées comme d'application mondiale ; disparité entre les « changelogs » publics de Facebook et les modifications identifiées via les jeux de données OTA.

Recours à des pratiques informelles ou arbitraires

Identifier des pratiques informelles d'autorégulation, autrement dit des décisions qui sont prises par les plateformes mais qui n'apparaissent pas expressément fondées sur une disposition des conditions d'utilisation. Ces pratiques informelles caractérisent généralement un aléa/un défaut de prévisibilité concernant les règles opposables aux utilisateurs ou un recours aux seules « valeurs » qui imprègnent la gouvernance de l'entreprise.

> Ex. : dans le cas de Facebook, recours au *crosscheck*/double standard de modération pour les personnalités publiques, avant la documentation de cette pratique en 2021 par le Conseil de surveillance de Meta/la lanceuse d'alerte Frances Haugen.
>
> Ex. : caractère non public des listes d'organisations dites « dangereuses », ne permettant pas de déterminer si ces dernières correspondent scrupuleusement aux critères fixés dans les standards relatifs aux contenus terroristes et extrêmement violents (TVEC, cf. *infra*).

Imperméabilité des instruments à l'égard du droit positif

Identifier des hypothèses dans lesquelles des conditions d'utilisation n'ont pas fait l'objet de modification substantielle malgré l'entrée en vigueur de réglementations ou l'adoption de sanctions visant les entreprises en question. L'absence d'amendement des conditions d'utilisation sur une longue durée (plusieurs années) peut refléter un défaut d'incorporation ou de prise en compte du cadre législatif en vigueur.

Amendements fréquents

Souligner des modifications marginales mais fréquentes des conditions d'utilisation, y compris cosmétiques. Cette démarche permet généralement de révéler l'existence d'un travail de veille juridique soutenu (ex. : ponctuation, numérotation, remplacement d'une notion par une autre dans un texte sans modifier la nature du droit ou de l'obligation).

Problèmes d'intelligibilité, de clarté ou de prévisibilité

Déterminer si les amendements aux condition d'utilisation relatifs à la modération de contenus illicites résolvent ou aggravent les problèmes d'intelligibilité de la « norme » édictée. La présence

d'un renvoi explicite à des législations et réglementations d'origine étatique facilite généralement l'interprétation des CGU et standards.

> Ex. : renvoi par les standards d'Instagram à la *NetzDG* allemande ou au règlement UE sur les contenus terroristes dans le but de clarifier le sens des prescriptions.

> Ex. : distinction textuelle entre les standards applicables aux utilisateurs ressortissants de l'UE et ceux applicables aux ressortissants extra-européens.

Concomitance entre des modifications de gouvernance et des amendements

Déterminer si les amendements aux conditions d'utilisation peuvent refléter des modifications structurelles au sein des entreprises administrant les réseaux sociaux, ou en résulter directement. Sont ainsi envisagées les changements de direction, suppression des services, réductions des effectifs, etc.

> Ex. : modifications des conditions d'utilisation depuis l'acquisition de Twitter par Elon Musk en novembre 2022.

PRINCIPAUX CONSTATS DU POINT DE VUE DES DROITS DE L'HOMME

- Les conditions d'utilisation des plateformes **ont un impact indéniable sur les droits des utilisateurs** dans le domaine de la vie privée, de la liberté d'expression ou du droit à la vie. La remise en cause par la Cour suprême des Etats-Unis, le 24 juin 2022[3], de l'arrêt *Roe v. Wade* a notamment démontré que les réseaux sociaux peuvent être exploités pour collecter les données d'utilisatrices qui souhaitent se rendre sur le territoire d'un autre État en vue de leur avortement[4]. Or, la pratique française et internationale établit que les droits reconnus aux individus par une variété d'instruments juridiques s'appliquent également en ligne. L'observation de l'évolution des CGU permet donc de constater que la teneur de ces droits est affectée par **l'ajout, la suppression ou la modification de règles précisant les comportements autorisés ou prohibés** en ligne.
- Certains réseaux sociaux **ont fait des efforts de transparence** concernant l'évolution de leurs conditions d'utilisation grâce à des fonctionnalités permettant la comparaison de différentes versions des textes (ex : le « changelog » de Meta, cf. encadré). Les utilisateurs se voient parfois notifier la mise à jour des conditions d'utilisation à l'ouverture de l'application de réseau social. Cela ne signifie pas pour autant que les utilisateurs les consultent, soient incités à le faire ou soient en mesure de comprendre les modifications opérées.
- Les termes employés pour décrire les droits et devoirs des utilisateurs **n'ont pas toujours de sens ou de portée concrète sur le terrain du droit**, ce qui rend difficile leur appréciation à la lumière des instruments juridiques pertinents sauf à les requalifier. À l'inverse, certains termes juridiques sont utilisés par les réseaux sociaux dans une acception différente de celles des droits étatiques. Il convient pour appréhender la portée réelle des conditions d'utilisation de s'attacher à la teneur du texte et aux modalités de leur mise en œuvre plutôt qu'aux termes employés.
- Sous réserve de cas spécifiques (ex : TikTok), les réseaux sociaux tendent à **appliquer leurs conditions d'utilisation de façon globale, selon des standards juridiques qui sont inspirés de la common law**, par exemple en termes de liberté d'expression ou d'identification des catégories de personnes protégées. Compte tenu des divergences entre les normes européennes et américaines, les réseaux sociaux alternent entre deux approches : soit ils adaptent leurs normes aux régions dans lesquelles ils opèrent, soit ils étendent les normes dérivées de la législation américaine à l'échelle mondiale. Dans le premier cas, les normes de modération seront adaptées pour se conformer à la législation locale, mais il en résultera probablement une approche fragmentée. Dans le second cas, les réseaux sociaux risquent de se voir infliger des sanctions par les autorités locales pour non-respect des lois en vigueur. En tout état de cause, cette transplantation de concepts et notions dans le système français et européen suscite des risques d'incohérence.
- L'analyse des relevés OTA et leur confrontation à la pratique démontre **l'existence de pratiques arbitraires en matière de modération des contenus**, c'est-à-dire non fondées sur des dispositions spécifiques des conditions d'utilisation ou non précédées de modifications de ces dernières. Twitter, en particulier, s'est à de nombreuses reprises écarté de ses propres standards après son acquisition par Elon Musk, suscitant un important manque de prévisibilité juridique du point de vue des utilisateurs.

[3] Dobbs v. Jackson Women's Health Organization.

[4] https://www.theguardian.com/us-news/2022/aug/10/facebook-user-data-abortion-nebraska-police ; https://mashable.com/article/police-using-facebook-google-user-data.

- Certaines **politiques de modération ne sont pas traduites en français** ou ne le sont que de manière tardive, ce qui empêche les utilisateurs non-anglophones d'accéder aisément à des informations essentielles sur les comportements autorisés ou non.
- Les rares références aux droits de l'homme ou au droit international, par exemple en matière de définition des contenus terroristes ou haineux, tendent à disparaître des conditions d'utilisation au profit de critères définis en interne par les entreprises de réseaux sociaux. Il n'existe qu'une correspondance très relative entre les engagements des réseaux sociaux en matière de droits de l'homme et leur mise en œuvre dans les conditions d'utilisation (cf. pour une illustration l'encadré sur les violences sexuelles et de genre).

Déclarations d'adhésion aux droits de l'homme

Plusieurs réseaux sociaux ont publié des engagements en matière de droits de l'homme, parfois disponibles en français. Ex : chez **Tiktok** https://www.tiktok.com/transparency/fr-fr/upholding-human-rights/ - « Notre philosophie s'inspire de plusieurs cadres internationaux relatifs aux droits de l'Homme et nous nous engageons à les respecter, notamment (1) la Charte internationale des droits de l'Homme (qui comprend la Déclaration universelle des droits de l'Homme [DUDH], le Pacte international relatif aux droits civils et politiques [PIDCP], le Pacte international relatif aux droits économiques, sociaux et culturels [PIDESC] et le Pacte international relatif aux droits civils et politiques [PIDCP]), et le Pacte international relatif aux droits économiques, sociaux et culturels [PIDESC]), (2) la Déclaration de l'Organisation internationale du travail [OIT] relative aux principes et droits fondamentaux au travail, (3) la Convention relative aux droits de l'enfant [CDE], et (4) les Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'Homme [UNGP] ».

Chez **Meta**, https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf : "We are committed to respecting human rights as set out in the United Nations Guiding Principles on Business and Human Rights (UNGPs). This commitment encompasses internationally recognized human rights as defined by the International Bill of Human Rights — which consists of the Universal Declaration of Human Rights; the International Covenant on Civil and Political Rights; and the International Covenant on Economic, Social and Cultural Rights — as well as the International Labour Organization Declaration on Fundamental Principles and Rights at Work".

Chez **Google**, https://about.google/human-rights/ : « Les initiatives de Google en faveur des droits civiques et des droits de l'homme s'inscrivent dans le cadre de notre Programme de défense des droits de l'homme, une fonction centrale qui veille à ce que nous respections, au sein de l'entreprise Google et à travers tous nos produits (y compris nos équipements matériels, la recherche Google, Cloud et YouTube), l'engagement pris concernant les Principes directeurs relatifs aux entreprises et aux droits de l'homme de l'ONU, les Principes de la Global Network Initiative et les autres mécanismes de protection des droits civiques et des droits de l'homme ».

Chez **Twitter/X**, https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice – "This is a global commitment, and while grounded in the United States Bill of Rights and the European Convention on Human Rights, it is informed by a number of additional sources including the members of our Trust and Safety Council, relationships with advocates and activists around the globe, and by works such as United Nations Principles on Business and Human Rights".

ILLUSTRATIONS

1. LIBERTE D'EXPRESSION

Les amendements apportés aux conditions d'utilisation des réseaux sociaux ont un impact sur la liberté d'expression, englobant la liberté de recevoir et de communiquer des informations et encadrée notamment à l'article 19 du Pacte international relatif aux droits civils et politiques (PIDCP), l'article 10 de la Convention européenne des droits de l'Homme ou encore l'article 13 de la Convention américaine des droits de l'homme. Comme l'a déjà exposé le Conseil des droits de l'Homme des Nations Unies (« Liberté d'opinion et d'expression », 8 juillet 2022, A/HRC/RES/50/15), les réseaux sociaux ont un rôle à jouer « *pour favoriser l'exercice du droit à la liberté d'opinion et d'expression et l'accès à l'information* » et « *sont tenu[s] de respecter les droits de l'homme, tant en ligne que hors ligne* »[5].

> V. les résolutions pertinentes de la Commission des droits de l'homme et du Conseil des droits de l'homme sur le droit à la liberté d'opinion et d'expression : résolutions 20/8 et 26/13, datées respectivement du 5 juillet 2012 et du 26 juin 2014, du Conseil, sur la promotion, la protection et l'exercice des droits de l'homme sur Internet, ainsi que les résolutions 12/16 du 2 octobre 2009, sur la liberté d'opinion et d'expression, 28/16 du 24 mars 2015, sur le droit à la vie privée à l'ère du numérique, et 23/2 du 13 juin 2013, sur le rôle de la liberté d'opinion et d'expression dans l'émancipation des femmes, et rappelant également les résolutions 68/167, datée du 18 décembre 2013, et 69/166, du 18 décembre 2014, sur le droit à la vie privée à l'ère du numérique, 70/184, du 22 décembre 2015, sur les technologies de l'information et de la communication au service du développement, et 70/125, du 16 décembre 2015, contenant le document final de la réunion de haut niveau de l'Assemblée générale sur l'examen d'ensemble de la mise en œuvre des textes issus du Sommet mondial sur la société de l'information, de l'Assemblée générale.

Après une succession d'ajustements depuis le début de la décennie, certaines plateformes, comme **Meta** (par l'intermédiaire de Facebook et Instagram), ont explicité leur volonté de favoriser davantage la liberté d'expression. En février 2023, Meta a clarifié sa définition de la valeur "*voice*" (« voix » ou « parole ») en expliquant que l'entreprise souhaite permettre aux utilisateurs de s'exprimer ouvertement sur les sujets qui leur tiennent à cœur, que ce soit à travers des commentaires écrits, des photos, de la musique ou d'autres formes d'expression artistique.

> *"L'objectif de nos Standards de la communauté est de créer un lieu d'expression qui donne la parole à tous. Meta souhaite que les utilisateurs puissent s'exprimer ouvertement sur les sujets qui comptent pour eux, que ce soit via des commentaires, des photos, de la musique ou d'autres moyens d'expression artistique, même si certains peuvent être en désaccord ou y trouver à redire. Dans certains cas, nous autorisons la publication de contenu qui va à l'encontre de nos Standards s'il est pertinent et d'intérêt public."*

Cette orientation suggère une volonté de favoriser la communication d'une diversité de vues sur la plateforme (chez TikTok il est question de « *développer l'imagination humaine par le biais de l'expression créative* »). Cependant, malgré cette volonté affichée, les plateformes telles que Meta continuent de définir de plus en plus précisément certains sujets sensibles, susceptibles de restrictions. Par exemple, Meta a établi des règles strictes concernant les contenus injurieux, incitant à la haine ou diffamatoires (cf. plus loin les développements sur ces aspects). Bien que la

[5] Article 19 du PIDCP : « Toute personne a droit à la liberté d'expression ; ce droit comprend la liberté de rechercher, de recevoir et de répandre des informations et des idées de toute espèce, sans considération de frontières, sous une forme orale, écrite, imprimée ou artistique, ou par tout autre moyen de son choix ». Article 21 du PIDCP : « L'exercice de ce droit ne peut faire l'objet que des seules restrictions imposées conformément à la loi et qui sont nécessaires dans une société démocratique, dans l'intérêt de la sécurité nationale, de la sûreté publique, de l'ordre public ou pour protéger la santé ou la moralité publiques, ou les droits et les libertés d'autrui ». Observation générale n° 34 du Comité des droits de l'homme, CCPR/C/GC/34, 12 sept. 2011.

lutte contre les discours haineux et offensants soit essentielle pour maintenir un environnement sûr et respectueux en ligne, il existe un risque que ces règles restrictives puissent limiter la liberté d'expression des utilisateurs. La ligne de démarcation entre ce qui relève d'une expression légitime et ce qui est considéré comme inapproprié apparaît parfois subjective et peut conduire à la censure de contenus qui pourraient, par ailleurs, être considérés comme comportant une forme légitime de critique ou de débat.

À cet égard, notons que les pratiques de modération par les réseaux sociaux ont dans certains cas été caractérisés par l'arbitraire. Par exemple, sur **Twitter**, de nombreux contenus critiquant ou parodiant Elon Musk ont été censurés sans justification, au titre de la lutte contre la désinformation. Elon Musk a ainsi annoncé par un tweet du 6 novembre 2022 que « *Tout compte Twitter se livrant à de l'usurpation d'identité sans spécifier clairement qu'il s'agit d'une 'parodie' sera définitivement banni* ». Ce tweet par lequel Elon Musk présente la nouvelle approche de Twitter en matière de liberté d'expression fait suite à la suspension du compte de l'humoriste américaine Kathy Griffin la veille. Celle-ci avait renommé son compte « Elon Musk » et parodié le style de publication du dirigeant.

Pourtant, ce n'est que bien plus tardivement, en avril 2023, que les règles pertinentes ont été formalisées et précisées par la « Politique en matière d'identités fallacieuses et trompeuses » :

> *« Il est interdit de détourner l'identité de personnes, de groupes ou d'organisations, ou d'utiliser une fausse identité à des fins de tromperie.*
>
> *Nous voulons que les utilisateurs de Twitter puissent y trouver des voix authentiques. Si nous n'exigeons pas que vous affichiez votre véritable nom ou une vraie photo de vous sur votre profil, votre compte ne doit pas utiliser de fausses informations de profil pour se présenter comme une personne ou une entité qui ne vous est pas affiliée, ce qui est susceptible d'induire les autres utilisateurs de Twitter en erreur. (…) Nous interdisons les comportements suivants dans le cadre de cette politique :*
>
> ***Usurpation d'identité***
>
> *Vous ne devez pas vous faire passer pour une personne, une organisation ou un groupe existant dans le but de tromper les autres quant à votre identité ou la personne que vous représentez. Les comptes qui enfreignent cette politique trompent quant à leur identité en utilisant au moins deux éléments d'une autre personne, comme son nom, son image ou de fausses affirmations d'affiliation à une autre personne ou organisation dans leur profil ou leurs Tweets.*
>
> ***Identités trompeuses***
>
> *Vous ne devez pas vous faire passer pour quelqu'un qui n'existe pas dans le but de tromper les autres quant à votre identité ou la personne que vous représentez. Cela inclut l'utilisation trompeuse d'au moins un élément de l'identité d'une autre personne sur votre profil ou dans vos Tweets, comme l'utilisation de l'image de quelqu'un d'autre ou le fait d'indiquer une affiliation à une personne ou entité existante qui ne correspond pas à la réalité. Nous considérons également des comptes comme trompeurs s'ils utilisent une image générée par ordinateur pour se faire passer pour quelqu'un qui n'existe pas ».*

Une exception est prévue en matière de parodie :

> *« Les comptes qui décrivent dans leur profil une autre personne ou organisation ou un autre groupe, et dont le but est de discuter d'informations à son sujet ou d'en partager, ou encore d'en faire la satire, n'enfreignent pas cette politique. Si ces comptes peuvent utiliser des éléments d'une autre identité, **ils incluent aussi dans leur profil certains mots ou d'autres indicateurs montrant qu'ils ne sont pas affiliés à leur objet**. Pour éviter de semer la confusion dans l'esprit des autres utilisateurs quant à leur affiliation, les comptes parodiques, de commentaires ou de fans doivent bien se différencier dans leur nom et dans leur biographie. Les comptes pour lesquels cette différenciation est insuffisante sont considérés comme enfreignant cette politique ».*

Sont ensuite décrits de façon assez précise les comportements qui seront couverts par cette exception :

> *« Nom du compte : le nom du compte doit clairement indiquer que le compte n'est pas affilié à l'objet décrit dans son profil. Pour ce faire, un compte peut intégrer dans son nom un mot tel que « parodie », « faux », « fan » ou « commentaire », par exemple. Cela doit être fait d'une manière qui puisse être comprise par tous et ne doit pas être contredit par d'autres termes évoquant une affiliation, comme le terme « officiel », par exemple. Veuillez noter que le nom du compte n'est pas son nom d'utilisateur (@nomdutilisateur).*
>
> *Biographie : la biographie doit clairement indiquer que le compte n'est pas affilié à l'objet décrit dans son profil. Cette non-affiliation peut être affirmée en intégrant des termes tels que « non affilié à », « parodie », « faux », « fan » ou « commentaire », par exemple. Cela doit être fait d'une manière qui puisse être comprise par tous ».*

La direction de Twitter n'en a pas moins censuré des comptes critiques d'Elon Musk sans fondement, notamment les comptes de plusieurs journalistes américains, le 16 décembre 2022. Cette mesure a suscité des réactions tant à l'échelle de l'Union européenne (« *News about arbitrary suspension of journalists on Twitter is worrying. EU's Digital Services Act requires respect of media freedom and fundamental rights. This is reinforced under our #MediaFreedomAct. @elonmusk should be aware of that*. *There are red lines. And sanctions, soon* », V. Jourova) que de la part du porte-parole du Secrétaire général des Nations Unies (« *The move sets a dangerous precedent at a time when journalists all over the world are facing censorship, physical threats, and even worse* », S. Dujarric). Il convient de souligner à cet égard, concernant la France et l'Europe, que la liberté d'expression englobe la satire et la caricature, tandis que l'exception de parodie protège dans une certaine mesure les droits des individus contre les revendications fondées sur le droit d'auteur[6].

Ainsi, les modifications des conditions d'utilisation des réseaux sociaux présentent des approches contradictoires en matière de liberté d'expression. Alors que certaines plateformes affichent une volonté d'encourager la liberté d'expression, elles établissent également des règles de plus en plus précises et restrictives qui peuvent y porter atteinte. Dans le même temps, certaines mesures restrictives prises contre les contenus publiés ne sont pas fondées sur les conditions d'utilisation et s'avèrent être arbitraires. Trouver un équilibre entre la régulation des contenus préjudiciables et la protection de la liberté d'expression reste un défi majeur pour les réseaux sociaux et nécessite une réflexion continue sur les politiques et les pratiques en vigueur.

Violences sexuelles, violences contre les femmes, violences basées sur le genre

Chez **TikTok**, concernant les violences sexuelles, sont interdits l'exploitation sexuelle ou la violence sexuelle et sexiste, y compris les actes sexuels non consentis, l'abus sexuel sur la base d'images, la sextorsion, l'abus physique et le harcèlement sexuel. Sont interdits l'exploitation humaine, y compris le trafic et la traite des êtres humains. Concernant les violences basées sur le genre : sont interdits les comportements et discours haineux ou la promotion d'idéologies haineuses. Sont visés les contenus ayant pour objectif d'attaquer une personne ou un groupe de personnes en raison de qualités ou attributs protégés, notamment le genre, l'identité de genre, le sexe. Dans son Centre de transparence, TikTok précise mettre tout en œuvre pour respecter les droits humains dans l'ensemble de ses activités et souligne que le respect des droits humains est une responsabilité partagée : par l'Etat et par le « *monde de l'entreprise, dont TikTok fait partie* ». TikTok précise que sa philosophie est « *fondée sur la Charte internationale des droits de l'Homme (qui inclut la Déclaration universelle des droits de l'Homme et la Déclaration de l'OIT relative aux principes et droits fondamentaux au travail) ainsi que sur les Principes directeurs relatifs aux entreprises et aux droits de l'Homme des Nations Unies.* »

[6] https://www.echr.coe.int/documents/d/echr/Guide_Art_10_FRA.

Chez **Facebook/Instagram**, ces comportements ne sont pas mentionnés explicitement, mais **Meta** se donne la possibilité de les supprimer. En ce qui concerne les violences sexuelles, est interdit le contenu qui demande, offre, ou admet offrir des services impliquant une violence de haute gravité (ex. la mutilation génitale féminine) ou soutient l'utilisation de tels services. Sera supprimé tout contenu qui menace de violence, de harcèlement ou d'exploitation sexuelle, ou décrit ou encourage ces pratiques, tout contenu affichant, prônant ou coordonnant des actes sexuels avec des personnes non consentantes. Seront supprimées les images relatant des incidents de violence sexuelle. En ce qui concerne les violences basées sur le genre, elles sont traitées sous l'angle des activités violentes ou haineuses (cf. encadré). Le harcèlement et l'intimidation sont également visés, incluant les attaques utilisant des termes péjoratifs à l'égard des femmes, le harcèlement sexuel, l'interdiction des commentaires sexualisés sévères (même si aucune définition n'est offerte d'un tel comportement). Facebook et Instagram s'appuient sur l'interdiction des discours haineux contre les "caractéristiques protégées" comme le sexe et l'identité de genre.

Chez **Twitter**, il n'est pas fait mention explicite des violences sexuelles, violences contre les femmes et les violences basées sur le genre. Concernant la violence sexuelle : les discours violents sont interdits. Il est interdit de menacer d'infliger des blessures physiques à autrui, ce qui inclut les menaces d'agression sexuelle. Les médias montrant la violence sexuelle sont interdits. Les violences associées à des actes sexuels, qu'elles soient réelles ou simulées, relèvent des comportements sexuels violents notamment le viol réel ou simulé ainsi que la "*violence sexualisée*". Concernant la violence basée sur le genre, sont interdites les attaques fondées sur le sexe, l'identité sexuelle, et le ciblage d'autrui avec des insultes, clichés et autres contenus répétés visant à dégrader des personnes ou à renforcer les stéréotypes négatifs ou préjudiciables au sujet d'une "*catégorie protégée*".

Enfin, chez **YouTube**, il n'est pas explicitement fait mention des violences sexuelles ou violences fondées sur le genre. Proférer des injures ou insultes basées sur des attributs intrinsèques est interdit. Ces attributs incluent aussi le « *statut de groupe protégé* » parmi lesquels : le genre et le sexe. Est ainsi interdite « *l'insulte extrême qui déshumanise un individu basé sur des attributs intrinsèques* ». Est interdit le discours haineux contre les individus et les groupes en raison de leur sexe ou leur genre.

Etude complète : https://grsomedia.files.wordpress.com/2023/10/rapport-vsg-boulvard-chollet-final.pdf

2. DESINFORMATION

Si les réseaux sociaux sont devenus des espaces d'exercice de la liberté d'expression et de débat public, il reste primordial de lutter contre la désinformation, comme l'expose le Conseil des droits de l'homme des Nations Unies (« Liberté d'opinion et d'expression », 8 juillet 2022, A/HRC/RES/50/15). Pourtant, en juin 2023, **YouTube** a modifié ses conditions d'utilisation concernant le contenu lié aux élections. La plateforme a décidé de mettre fin à sa politique de suppression de la désinformation concernant toutes les élections présidentielles américaines, y compris celle de 2020. Cela représente un véritable retour en arrière par rapport aux positions antérieures de YouTube, qui avait pris des mesures pour lutter contre la désinformation lors des élections et qui de manière générale évoquait une véritable « *chasse aux fake news* ». Cette décision vise, selon YouTube, à accorder une plus grande liberté d'expression aux utilisateurs, même au détriment de la véracité des informations diffusées. La plateforme explique en effet que :

> « *La capacité à débattre de façon libre d'idées politiques, même celles qui sont controversées ou basées sur des hypothèses fausses, est essentielle au fonctionnement d'une société démocratique, en particulier en pleine saison électorale* ».

Ces modifications des conditions générales d'utilisation des réseaux sociaux en ce qui concerne les personnalités publiques et la désinformation soulèvent des questions sur l'équilibre entre la liberté d'expression et la responsabilité des plateformes dans la diffusion de contenus précis et fiables. Alors que certains voient ces changements comme une opportunité pour les utilisateurs de s'exprimer librement et de critiquer les personnalités publiques, il faut également s'inquiéter de leur impact sur la qualité de l'information et la confiance du public dans les médias sociaux (cf. sections sur la parodie et sur les processus démocratiques).

3. CONTENUS SENSIBLES, EXTREMES OU HAINEUX

Ces contenus sont visés par différentes dénominations au sein des conditions d'utilisation, souvent dans des sections relatives à la sécurité : contenus dangereux, contenus répréhensibles, contenus violents, comportements criminels, activités illicites ou illégales. Pour le présent rapport, le choix a été fait de regrouper ces modifications qui sont les plus substantielles sur la période étudiée.

CONTENUS SENSIBLES

La notion de contenu sensible renvoie à une variété de publications susceptibles de heurter les utilisateurs par leur violence **physique, psychologique ou morale**. Les conditions d'utilisation des réseaux sociaux étudiés renvoient notamment à la diffusion de comportements suicidaires, à l'automutilation ou tout contenu attentatoire au bien-être ou à la dignité de la personne représentée ou de l'utilisateur le visionnant. Ces contenus ne sont pas toujours illicites dans l'ensemble des ordres juridiques mais leur publication va à l'encontre des valeurs promues par certains réseaux sociaux ou sont perçus comme particulièrement choquants (ex : chez Facebook et Instagram, les actes de torture commis contre un ou plusieurs individus, lesquels doivent *a minima* être accompagnés d'un avertissement).

Twitter a par exemple modifié en février 2023 sa politique concernant la représentation de personnes décédées. Auparavant était prévue la suspension immédiate d'utilisateurs publiant de telles images à des fins dites « sadiques ». Le caractère automatique de la suspension du compte disparaît. « *We will* » devient « *we may* », supposant l'appréciation de la situation par la plateforme[7]. Les utilisateurs publiant de tels contenus par manque de vigilance, pour exprimer leur condamnation de faits, leur dégoût, leur état de choc, étaient relativement protégés. Les possibilités de retrait des contenus (représentant la personne décédée avant ou après son décès) sont étendues, y compris à la demande des proches. Sont compris les partages, la diffusion de contenus en direct ou la diffusion par les images de profil.

> *« Le partage d'images ou de vidéos d'une personne décédée peut être à l'origine d'une grande souffrance pour sa famille et peut également avoir un impact négatif sur le bien-être de ceux qui voient ce contenu. Par respect pour la personne décédée et les personnes touchées par sa mort, et pour réduire l'impact de l'exposition non souhaitée à des médias crus, nous pouvons vous demander de supprimer les images et les vidéos montrant le décès d'un individu identifiable.*
>
> *En vertu de cette politique, nous pouvons vous demander de supprimer les images ou vidéos prises au moment du décès d'une personne, ou juste avant ou après celui-ci, si nous recevons une demande en ce sens de sa famille ou d'un représentant autorisé. Il est également interdit de partager à des fins sadiques des médias représentant une personne décédée, ou des images ou vidéos excessivement macabres.*
>
> *Parmi les types de contenus qui peuvent être concernés par cette politique, citons notamment :*

[7] Pour un exemple chez YouTube concernant la distinction entre will et may : publication le 7 juillet 2022 d'une vidéo relative aux modalités de traitement des requêtes consécutives à une suppression de vidéo : https://www.youtube.com/watch?v=fxJG1WCk2FI&t=73s

- *les images ou vidéos dans lesquelles une personne raisonnablement identifiable est clairement décédée ;*
- *les images ou vidéos représentant le meurtre d'une personne identifiable ;*
- *les médias excessivement macabres représentant le décès d'une personne identifiable ;*
- *les images ou vidéos d'une personne décédée identifiable, partagées à des fins sadiques, notamment des médias accompagnés d'un contenu qui :*
    - *se moque du défunt de quelque manière que ce soit ;*
    - *prend plaisir à la souffrance du défunt. »*

La règle est modifiée en mars-avril 2023 pour préciser les actions envisageables concernant les images ne constituant pas des violations, en raison de « *facteurs d'intérêt public* ». Outre le fait de ne pas retirer cette image, Twitter dit être susceptible de ne pas réduire la visibilité du contenu concerné dans ce contexte. Il s'agit de trouver un équilibre précaire entre la protection de la dignité de la personne défunte, la protection de ses proches, la protection du public et le droit de ce dernier d'être informé. Le réseau social se retrouve *de facto* en position d'articuler protection de la vie privée, protection de la dignité de la personne humaine et liberté d'expression.

Les politiques sur les comportements suicidaires et l'automutilation ont également fait l'objet d'amendements. En avril 2023, **YouTube** modifie sa politique pour y intégrer les troubles d'ordre alimentaire, considérés comme des contenus sensibles susceptibles de heurter la communauté des utilisateurs. YouTube s'en explique dans une publication de blog et affirme avoir bénéficié de l'expertise d'ONG. Sont ainsi encadrés :

> "*content about eating disorders that feature imitable behavior, or behavior that we worked with experts to determine can lead at-risk viewers to imitate. This could include videos that show or describe: Disordered eating behaviors, such as purging after eating or severely restricting calories; Weight-based bullying in the context of eating disorders*".

Ces publications ne sont toutefois pas supprimées lorsqu'elles n'encouragent pas les troubles d'ordre alimentaire. Sont privilégiés la mise en place d'avertissements, de restrictions d'accès aux mineurs ou de liens vers des ressources documentaires.

Utilisateurs mineurs

Le mineur, considéré comme une personne vulnérable (notamment dans le DSA), doit bénéficier d'une protection renforcée sur les réseaux sociaux. Ainsi, les réseaux sociaux ainsi que les institutions publiques ont abordé la question de l'âge limite d'accès aux réseaux sociaux. La majorité des réseaux sociaux semblait avoir tranché cette question en établissant l'âge minimum à 13 ans. Seul YouTube ne suivait pas cette pratique. Cependant, en décembre 2022, YouTube est revenue sur sa position pour passer l'âge minimum de 15 à 13 ans. Si ce changement permet à YouTube de s'aligner sur les autres réseaux sociaux, certains États ne semblent pas partager cette position. En France par exemple, la loi du 7 juillet 2023 visant à instaurer une majorité numérique et à lutter contre la haine en ligne établit l'âge minimum pour l'inscription sur des services de réseaux sociaux à 15 ans.

CONTENUS TERRORISTES, EXTREMISTES ET VIOLENTS

Les contenus terroristes et extrémistes correspondent à des comportements qui ne sont pas définis de façon consensuelle à l'échelle internationale et qui sont appréhendés par une variété

d'instruments [8] . Il existe à l'échelle des Nations Unies et de l'UE des listes évolutives d'organisations désignées comme terroristes et visées par des sanctions ciblées. En outre, l'UE s'est dotée d'un règlement qui offre quelques critères mais qui ménage une certaine marge d'appréciation en fonction de la situation appréhendée. Adopté le 29 avril 2021, le **règlement (UE) 2021/784 relatif à la lutte contre la diffusion des contenus à caractère terroriste en ligne** est entré en application le 7 juin 2022. Le règlement impose un devoir de vigilance aux plateformes numériques susceptibles de constituer des espaces privilégiés pour la dissémination de contenus illicites, dont les réseaux sociaux. Il les soumet à une obligation de prompt retrait (ou blocage) des contenus illicites présentant un caractère terroriste, ce retrait devant être opéré dans un délai d'une heure suivant la réception d'une injonction adressée par l'autorité compétente d'un État membre (art. 3). Les comportements terroristes sont envisagés (cons. 11 ; art. 2) par renvoi à la directive (UE) 2017/541 relative à la lutte contre le terrorisme comme englobant :

> *« le matériel qui incite ou invite quelqu'un à commettre des infractions terroristes ou à contribuer à la commission de telles infractions, invite quelqu'un à participer aux activités d'un groupe terroriste ou glorifie les activités terroristes y compris en diffusant du matériel représentant une attaque terroriste. La définition devrait également englober le matériel qui fournit des instructions concernant la fabrication ou l'utilisation d'explosifs, d'armes à feu ou d'autres armes, ou de substances nocives ou dangereuses, ainsi que de substances chimiques, biologiques, radiologiques et nucléaires (CBRN), ou concernant d'autres méthodes ou techniques spécifiques, y compris le choix de cibles, aux fins de la commission ou de la contribution à la commission d'infractions terroristes ».*

Sont visés à la fois les textes, images, enregistrements sonores, vidéos et transmissions en direct d'infractions terroristes, notamment lorsqu'ils entraînent le risque de commission d'infractions analogues. Sont en revanche exclus du champ d'application du règlement les contenus présentant un caractère éducatif, journalistique, artistique ou relevant d'une activité de recherche scientifique (art. 1er).

De leur côté, les réseaux sociaux optent fréquemment dans leurs conditions d'utilisation pour des approches qui associent discours terroristes, extrémistes, violents, dangereux et haineux et qui se reposent parfois sur des listes (internes) d'organisations identifiées comme telles. **Facebook** a notamment été critiqué à la suite de la fuite de sa liste d'individus et d'organisations dangereux (octobre 2021), celle-ci s'inscrivant dans le prolongement des enjeux de politique extérieure états-unienne plutôt que de tenir compte des travaux pertinents d'organisations internationales.

De son côté, **Twitter** se départit depuis la fin 2022 d'une partie des référentiels pertinents de droit international sur lesquels elle s'appuyait auparavant. Twitter développe désormais une conception interne de ce que sont les contenus terroristes et extrémistes. En janvier 2023, Twitter a modifié ses politiques concernant les organisations dites « violentes ». La nouvelle politique englobe alors les organisations violentes *et* haineuses. À la faveur de cette modification, Twitter ne se réfère plus aux listes nationales et internationales désignant les organisations terroristes. La notion d'organisation haineuse semble désormais englober les organisations terroristes et violentes et renvoie, à des fins de définition, à une autre politique portant sur les auteurs d'attaques violentes.

> *« Nous supprimerons tout compte tenu par des auteurs d'attaques terroristes, extrémistes violentes ou violentes de masse, ainsi que tout compte glorifiant ces auteurs, ou dont le but est de partager*

[8] Convention pour la répression de la capture illicite d'aéronefs, 16 décembre 1970 ; Convention pour la répression d'actes illicites dirigés contre la sécurité de l'aviation civile, 22 mars 1971 ; Convention pour la répression d'actes illicites contre la sécurité de la navigation maritime, 10 mars 1988 ; Convention internationale pour la répression des attentats terroristes à l'explosif, 15 déc. 1997 ; Convention internationale pour la répression du financement du terrorisme, 9 déc. 1999.

> *des manifestes et/ou des liens tiers vers des pages hébergeant du contenu associé. Nous pouvons aussi supprimer les Tweets diffusant des manifestes ou d'autres contenus produits par ces auteurs ».*

Certains critères relatifs à l'impact des activités terroristes ou extrémistes sur le bien-être des utilisateurs disparaissent des conditions d'utilisation. Twitter a également revu sa définition des organisations ou entités extrémistes recourant à des actes de violence (*violent extremist groups*) pour la rendre plus englobante, en supprimant certains critères (affiliation, promotion à des causes extrémistes, comportements en dehors de Twitter, etc.). La publication de symboles extrémistes n'est plus explicitement listée comme une violation de la politique. De manière intéressante, la politique considère désormais que toute discussion portant sur ces contenus illicites ne constituera pas une violation dès lors qu'elle est menée « *à des fins manifestement pédagogiques, de documentation et/ou à des fins d'information* » (*newsworthiness*). Sont susceptibles d'être autorisés :

> « *les contenus informatifs, si toutefois : ils ne donnent pas de suggestions sur la manière de trouver une arme et de choisir des cibles ; ils ne partagent pas de slogans, symboles ou mèmes haineux, ni de théories du complot haineuses ; ils n'exposent pas l'idéologie de l'auteur d'une attaque, ses choix tactiques ni son plan d'attaque* ».

La pratique récente de Twitter a toutefois démontré l'existence de décisions arbitraires mises en œuvre depuis la prise de contrôle du réseau social par Elon Musk. Sous prétexte d'une « amnistie », la nouvelle direction a rétabli des milliers de comptes qui avaient été bannis véhiculant des contenus terroristes, extrémistes ou complotistes. De nouveaux comptes associés à des organisations terroristes sont également apparus.

En parallèle, **Facebook et Instagram** (avril 2023) ont modifié plus à la marge leurs politiques. Le groupe Meta vise les contenus violents constituant une violation de ses politiques. Le contenu généré par l'auteur d'une l'attaque et les images de tiers représentant le moment de ces attaques sur des victimes visibles risquent la suppression. De façon peut-être plus cosmétique, Meta remplace dans certains pans de sa politique la notion d'« organisation » incitant à la haine par celle d'« entité » incitant à la haine. *L'organisation* correspondait à un groupe de trois individus ou plus qui :

> "*is organized under a name, sign, or symbol; and; has an ideology, statements, or physical actions that attack individuals based on characteristics, including race, religious affiliation, national origin, disability, ethnicity, gender, sex, sexual orientation, or serious disease*".

La notion *d'entité* incitant à la haine correspond désormais à une :

> "*organization or individual that spreads and encourages hate against others based on their protected characteristics. The entity's activities are characterized by at least some of the following behaviors:*
>
> - *Violence, threatening rhetoric, or dangerous forms of harassment targeting people based on their protected characteristics;*
> - *Repeated use of hate speech;*
> - *Representation of Hate Ideologies or other designated Hate Entities, and/or*
> - *Glorification or substantive support of other designated Hate Entities or Hate Ideologies*"

À l'instar de la pratique de Twitter, c'est par le renvoi à une politique sur le discours de haine, par l'identification de leur propension à diffuser ou promouvoir un comportement haineux (glorification, soutien) que sont identifiés les organisations dont les contenus peuvent être

supprimés. Les cibles des comportements sont également centrales dans l'analyse du comportement, sous l'angle des atteintes portées aux « catégories » ou « caractéristiques » protégées. L'approche retenue par les réseaux sociaux se fait de plus en plus fonctionnelle : il s'agit davantage de viser les comportements en tant que tels plutôt que de se limiter à l'identification formelle des organisations.

Les caractéristiques protégées

Les réseaux sociaux étudiés s'appuient sur la notion de caractéristiques protégées (Facebook, Instagram, YouTube), catégories protégées (Twitter), d'attributs protégés ou de qualités protégés (TikTok) pour structurer leurs politiques. Ces notions sont très imprégnées de conceptions nord-américaines du droit de la non-discrimination et visent notamment les personnes attaquées en raison de leur ethnicité, origine nationale, religion, caste, sexe, orientation sexuelle, identité, état de santé ou handicap ou statut d'immigré. Les instruments pertinents du droit français, notamment l'article 225-1 du Code pénal, visent ces catégories sous l'angle des critères prohibés de discrimination : « *Constitue une discrimination toute distinction opérée entre les personnes physiques sur le fondement de leur origine, de leur sexe, de leur situation de famille, de leur grossesse, de leur apparence physique, de la particulière vulnérabilité résultant de leur situation économique, apparente ou connue de son auteur, de leur patronyme, de leur lieu de résidence, de leur état de santé, de leur perte d'autonomie, de leur handicap, de leurs caractéristiques génétiques, de leurs mœurs, de leur orientation sexuelle, de leur identité de genre, de leur âge, de leurs opinions politiques, de leurs activités syndicales, de leur qualité de lanceur d'alerte, de facilitateur ou de personne en lien avec un lanceur d'alerte* ».

4. PROCESSUS DEMOCRATIQUES ET PERSONNALITES PUBLIQUES

Les réseaux sociaux jouent un rôle crucial dans le processus démocratique et la communication de l'État. Pendant longtemps, ces plateformes ont semblé lutter contre la désinformation, établissant des règles strictes à ce sujet. Cependant, les modifications opérées dans les conditions d'utilisation suggèrent un changement de politique dans cette lutte contre la désinformation.

Les conditions générales d'utilisation des réseaux sociaux sont un véritable marqueur de la relation particulière qui se noue entre les États et les plateformes. **Facebook et Instagram** par exemple, après avoir opéré une distinction de réglementation entre les personnalités publiques et les personnes privées, ont ajouté en mars 2023 une définition claire de ce qu'elles considèrent comme une personnalité publique. Cette définition inclut les responsables gouvernementaux au niveau national et local, les candidats politiques à ces fonctions, les personnes ayant plus d'un million de fans ou de followers sur les réseaux sociaux, ainsi que celles bénéficiant d'une importante couverture médiatique.

> "*Nous distinguons les personnalités publiques des personnes privées, car notre plateforme prône l'échange, ce qui implique souvent des commentaires critiques concernant des individus qui attirent l'attention des médias ou du grand public. En ce qui concerne les personnalités publiques, nous supprimons les attaques graves et certaines attaques qui identifient directement la personnalité publique dans une publication ou un commentaire. Nous définissons les personnalités publiques comme les représentants des gouvernements au niveau national et étatique, les candidats politiques à ces fonctions, les personnes ayant plus d'un million de fans ou de followers sur les réseaux sociaux et les personnes qui bénéficient d'une couverture médiatique importante*."

Celle-ci permet alors d'accorder plus de liberté aux utilisateurs dans leurs messages envers les personnalités publiques, qui ne seront supprimés que dans des cas d'attaques particulièrement graves. Cette évolution suggère un changement de politique qui pourrait avoir des implications sur la manière dont les critiques ou les débats politiques sont traités sur ces plateformes.

Le rétablissement du compte de Donald Trump

Après avoir acquis **Twitter**, Elon Musk a affirmé son souhait de transformer la plateforme en véritable enceinte de démocratie participative. Il a donc multiplié les sondages, relatifs notamment à l'opportunité de rétablir des comptes d'utilisateurs bannis par ses prédécesseurs à la suite de l'assaut du Capitole du 6 janvier 2021, entre autres l'ancien président Donald Trump. 52% des votants étant favorables au rétablissement du compte, Elon Musk s'est exécuté le 20 novembre 2022 : « *The people have spoken. Trump will be reinstated. Vox Populi, Vox Dei*" (tweet).

Du côté de **Meta**, les comptes de l'ancien président ont été rétablis début 2023. Nick Clegg, président en charge des affaires mondiales, s'en est justifié dans une publication de blog en soulignant l'intérêt du public à avoir connaissance de ces publications : "*To assess whether the serious risk to public safety that existed in January 2021 has sufficiently receded, we have evaluated the current environment according to our Crisis Policy Protocol, which included looking at the conduct of the US 2022 midterm elections, and expert assessments on the current security environment. Our determination is that the risk has sufficiently receded, and that we should therefore adhere to the two-year timeline we set out. As such, we will be reinstating Mr. Trump's Facebook and Instagram accounts in the coming weeks. However, we are doing so with new guardrails in place to deter repeat offenses*". Nick Clegg s'appuyait ainsi sur la nouvelle politique relative aux contenus d'intérêt médiatique (newsworthy content) https://transparency.fb.com/en-gb/features/approach-to-newsworthy-content/

YouTube a rétabli ses comptes le 17 mars 2023 au nom du pluralisme en période électorale : *"Starting today, the Donald J. Trump channel is no longer restricted and can upload new content. We carefully evaluated the continued risk of real-world violence, while balancing the chance for voters to hear equally from major national candidates in the run up to an election"*.

5. PROTECTION DE LA VIE PRIVEE / DES DONNEES A CARACTERE PERSONNEL

Dans cette partie, nous nous concentrerons sur l'impact de l'évolution des CGU sur la protection de la vie privée, « *en vertu [de laquelle] nul ne peut être l'objet d'immixtions arbitraires ou illégales en lien avec son domicile ou sa correspondance ou dans sa vie privée et sa vie familiale* ». Ce droit est encadré par des instruments internationaux tels que l'article 17 du Pacte international relatif aux droits civils et politiques (PIDCP), l'article 8 de la Convention européenne des droits de l'homme, l'article 11 de la Convention américaine relative aux droits de l'homme ou le RGPD dans sa globalité.

Comme a déjà pu l'exposer l'Assemblée générale des Nations Unies, la collecte et l'utilisation des données des utilisateurs constitue l'une des principales préoccupations relatives à la protection de la vie privée. Si les organisations internationales expriment de façon croissante leur préoccupation face au traitement des données personnelles, les réseaux sociaux tendent quant à eux vers une utilisation toujours plus importante de ces données[9].

TikTok, par exemple, a étendu sa politique de collecte de données en avril 2023. Cette extension inclut désormais toutes les interactions avec un chatbot ou un assistant virtuel, ainsi que la synchronisation des contacts. Cela signifie que TikTok recueille une quantité encore plus importante de données personnelles de ses utilisateurs. Ces pratiques pourraient potentiellement porter atteinte au droit au respect de la vie privée, sans compter que les utilisateurs ne sont peut-être pas pleinement conscients de l'étendue de la collecte de leurs données et des implications

[9] En matière de coopération avec les autorités judiciaires, en mai 2023, Twitter a supprimé l'expression « DM » de la liste des contenus pour lesquels les autorités doivent fournir un mandat. Le sens et l'impact de cette modification ne sont pas clairs dans la mesure où elle reviendrait à simplifier la communication d'échanges privés aux autorités.

https://help.twitter.com/en/rules-and-policies/twitter-legal-faqs

qui en découlent. TikTok a également fait évoluer sa politique de partage des données. En effet, les données des utilisateurs peuvent à présent être partagées pour faciliter des recherches menées par des chercheurs indépendants, des recherches historiques ou scientifiques. Cependant, les contours précis de ce partage ne sont pas clairement définis, de sorte qu'il existe un risque que des informations sensibles soient partagées avec des tiers sans le consentement libre et éclairé des utilisateurs. Il s'agit peut-être d'une mise en œuvre anticipée de l'article 40 du Règlement sur les services numériques, prévoyant à son §4 que « *Sur demande motivée du coordinateur pour les services numériques de l'État membre d'établissement, les fournisseurs de très grandes plateformes en ligne ou de très grands moteurs de recherche en ligne fournissent, dans un délai raisonnable spécifié dans la demande, l'accès aux données à des chercheurs agréés qui satisfont aux exigences énoncées au paragraphe 8 du présent article, à la seule fin de procéder à des recherches contribuant à la détection, au recensement et à la compréhension des risques systémiques dans l'Union* (…) ».

De son côté, **Twitter** a introduit en décembre 2022 le partage d'informations "*live*" dans le cadre de la lutte contre les atteintes aux règles de la plateforme.

> "Nous prenons ceci en compte, car certains types d'informations privées ou publiées en temps réel comportent des risques plus élevés que d'autres en cas de partage sans autorisation. Notre objectif principal est de protéger les personnes de tout préjudice physique potentiel résultant du partage de leurs informations. Nous considérons donc que les informations telles que la localisation géographique et le numéro de téléphone présentent un risque plus élevé que les autres types d'informations. Pour ce qui est des informations publiées en temps réel ou datant du même jour, il est possible que la personne se trouve encore à l'endroit indiqué."

Bien que cette mesure puisse sembler nécessaire pour garantir la sécurité en ligne et prévenir les abus, elle peut également avoir des répercussions sur le droit à la vie privée des utilisateurs. La surveillance constante des activités en ligne peut être considérée comme une atteinte à la vie privée, surtout si elle est effectuée sans une supervision adéquate ou des garanties suffisantes pour protéger les droits fondamentaux des utilisateurs.

Ainsi, les évolutions des conditions générales d'utilisation des réseaux sociaux ont des implications significatives sur le droit à la vie privée des utilisateurs. Les politiques de collecte et de partage des données peuvent potentiellement porter atteinte à ce droit fondamental, ce qui souligne la nécessité de réglementer ces pratiques et de renforcer la protection des droits de l'homme dans le contexte numérique.

RECOMMANDATIONS

- Il est nécessaire que les pouvoirs publics, organisations de la société civile et chercheurs de différentes disciplines poursuivent et systématisent les **démarches de suivi de l'évolution des conditions d'utilisation**. Le volume important des conditions d'utilisation justifie des démarches parallèles de veille à l'égard d'enjeux spécifiques : protection de la vie privée, protection de la liberté d'expression, droits des consommateurs ou lutte contre la désinformation.
- L'étude exploratoire a soulevé de nombreux **enjeux linguistiques**, notamment l'indisponibilité des conditions d'utilisation dans certaines langues cibles ou des disparités entre versions. Les réseaux sociaux devraient fournir les conditions d'utilisation dans la langue cible, conformément aux obligations fixées par les instruments pertinents, notamment l'article 14 du Règlement européen sur les services numériques.
- **L'énumération des comportements interdits** présente certainement un intérêt pour ce qui est de leur compréhension par les utilisateurs des services. D'un point de vue juridique en revanche, elle présente l'inconvénient d'entraîner des **risques d'actualisations fréquentes**, lesquelles ne sont pas toujours notifiées aux utilisateurs. Elles créent aussi des risques de chevauchements à l'égard de différentes qualifications issues du droit positif ou, au contraire, de contradictions. Il est souhaitable que les listes fournies soient considérées comme des faisceaux d'indices destinés à éclairer les différentes parties prenantes (ex. : pour l'identification de discours dits haineux ou de comportements considérés comme extrémistes ou terroristes) et **ne puissent pas mettre en échec les catégories issues du droit positif**.
- Le recours à des **publications de blog ou autres communiqués de presse** comme canaux autonomes d'amendement des conditions d'utilisation est problématique. En effet, les utilisateurs n'ont pas nécessairement connaissance de leur existence, de sorte qu'ils ne disposent pas de tous les éléments susceptibles d'éclairer leur compréhension des règles. En outre, ces communications ne sont pas systématiquement traduites depuis l'anglais.
- Compte tenu de l'adoption de déclarations par lesquelles elles s'engagent à se conformer aux droits de l'homme, les entreprises de réseaux sociaux devraient autant que possible faire correspondre le contenu de leurs conditions d'utilisation aux prescriptions issues des instruments internationaux pertinents. L'inverse est constaté chez certains réseaux sociaux qui **suppriment des références au droit international et à la pratique des organisations internationales**.
- De nombreux pans des conditions d'utilisation n'ont pas été abordés dans cette étude dans la mesure où certaines politiques n'ont pas été amendées durant la période de veille. Les conditions d'utilisation n'en méritent pas moins d'être scrutées dans leur globalité, par les chercheurs, les autorités compétentes et la société civile.

CONTACT

COURRIEL :

Valère Ndior
Université de Bretagne occidentale, Lab-LEX
12 rue Kergoat
29200 Brest
ndior [at] univ-brest.fr

SITES INTERNET :

Laboratoire Lab-LEX

https://www.univ-brest.fr/lab-lex/fr

Programme Gouvernance et régulation des réseaux sociaux

https://grsomedia.wordpress.com

Centre GEODE

https://geode.science

Couverture rapport