

# Questionnaire 2020 aux opérateurs de plateformes en ligne soumis au titre III de la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information

## Twitter

### Questions liminaires : présentation du service

- **Nom du service** : Twitter
- **Nom et raison sociale de la société opératrice** : Twitter International Company
- **Identité et fonction du représentant légal exerçant les fonctions d'interlocuteur référent sur le territoire français au sens de l'article 13 de la loi du 22 décembre 2018** : Audrey Herblin-Stoop - Public Policy Director
- **Nombre de visiteurs uniques par mois de la plateforme en ligne, calculé sur la base de l'année précédant l'année d'exercice (2019)** : Notre mesure d'audience est le nombre d'utilisateurs actifs quotidiens monétisables (mDAU). Nous mesurons cette audience à l'échelle mondiale. Notre dernier chiffre pour le troisième trimestre de l'année 2019 est de 152 millions de mDAU.
- **Modalités de mise à disposition du service auprès du public français** :
- **Disponibilité** :
  - **sur internet (URL du site internet)** : [www.twitter.com](http://www.twitter.com)
  - **via des applications (systèmes d'exploitation concernés)** : IOs, Android
  - **via des assistants vocaux** : N/A
- **Langue du service sur le territoire français (par défaut)** : français  
**Outre la langue, le service disponible sur le territoire français constitue-t-il une version spécifique du service (pour le territoire français ou pour le public francophone) ?** : non
- **Description du service** :
  - **Objet** : servir la conversation publique
  - **Population(s) ciblée(s) (tranches d'âges, communautés d'intérêt...)** : toutes (15 ans et plus)
  - **Types de contenus qui peuvent être publiés et consommés sur le service (vidéos, textes, sons, images, jeux vidéo...)** : texte, images, vidéos courtes, vidéos live (en temps réel)
  - **Thématique(s)** : toutes les thématiques peuvent être discutées sur notre plateforme à condition qu'elles respectent nos conditions générales d'utilisation
- **Modèle économique du service (types d'activités rémunératrices pour l'opérateur tirées de l'exploitation du service en France)** : publicité et licences de données
- **Effectifs employés aux activités du service dans le monde et en France** : Twitter emploie 5500 personnes dans le monde. L'entité française, Twitter France SAS, emploie 30 salariés.
- **Comptes de résultats pour l'activité du service en France en 2020** : les résultats financiers pour 2020 ne sont pas encore publiés. Les derniers résultats financiers que nous avons déposés pour Twitter France SAS étaient pour l'exercice clos le 31 décembre 2019 et ceux-ci sont accessibles au public. Ils montrent les résultats suivants :



- **Chiffres d'affaires** : 13 662 654 €
  - **Produits d'exploitation** : 14 276 959 €
  - **Charges d'exploitation** : 13 262 614 €
  - **Résultat net** : 1 014 345 €
- **Autres informations ou spécificités** : Twitter, Inc. est tout ce qui se passe dans le monde et ce dont les gens parlent en ce moment. Des dernières nouvelles et des divertissements, aux sports, à la politique et aux intérêts quotidiens, Twitter montre toutes les facettes d'un même sujet. Participer à la conversation ouverte. Regarder les événements en direct. Disponible dans plus de 40 langues dans le monde entier, le service est accessible via twitter.com et un ensemble d'appareils mobiles. Twitter est une plateforme mondiale d'expression publique et de conversation en temps réel. Twitter permet aux gens de consommer, créer, distribuer et découvrir du contenu en démocratisant la création et la distribution de ce dernier. En décembre 2020, nous avons pris la décision de mettre un terme à notre application de diffusion de contenu en direct Periscope en tant qu'application distincte, d'ici mars 2021. Par conséquent, les fonctionnalités de Periscope sont progressivement intégrées à Twitter.

## Questions générales

### Définition des fausses informations

L'opérateur indiquera s'il a adopté une définition des fausses informations et/ou de notions proches ; le cas échéant, il la fournira, il précisera à quelle fin (travail de modération, règles d'utilisation du service...) et si cette définition vaut globalement ou diffère selon les zones géographiques de mise à disposition du service.

Afin de servir la conversation publique, l'objectif de Twitter est de faciliter la recherche d'informations crédibles sur la plateforme et de limiter la diffusion de contenus potentiellement dangereux et trompeurs. C'est pourquoi nous avons adopté des définitions de travail de plusieurs concepts clés afin de prendre en compte le fait que les contenus faux ou trompeurs peuvent prendre de nombreuses formes. Les trois grandes catégories de concepts suivantes - que nous appliquons universellement à l'ensemble des services de Twitter à travers le monde - nous aident à prendre différents types d'actions sur les contenus problématiques dans notre travail quotidien de modération :

- Informations trompeuses - déclarations ou affirmations dont le caractère faux ou trompeur a été confirmé par des experts en la matière, tels que les autorités de santé publique.
- Affirmations contestées - déclarations ou affirmations dont l'exactitude, la véracité ou la crédibilité sont contestées ou encore inconnues à ce jour.
- Affirmations non vérifiées - informations (qui peuvent être vraies ou fausses) qui ne sont pas confirmées au moment où elles sont partagées.

Veillez noter que notre travail ne se limite pas à ces concepts car nous disposons d'une série d'autres politiques pour lutter contre la diffusion d'informations fausses et trompeuses, y compris en matière de médias synthétiques et manipulés pour lesquels nous fournissons des informations plus détaillées plus loin dans le questionnaire. Nous mettons régulièrement à jour nos règles et politiques afin de mieux refléter et combattre les nouvelles menaces émergentes et ainsi garder une longueur d'avance sur les nouvelles tactiques employées par les acteurs malveillants qui tenteraient de manipuler notre plateforme et perturber la conversation publique.

### Lutte contre les fausses informations sur le service

Il indiquera si les fausses informations (ou notion proche) sont interdites sur son service et le cas échéant, où l'utilisateur peut trouver l'information relative à cette interdiction.

Comme indiqué ci-dessus, nous prenons différents types de mesures en fonction des trois catégories suivantes de contenu potentiellement préjudiciable et trompeur (informations trompeuses, affirmations contestées et affirmations non vérifiées). Ces informations sont facilement accessibles à tous sur notre site web (via les pages



relatives à notre politique en matière d'intégrité civique et d'informations trompeuses sur la COVID-19) et sur ce [billet de blog](#). Ces concepts font partie intégrante de nos politiques concernant la diffusion de fausses informations sur Twitter, notamment autour de la désinformation liée à la COVID-19, à l'intégrité civique et aux médias synthétiques et manipulés. Cependant, Twitter n'a pas vocation à évaluer la véracité de tous les contenus publiés sur sa plateforme, car cela rendrait celle-ci inopérante.

**L'opérateur est invité à indiquer au Conseil le nombre de contenus identifiés comme comportant une fausse information sur son service en 2020, toutes sources de détection confondues (interne, utilisateurs, fact-checkers...), leur part dans l'ensemble des contenus du service (en pourcentage) et le nombre d'émetteurs et de réémetteurs de ces contenus, ainsi que le nombre de ces contenus et de ces émetteurs ayant fait l'objet d'une mesure en conséquence, en déclinant en fonction du type de mesures adoptées. Il fera état des évolutions de ces chiffres entre 2019 et 2020.**

La transparence est au cœur du travail que nous faisons chez Twitter. Nous avons récemment mis à jour notre [Centre de transparence](#), lequel fournit un niveau de détails élevé sur tous les domaines de nos politiques ainsi que nos mesures d'application. En ce qui concerne les contenus identifiés comme contenant des fausses informations sur Twitter, nous avons pris des mesures d'application fondées sur nos politiques en matière d'intégrité civique et d'informations trompeuses sur la COVID-19. Ainsi, à l'échelle mondiale, de janvier à juin 2020, nous avons pris des mesures à l'encontre de 2 351 comptes pour violation de notre politique en matière d'intégrité civique (soit une augmentation de 37 % par rapport à la période de référence précédente) et nous avons supprimé 2 710 contenus individuels. Nous avons également suspendu ou exigé des retraits de contenu de la part de 4 658 comptes dans le monde pour violation de notre politique d'informations trompeuses sur la COVID-19 au cours de cette même période (veuillez noter qu'il n'est pas possible de comparer les changements avec la période de référence précédente, étant donné que cette politique était appliquée pour la première fois chez Twitter). Ces chiffres doivent être analysés dans un contexte plus large où plusieurs centaines de millions de Tweets sont partagés chaque jour sur Twitter à travers le monde, soit plusieurs dizaines de milliards de Tweets au cours de cette période de référence de 6 mois. En outre, ce chiffre ne tient pas compte des autres mesures d'application que nous pouvons prendre, notamment l'apposition d'un libellé ou d'un message d'avertissement sur un compte ou un contenu.

**En règle générale, l'opérateur a-t-il accru ou renforcé ses moyens de lutte contre les fausses informations en 2020 ? Il pourra résumer les éventuelles évolutions de sa politique en la matière et les enseignements qu'il en tire, et fournira des éléments quantitatifs témoignant de ces renforcements et de leurs effets. Il distinguera, en sus, les nouveaux dispositifs, humains ou techniques qu'il aurait mis en œuvre en 2020 et qui auraient contribué à cette lutte.**

Alors que le monde entier faisait face - et continue de faire face - à une urgence de santé publique sans précédent, nous avons fait preuve de transparence quant aux mesures d'urgence que nous avons mises en place dès le début de la crise pour servir la conversation publique en ce moment critique. Cette approche s'est appuyée sur le travail précieux de nos partenaires de confiance, notamment les autorités de santé publique, les organisations non gouvernementales et les gouvernements du monde entier.

Ainsi, nos mesures visant à protéger la conversation publique se sont concentrées sur deux piliers majeurs : 1) l'adaptation de nos politiques pour faire face aux menaces émergentes, notamment en ce qui concerne la COVID-19 et les élections ; 2) le maintien de notre service existant, la poursuite de l'application de nos règles et la protection de la plateforme dans le contexte de la transition de nos équipes vers le télétravail.

Tout d'abord, dès mars 2020, nous avons élargi notre définition de la notion de préjudice pour prendre en compte les contenus qui vont directement à l'encontre des recommandations de sources faisant autorité en matière de santé publique mondiale et locale. Par conséquent, il est désormais interdit d'utiliser les services de Twitter pour partager des informations fausses ou trompeuses sur la COVID-19, susceptibles de causer un préjudice. Alors même que la compréhension scientifique de la pandémie de COVID-19 continue de se développer, nous avons observé l'émergence de théories conspirationnistes persistantes, de discours alarmistes non fondés sur des recherches ou des rapports crédibles, et d'un large éventail de rumeurs non fondées, qui, si elles ne sont pas contextualisées, peuvent empêcher le public de prendre des décisions éclairées concernant des enjeux sanitaires, et mettent en danger les individus, les familles et les communautés. Ainsi, tout contenu qui



est manifestement faux ou trompeur et qui peut entraîner un risque important de préjudice (comme une exposition accrue au virus ou des effets néfastes sur les systèmes de santé publique) ne peut être partagé sur Twitter. Cela inclut le partage de contenus susceptibles d'induire les gens en erreur sur la nature de la COVID-19, l'efficacité et/ou la sécurité des mesures préventives, des traitements ou d'autres précautions pour atténuer ou traiter la maladie, les réglementations, restrictions ou exemptions officielles relatives aux avis sanitaires, la prévalence du virus ou le risque d'infection ou de décès associé à la COVID-19. En outre, en mai 2020, nous avons annoncé qu'en fonction de la gravité potentielle du préjudice et du type d'informations trompeuses, nous pouvons désormais libeller les Tweets qui partagent des informations trompeuses sur la COVID-19 et ce, afin de réduire leur viralité et fournir un contexte supplémentaire. Ces avertissements informent les individus - avant même qu'ils ne consultent le Tweet - que le contenu qu'ils s'approprient à voir est en contradiction avec les conseils d'experts en santé publique. Ces libellés renvoient à une page créée par Twitter ou à une source externe de confiance contenant des informations supplémentaires sur les affirmations contenues dans le Tweet. En effet, plutôt que de supprimer davantage de contenu, nous visons à fournir un contexte supplémentaire sous la forme de libellés et d'avertissements, qui peuvent limiter davantage la visibilité ou les possibilités d'interaction avec ces Tweets. Enfin, en décembre 2020, alors que le monde se préparait au déploiement des vaccins, nous avons étendu notre politique de désinformation aux vaccins contre la COVID-19. Ainsi, nous exigeons le retrait de Tweets qui avancent des affirmations fausses ou trompeuses et nuisibles sur les vaccins contre la COVID-19, y compris :

- Les fausses affirmations qui suggèrent que la vaccination et les vaccins sont utilisés pour causer intentionnellement des dommages aux populations ou les contrôler, y compris les déclarations sur les vaccins qui invoquent une conspiration délibérée ;
- Les affirmations fausses qui ont été largement démenties concernant les effets ou impacts négatifs de la vaccination ;
- Les affirmations fausses selon lesquelles la COVID-19 n'est pas réelle ou grave, et donc que la vaccination est inutile.

Nous pouvons libeller ou placer un message d'avertissement sur les Tweets qui avancent des rumeurs non fondées, des affirmations contestées, ainsi que des informations incomplètes ou hors contexte sur les vaccins. Les Tweets libellés dans le cadre de cette politique élargie peuvent contenir un lien vers des informations de santé publique faisant autorité ou vers les règles de Twitter, afin de fournir aux gens un contexte supplémentaire et des informations faisant autorité sur la COVID-19.

D'autre part, en plus de ces nouvelles politiques sur la COVID-19, nous nous sommes attachés à mettre à jour et à étendre notre politique en matière d'intégrité civique, notamment dans le contexte crucial des élections américaines de 2020. En effet, la conversation qui se déroule sur Twitter n'est jamais aussi importante que pendant les élections. C'est sur Twitter que les gens viennent entendre directement les élus et les candidats aux élections, qu'ils découvrent les dernières informations et, de plus en plus, qu'ils s'informent au sujet de la tenue pratique du scrutin. Alors que de plus en plus de personnes cherchent de nouveaux moyens de voter et d'exprimer leurs droits civiques fondamentaux en toute sécurité pendant la pandémie de COVID-19, nous avons observé que le besoin de ce type d'informations a crû significativement. Notre politique en matière d'intégrité civique existante cible les types de contenus les plus directement nuisibles, à savoir ceux liés aux :

- Fausses informations ou déclarations sur la manière de participer au scrutin
- Contenu susceptible d'intimider ou de supprimer la participation
- Fausse affiliation

Toutefois, compte tenu de la situation particulière dans laquelle les scrutins allaient se dérouler en 2020, et conformément à notre engagement à protéger l'intégrité de la conversation électorale, nous avons élargi ce cadre existant. Nous libellons ou supprimons désormais les informations fausses ou trompeuses destinées à saper la confiance du public dans une élection ou tout autre processus civique. Cela inclut, sans s'y limiter, les éléments suivants :

- Les informations fausses ou trompeuses qui sèment la confusion sur les lois et réglementations liées à un processus civique, ou sur les fonctionnaires et institutions qui exécutent ces processus civiques.
- Les affirmations contestées qui pourraient ébranler la confiance dans le processus lui-même, par exemple, des informations non vérifiées sur la fraude électorale, la falsification des bulletins de vote, le décompte des voix ou la certification des résultats des élections.
- Les affirmations trompeuses sur les résultats ou l'issue d'un processus civique qui appellent ou pourraient conduire à une interférence avec la mise en œuvre des résultats du processus, par exemple



revendiquer la victoire avant que les résultats des élections n'aient été certifiés, inciter à une conduite illégale pour empêcher un transfert pacifique du pouvoir ou une succession ordonnée.

- Le fait pour toute personne sur Twitter, y compris des candidats à un poste, de revendiquer la victoire d'une élection avant que celle-ci ne soit officiellement proclamée. Afin de déterminer les résultats d'une élection dans le contexte des élections américaines de 2020, notre approche expérimentale consistait à exiger : 1) soit une annonce de la part des responsables électoraux d'un État (tels que déterminés par la *National Association of Secretaries of State* et la *National Association of State Election Directors*) ; 2) soit une projection électorale publique d'au moins deux organes d'information nationaux faisant autorité et procédant à une couverture électorale indépendante.

Conformément à notre approche existante de mise en application de cette politique, les Tweets qui sont libellés en vertu de cette politique élargie ont une visibilité réduite sur Twitter, ce qui signifie que nous ne les amplifions pas sur la plateforme. Toutefois, les personnes qui suivent le compte visé sont toujours en mesure de voir le Tweet et de le retweeter.

Deuxièmement, avec la transition obligatoire vers le télétravail dans la plupart des pays où nos équipes sont basées et opèrent, nous avons dû faire pivoter sur le plan opérationnel nos principaux efforts pour assurer la sécurité des personnes sur Twitter. C'est pourquoi nous avons accru notre utilisation du machine learning et de l'automatisation pour prendre un large éventail de mesures sur les contenus potentiellement abusifs et trompeurs, y compris les fausses informations. Voici quelques-unes des actions et moyens supplémentaires que nous avons mis en place pour protéger la conversation publique dans ce moment critique :

- Nous avons construit des systèmes qui permettent à nos équipes de continuer à faire respecter nos règles à distance, dans le monde entier.
- Nous avons renforcé l'assistance aux employés et le soutien au bien-être de tous ceux qui participent à ce travail essentiel, et avons veillé à ce que la confidentialité et la sécurité des personnes restent une priorité absolue.
- Nous avons mis en place un système mondial de triage de la gravité du contenu afin de donner la priorité aux violations potentielles des règles qui présentent le plus grand risque de préjudice et ce, afin de réduire la charge qui pèsent sur les victimes et les personnes qui signalent des contenus.
- Nous effectuons des contrôles quotidiens de qualité de nos processus d'application sur les contenus afin de nous assurer que nous sommes toujours en mesure de répondre à cette épidémie mondiale qui évolue rapidement.
- Nous nous engageons auprès de nos partenaires dans le monde entier pour que les canaux de signalement restent toujours disponibles et que les cas les plus urgents puissent être portés à notre attention dans les meilleurs délais.

**L'opérateur est invité à partager la liste des 20 contenus relevant d'une communication commerciale (publicitaires, sponsorisés et en partenariat) et celle des 20 contenus n'en relevant pas détectés comme contenant une fausse information et qui ont suscité le plus d'interactions de la part des utilisateurs en France en 2020 sur le service qu'il opère. Il indiquera pourquoi ces contenus ont été qualifiés ainsi et quelques indicateurs relatifs à ces interactions. Les titres ou thématiques abordées par ces contenus, leur date de première apparition ainsi que les liens permettant au régulateur de les consulter seront précisés à chaque fois. L'opérateur pourra également indiquer si les tendances observées dans ces deux listes diffèrent sensiblement de ses observations lors des années précédentes.**

La transparence est au cœur du travail que nous faisons chez Twitter. Notre [Centre de transparence](#) nouvellement mis à jour fournit un niveau élevé de détails dans tous les domaines de nos politiques et de nos mesures d'application. Toutefois, pour un certain nombre de raisons, notamment relatives à la protection de la vie privée et des données de nos utilisateurs, ainsi que des difficultés techniques, nous ne sommes pas en mesure de conserver et de divulguer le type d'informations demandé par le Conseil dans cette question.

### ***Respect de la liberté d'expression***

**En se fondant sur des cas de fausses informations regardées comme telles par ses équipes au cours de l'année passée, l'opérateur est invité à exposer comment sa politique de modération s'articule avec la liberté d'expression et de communication sur son ou ses services.**



La mission de Twitter est de permettre aux gens de créer, de partager des idées et des informations, et d'exprimer leurs opinions et leurs croyances sans aucune barrière. La liberté d'expression fait partie intégrante des droits de l'Homme. Chacun a sa propre voix et le droit de l'exprimer. Notre rôle est donc de servir la conversation publique, ce qui nécessite l'expression d'un large éventail de points de vue et d'opinions. Nous voulons donner aux gens les moyens de voir toutes les facettes d'une question donnée et encourageons les discussions ouvertes autour d'opinions et de perspectives divergentes. Cette approche permet la coexistence de nombreuses formes de discours sur notre plateforme et favorise en particulier le contre-discours : discours présentant des faits pour corriger des déclarations inexactes ou des perceptions erronées, soulignant l'hypocrisie ou les contradictions de certains acteurs, mettant en garde contre les conséquences hors ligne ou en ligne de certains discours, dénonçant les discours haineux ou dangereux, ou contribuant à changer les opinions et à désamorcer des situations.

### *Mesures futures*

**L'opérateur est invité à faire part des actions qu'il envisage de mener à l'avenir afin de lutter contre la diffusion de fausses informations sur son service.**

Nous nous efforçons de servir la conversation publique, et de faire notre travail ouvertement avec les personnes qui utilisent notre service au quotidien et nos partenaires mondiaux dans le cadre de notre [Conseil de confiance et de sécurité](#). Par conséquent, nous nous engageons à réviser et à mettre à jour constamment nos règles - comme nous l'avons fait lors de la crise sanitaire du COVID-19 et des élections américaines de 2020 - pour mieux refléter et combattre les nouvelles menaces de désinformation émergentes et garder une longueur d'avance sur les tactiques employées par les acteurs malveillants qui cherchent à manipuler notre plateforme. Nous investissons également massivement dans la technologie pour mieux faire appliquer nos politiques de manière cohérente dans le monde entier.



## 1. Dispositif de signalement de fausses informations

### *Description du dispositif*

**1.1 L'opérateur indiquera si son service dispose d'un dispositif de signalement des contenus publiés et si ce dispositif permet à l'utilisateur de signaler de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité du scrutin. Il indiquera le libellé choisi pour ce motif et les raisons qui sous-tendent le choix de ce libellé. Il pourra illustrer sa réponse d'exemples.**

Twitter offre la possibilité de signaler des Tweets, comptes et messages privés qui comportent notamment du spam, du contenu abusif ou nuisible, de l'usurpation d'identité, des violations de droits d'auteur ou de marque. En outre, les utilisateurs peuvent signaler tout Tweet qu'ils voient sur Twitter, y compris les Tweets sur le fil d'actualité, dans l'onglet Notifications ou dans la barre de recherche Twitter. Le signalement d'un Tweet permet aux personnes d'indiquer le Tweet spécifique qui, selon elles, est en infraction aux règles de Twitter. Le mécanisme de signalement est directement accessible au niveau du [Tweet](#), au niveau du compte et depuis la page du [Centre d'aide](#).

En ce qui concerne les fausses informations, Twitter prévoit deux canaux de signalement distincts pour les utilisateurs français :

1. Les fausses informations sur des élections ou l'inscription sur les listes électorales.

Les Règles de Twitter interdisent la publication de contenus présentant de fausses informations sur le vote ou l'inscription sur les listes électorales. Si un utilisateur signale ce type de contenu, nous examinerons systématiquement le Tweet signalé. Si nous déterminons que le Tweet est en violation de nos politiques, nous prendrons des mesures (allant de la demande de suppression du contenu en question à la suspension permanente du compte). L'auteur du signalement recevra un suivi de la part Twitter si nous avons besoin de plus d'informations de sa part ou si nous prenons des mesures concernant le Tweet signalé.

2. Les fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un des scrutins.

La loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information exige que nous fournissions également un moyen pour les personnes de signaler les fausses informations susceptibles d'altérer la sincérité d'un scrutin ou de troubler l'ordre public. Ces signalements fournissent des informations utiles pour guider l'application de nos Règles relatives aux fausses informations en France. Nous les utilisons pour informer la manière dont nous défendons notre plateforme contre les tentatives de manipulation. Toutefois, nous n'examinons pas ces signalements spécifiques et ne prenons pas de mesures à leur égard sur une base individuelle.

**Le Conseil souhaite connaître le nombre total de motifs de signalement proposés à l'utilisateur, et si ces motifs varient selon le type de contenus (texte, vidéo, commentaire, compte). Il souhaite également connaître leur arborescence et s'il existe une possibilité de saisir un motif autre que ceux proposés.**

Toute personne peut signaler directement des Tweets, comptes ou messages privés, avec différents motifs de signalement pour chaque type de contenu. Pour les Tweets, il existe 6 motifs de signalement (peu importe que le Tweet comporte du texte, une photo, une vidéo ou une combinaison de ces éléments), à savoir : "Ce Tweet ne m'intéresse pas", "Il est suspect ou constitue un spam", "Il contient une photo ou une vidéo sensible", "Les propos tenus sont inappropriés ou dangereux", "Il induit en erreur

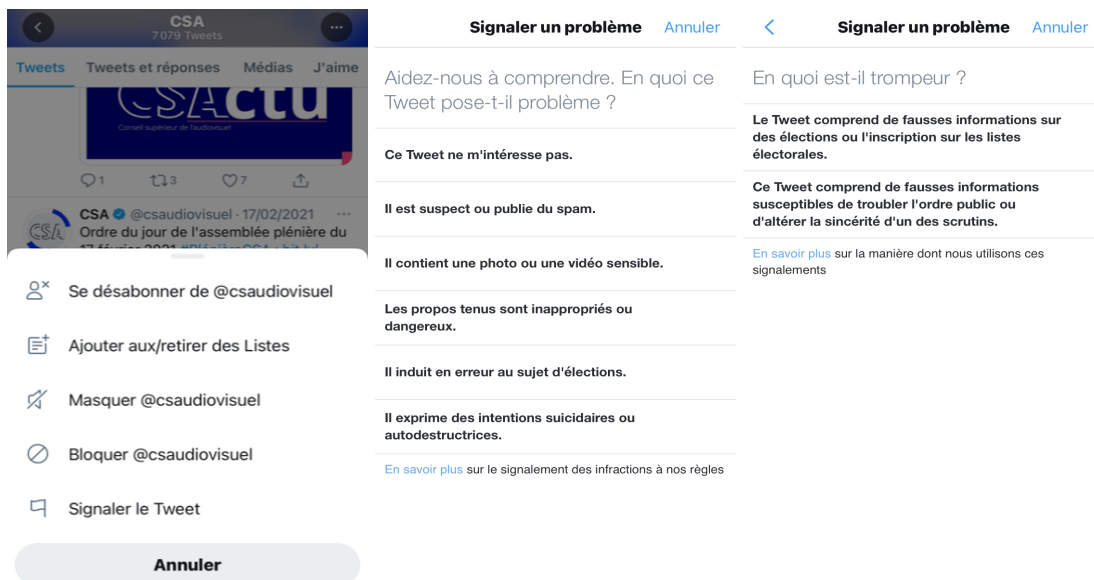


au sujet d'élections", "Il exprime des intentions suicidaires ou autodestructrices". Au niveau du compte, les internautes ont le choix entre sept motifs de signalement, dont "Ce compte ne m'intéresse pas ", "Il est suspect ou publie un spam", "Le compte semble avoir été piraté", "Il se fait passer pour moi ou pour quelqu'un d'autre", "Ses informations et/ou ses images comportent un contenu inapproprié ou haineux", "Ses Tweets sont inappropriés ou haineux", "Il exprime des intentions suicidaires ou autodestructrices". Pour les messages privés, les personnes peuvent signaler un message individuel ou une conversation entière. Il existe deux motifs pour le faire : "Il s'agit d'un spam" et "Les propos tenus sont inappropriés ou dangereux".

## 1.2 Le Conseil souhaite que l'opérateur lui fournisse des informations sur la visibilité, l'accessibilité et l'ergonomie de ce dispositif, et notamment :

- sa présentation visuelle ;
- son affichage (visibilité par défaut ou non, visibilité par tous les utilisateurs ou seulement par les utilisateurs titulaires d'un compte) ;
- son emplacement à proximité immédiate du contenu ou non ;
- le nombre d'actions nécessaires pour accéder au dispositif de signalement général, et à celui de signalement d'une fausse information ;
- le nombre d'actions et d'hyperliens nécessaires à l'envoi du signalement d'une fausse information depuis la page du contenu et, le cas échéant, les efforts déployés par l'opérateur pour réduire de façon pertinente le nombre d'actions nécessaires pour l'envoi du signalement ;
- l'existence d'un dispositif de signalement facilement accessible adapté notamment aux personnes en situation de handicap.

Twitter propose un mécanisme de signalement directement accessible au niveau du Tweet en 1 clic (dans le coin supérieur droit). Le bouton dédié au signalement est clair : "Signaler ce Tweet" et identifié par un drapeau. Le signalement peut être complété en 5 étapes, dont une proposant d'ajouter un commentaire supplémentaire et la possibilité d'ajouter facilement des Tweets supplémentaires au signalement.





**Signaler un problème** Annuler

Ajoutez jusqu'à 5 Tweets à ce signalement.

Le signalement simultané de plusieurs Tweets nous aide à résoudre les problèmes plus rapidement.

Les communications sur ce signalement peuvent reprendre ces Tweets. [En savoir plus](#)

**CSA @csaudiovisuel**  
Ordre du jour de l'assemblée plénière du 17 février 2021 #PlénièreCSA :  
<https://t.co/dIM1h9PnpA>  
<https://t.co/CuY6tfwKr3>

**remettons de la diversité ette contre les discrimine**  
#JournéeZéroDiscrimination | Retrouvez les ressources pédagogiques #EMI du @CSAudiovisuel sur la lutte contre les discriminations et la juste représentation de la

Passer Ajouter 1

**Signaler un problème** OK

**Merci de nous avoir informés.**

Si nous constatons que ce compte agit en violation des [Règles de Twitter](#) alors nous ferons le nécessaire.

En attendant, vous pouvez [améliorer votre expérience](#) Twitter en utilisant les options suivantes :

**Bloquer @csaudiovisuel**

Quand vous bloquez un compte, il ne peut pas vous suivre ni voir vos Tweets. Par ailleurs, vous ne voyez pas ses Tweets ni ses notifications.

**Masquer @csaudiovisuel**

Vous pouvez aussi masquer un compte. Si vous ne le suivez pas, vous ne verrez plus ses Tweets, ni ses Retweets ni ses notifications. Si vous le suivez

Nous sommes presque prêts à envoyer votre signalement relatif à @csaudiovisuel.

Si vous le souhaitez, vous pouvez nous en dire plus sur ce qui s'est passé.

**Envoyer le signalement à Twitter**

### 1.2.1 L'opérateur précisera si le dispositif décrit est commun à toutes les déclinaisons du service (site internet, application) et à tous les types de contenus (texte, vidéo, image, commentaires, contenus sponsorisés, compte utilisateur) accessibles sur le site.

Twitter s'efforce de fournir un environnement dans lequel les gens peuvent se sentir libres de s'exprimer. À cet effet, nous voulons permettre aux gens de signaler facilement les contenus qui, selon eux, pourraient enfreindre nos règles. À ce titre, le mécanisme de signalement relatif aux fausses informations est commun à toutes les versions du service Twitter, même s'il peut y avoir de légères variations visuelles (par exemple, des icônes) en fonction du système d'exploitation (IOS ou Android) et du type d'appareil (desktop vs. mobile).

### 1.2.2 Il mentionnera les fonctionnalités supplémentaires, développées ou en cours de développement, permettant d'étayer un signalement (possibilité de signaler plusieurs contenus conjointement, d'associer le signalement d'explications, d'une pièce jointe ou d'un lien, etc.).

Après avoir accédé aux différents canaux de signalement, et afin de finaliser ce dernier, les personnes devront identifier au sein d'une liste la raison de leur(s) signalement(s). Elles pourront ensuite fournir des informations supplémentaires sur le problème qu'elles signalent. Nous pouvons également leur demander de sélectionner d'autres Tweets du compte qu'elles signalent afin que nous disposions d'un meilleur contexte pour évaluer le signalement.

Nous incluons le texte des Tweets signalés dans nos courriels de suivi et nos notifications. Une fois le signalement soumis, nous fournissons des recommandations sur les mesures supplémentaires que toute personne peut prendre pour améliorer son expérience, comme par exemple, la mise en sourdine ou le blocage du compte qui a publié le contenu signalé.

### 1.2.3 L'opérateur indiquera au Conseil les éventuels retours des utilisateurs concernant l'accessibilité, la visibilité et l'ergonomie du dispositif de signalement déployé sur son service.

La sérénité de la conversation publique est la priorité absolue de Twitter. Pour atteindre cet objectif, les signalements de la part des personnes qui utilisent notre service sont très importants pour nous. Il est essentiel que nos outils de signalement soient facilement accessibles et largement compris par les

personnes qui utilisent notre service afin d'y répondre au mieux, le plus rapidement et de manière aussi transparente que possible. Au fil des ans, nous avons travaillé pour que notre mécanisme de signalement soit facile à utiliser et à comprendre. Twitter s'engage à élaborer ses produits et ses politiques de manière ouverte, en tenant compte des commentaires de toutes les personnes, communautés et organisations spécialisées impliquées.

### **Procédure de traitement des signalements**

**1.3 L'opérateur décrira l'ensemble de la procédure de traitement du signalement d'une fausse information propre à garantir sa pleine effectivité tel que prévu à l'article 11 de la loi du 22 décembre 2018.**

**1.3.1 Il indiquera les conditions de mise en instruction d'un signalement, en précisant notamment si l'instruction des signalements est systématique ou non. Le cas échéant, il portera à la connaissance du Conseil les critères déclenchant cette mise en instruction. Il précisera les conditions d'élaboration de ces critères (par qui, quand et comment sont-ils établis ou révisés, les objectifs qu'ils poursuivent) et leurs adaptations éventuelles au cours de l'année 2020.**

Comme indiqué en réponse à la question 1.1, Twitter prévoit deux canaux de signalement distincts pour les utilisateurs français concernant les fausses informations. Chaque signalement envoyé par nos utilisateurs sera systématiquement examiné et fera l'objet d'une réponse en temps utile.

1. Les fausses informations sur des élections ou l'inscription sur les listes électorales.

Les Règles Twitter interdisent la publication de contenu fournissant de fausses informations sur le vote ou l'inscription sur les listes électorales. Si une personne signale ce type de contenu, nous examinerons le Tweet signalé. Si nous déterminons que le Tweet est en violation de nos politiques, nous prendrons des mesures (allant de la demande de suppression du contenu en question à la suspension permanente du compte).

2. Les fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un des scrutins.

La loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information exige que nous fournissions également un moyen pour les personnes de signaler les fausses informations susceptibles d'altérer la sincérité d'un scrutin ou de troubler l'ordre public.

**1.3.2 Il détaillera les moyens déployés pour l'instruction, et notamment :**

- les moyens humains, en précisant le nombre de personnes en charge de cette activité, leur statut (collaborateur ou prestataire de la société) leur formation (initiale et continue), leur localisation géographique, leur langue de travail ainsi que les modalités d'organisation de leur travail (notamment si elles impliquent une astreinte) ;
- les sources dont disposent les modérateurs afin d'effectuer le contrôle des contenus signalés, en précisant leur nature (juridique, journalistique, académique) et l'articulation de ces sources entre elles ;
- l'articulation entre l'intervention humaine et l'intervention automatique ; en particulier, une intervention humaine est-elle systématique dans l'instruction et, le cas échéant, dans la décision d'intervenir à l'encontre du contenu ou du compte ;
- les dispositifs spécifiques mis en place afin de garantir une modération efficace lors de situations de crise (formation des équipes de modération aux sujets d'actualité, moyens automatisés renforcés, partenariats exceptionnels avec les *fact-checkers*...).



Nos règles sont là pour garantir que tout le monde puisse participer à la conversation publique librement et en toute sécurité, à condition de respecter les conditions d'utilisation de Twitter. Lorsqu'il s'agit de faire respecter nos politiques, nous disposons d'équipes internationales qui travaillent 24 heures sur 24 pour examiner les signalements et aider à faire respecter nos règles de manière cohérente. Ces équipes utilisent également la technologie pour aider à déceler les comportements susceptibles de nuire à la conversation publique.

En tant que service ouvert avec des centaines de millions de Tweets partagés chaque jour, la technologie est essentielle pour notre capacité à répondre à l'échelle et à améliorer l'expérience de chacun sur le service, et nous investissons en conséquence. C'est bien la technologie qui a le plus grand impact sur Twitter.

L'une des caractéristiques sous-jacentes de notre approche est que nous adoptons une approche axée sur le comportement. En d'autres termes, nous examinons le comportement des comptes avant de nous intéresser au contenu qu'ils publient. C'est ainsi que nous sommes en mesure de déployer nos efforts à l'échelle mondiale. Nous utilisons une technologie de détection de contenu étendue pour identifier et actionner des contenus nuisibles et abusifs intégrés dans diverses formes de médias sur la plateforme.

Nous avons toujours utilisé une combinaison de machine learning et d'examen humain. Nos systèmes sont en mesure de présenter le contenu aux modérateurs humains qui utilisent un contexte important pour prendre des décisions sur les violations potentielles de nos Règles. Les modérateurs de contenu de Twitter sont continuellement formés sur la manière d'appliquer nos Règles de manière cohérente et impartiale, ce qui garantit une prise de décision objective et judicieuse. Nous disposons également d'une [procédure d'appel](#) pour toute erreur potentielle qui pourrait se produire.

Nous investissons dans la technologie afin d'améliorer l'expérience de chacun sur la plateforme en rendant moins visibles les contenus qui détournent la conversation publique (tels que les spams et les abus). L'objectif est de réduire la charge de travail du public et des victimes en matière de signalement. Nos investissements dans les technologies proactives continuent de réduire la charge de travail pour les personnes qui signalent des abus à Twitter en faisant apparaître davantage de contenus à examiner pour nos modérateurs.

En ce qui concerne les abus, cette stratégie à deux volets nous permet aujourd'hui d'identifier de manière proactive plus d'un Tweet sur deux sur lequel nous prenons des mesures, plutôt que de s'appuyer uniquement sur les signalements faits à Twitter. Ce chiffre était d'environ un sur cinq en 2018, ce qui montre une forte réduction de la charge de signalement pour les victimes.

La technologie joue un rôle déterminant dans des domaines clés tels que le terrorisme, la pédopornographie ou les tentatives de manipulation de la plateforme, par exemple :

- Plus de 90% des contenus terroristes et pédopornographiques sont désormais supprimés de manière proactive à l'aide de notre technologie interne construite à cet effet, et la majorité des comptes sont suspendus avant leur premier Tweet. Les signalements des utilisateurs ne représentent qu'une infime partie de ces suppressions. Les demandes de retrait de contenu de la part des forces de l'ordre est encore plus faible - moins 1%.
- Notre équipe dédiée à "l'intégrité" de la plateforme, en partenariat avec un large éventail d'organisations et de personnes engagées au sein de Twitter, continue à investir massivement pour identifier et combattre les acteurs malveillants qui tentent de manipuler notre service.
- Nous luttons contre les tentatives de manipulation de la plateforme (y compris le spam et l'automatisation malveillante) de manière stratégique et à grande échelle. En tant que défi mondial, la technologie est au cœur de notre capacité de réponse et nous aide à déceler les utilisations abusives de notre service avant même qu'elles n'aient un impact sur l'expérience de chacun.



La technologie ne sert pas seulement à supprimer du contenu. Elle est essentielle pour nous aider à identifier les problèmes, à hiérarchiser les signalements, à prendre plusieurs mesures en même temps, à trouver des contenus qui n'ont pas été signalés. Nos agents peuvent en faire plus à mesure que leur technologie et leurs outils s'améliorent, et nous investissons chaque jour dans ce domaine.

Nous ne résoudrons pas le défi de la modération à grande échelle avec uniquement plus de moyens humains. Nous avons constaté que nous sommes beaucoup plus efficaces dans la lutte contre les contenus préjudiciables en utilisant davantage la technologie et en augmentant proportionnellement nos équipes.

En matière de modération, nous ne sommes pas axés sur les personnes, mais sur la technologie. C'est ce qui fonctionne, ce qui est évolutif et ce qui est humain compte tenu des types de contenu auxquels ces équipes peuvent être amenées à être confrontées. Nous avons récemment doublé le nombre de personnes chargées de faire respecter nos règles. Chez Twitter, 1 867 personnes se consacrent exclusivement à l'application de nos politiques et à la modération des contenus. Ce chiffre représente plus d'un tiers de l'ensemble de nos effectifs mondiaux. Ce nombre s'ajoute bien sûr aux ingénieurs et aux équipes produits qui travaillent à l'amélioration de la sérénité de la conversation sur Twitter et, bien sûr, au renforcement de notre capacité de machine learning.

Tout au long de la phase de formation initiale, nos modérateurs de contenu doivent passer plusieurs examens de certification avec un système de notation strict. Une partie de ce processus d'examen permet de vérifier la qualité et la précision de la modération, et nous aide à identifier tout risque de partialité et/ou de faux positifs. Des formations régulières sont également dispensées et des contrôles ponctuels sont effectués afin de garantir un niveau élevé de modération à tout moment. En particulier, étant donné que 79% de nos utilisateurs se trouvent en dehors des États-Unis, une grande partie de la formation porte sur la sensibilité aux différences culturelles et linguistiques à travers le monde.

Le bien-être des personnes qui examinent le contenu est une préoccupation majeure pour nos équipes et notre priorité absolue est de veiller à ce que notre personnel et nos partenaires soient traités avec empathie, attention et respect. Nous évaluons en permanence les normes de nos partenaires et restons déterminés à protéger le bien-être des équipes chargées de ce rôle important. Nous disposons d'une gamme complète de services de soutien pour nos employés, y compris nos modérateurs de contenu. Dans le cadre de notre travail avec des tiers, nous exigeons dans nos contrats la fourniture de services de soutien, y compris durant toute une période de temps après qu'une personne a changé de rôle.

À long terme, l'un des investissements les plus précieux que nous puissions faire est dans la technologie. Plus nous pouvons exploiter la technologie pour minimiser l'exposition au contenu, moins nos employés et nos sous-traitants seront en contact avec celui-ci. Nous continuerons à soutenir ceux qui occupent ces fonctions clés.

**1.3.3 L'opérateur indiquera s'il a mis en place des moyens humains et/ou automatisés de détection proactive des contenus diffusant de fausses informations. Il indiquera également si le processus d'instruction des contenus détectés est en tout point identique à celui des contenus signalés par les utilisateurs ou sur quels points il diffère le cas échéant.**

Servir la conversation publique reste notre mission principale, et nous continuons à investir et à travailler pour construire des outils et offrir du contexte afin que les gens puissent trouver des informations fiables et crédibles sur Twitter.

Des centaines de millions de Tweets sont partagés chaque jour. Notre objectif est donc de supprimer le contenu problématique à grande échelle aussi rapidement que possible sans faire peser une charge de signalement trop lourde sur les victimes. À cette fin, la technologie est essentielle pour nous aider à



améliorer l'expérience de chacun. Aujourd'hui, plus de la moitié des Tweets sur lesquels nous intervenons sont décelés de manière proactive par notre technologie pour un examen humain. Cette technologie utilise une approche comportementale pour identifier les modèles de Tweets potentiellement problématiques.

En ce qui concerne les fausses informations sur la COVID-19, nos équipes utilisent et améliorent nos systèmes internes pour faire apparaître de manière proactive les contenus liés à la COVID-19. Ces systèmes permettent de s'assurer que nous n'amplifions pas les Tweets comportant ces avertissements ou libellés et de déceler rapidement les contenus à forte visibilité. En outre, nous continuons à nous appuyer sur des partenaires de confiance pour identifier les contenus susceptibles de causer des dommages hors ligne. Compte tenu de la situation dynamique, nous accordons la priorité à l'examen et à la labellisation des contenus susceptibles d'entraîner une exposition ou une viralité accrue.

Nous avons beaucoup appris en déployant ces nouveaux libellés, et nous restons ouverts aux ajustements pour explorer la labellisation de différents types d'informations trompeuses. Ce processus est en cours et nous nous efforcerons de faire en sorte que ces libellés et avertissements, ainsi que d'autres, apparaissent sur Twitter.

**1.3.4 L'opérateur présentera également les différentes étapes de l'instruction (décision de mise en instruction d'un signalement, examen du contenu, qualification du contenu, décision prise sur un contenu) et il indiquera s'il existe une procédure d'instruction spécifique en période électorale ou de situation de crise (ex : crise sanitaire).**

Tout d'abord, un cadre d'analyse axé uniquement sur le contenu se révèle trop simpliste pour aborder les nombreux défis liés aux fausses informations. Il ignore à la fois les acteurs et le comportement derrière les réseaux d'acteurs qui peuvent tenter de disséminer en masse des fausses informations. Ainsi, Twitter peut mener une série d'enquêtes par des moyens humains et automatisés, le tout dans un cadre d'analyse axé sur l'acteur, le comportement et le contenu. Il existe un certain nombre de détections automatisées qui se déclenchent sur la base de comportements inauthentiques en réseaux qui cherchent à promouvoir des fausses informations, en violation de notre politique en matière de manipulation de la plateforme et de spam. Les équipes d'investigation avancées qui recherchent ces acteurs et ces menaces peuvent supprimer et divulguer des opérations de désinformation soutenues par des États qui cherchent à créer des récits trompeurs par la désinformation. Toutefois, elles peuvent décider de laisser le même contenu affiché (ou prendre des mesures moindres) sur un citoyen qui partage sans intention la même information. Par conséquent, nous essayons non seulement d'évaluer le contenu partagé, mais aussi, dans la mesure du possible, le contexte dans lequel il est partagé, en tenant compte de l'acteur et du comportement derrière le contenu lors de la prise de décision, ainsi que de l'évaluation critique des dommages potentiels. Twitter s'appuie sur des pipelines de détection automatisés et de machine learning pour soutenir un effort d'examen humain des informations fausses, trompeuses et nuisibles (telles que déterminées par notre cadre d'analyse en matière de potentiel de préjudice). Au cours de l'examen du contenu, nous pouvons prendre diverses mesures en fonction du risque de préjudice pour les utilisateurs, notamment la labellisation et le retrait du contenu.

**1.3.5 Concernant l'issue de l'instruction, l'opérateur présentera également de façon détaillée :**

- **la liste des actions qu'il peut mettre en œuvre vis-à-vis des contenus signalés dont il a considéré qu'ils contenaient de fausses informations, ainsi que des comptes sur lesquels ils ont été publiés ;**
- **les critères amenant à décider de telle ou telle action ;**
- **les modalités de mise en œuvre de la décision.**

Comme indiqué précédemment, nous disposons d'une série d'options d'application allant au-delà de la simple suppression de contenu, notamment des libellés et des avertissements. Plus précisément, ces



options alternatives s'appliquent à trois domaines majeurs des Règles de Twitter : 1) notre politique de désinformation sur la COVID-19; 2) notre politique en matière d'intégrité civique et 3) notre politique relatives aux médias synthétiques et manipulés.

Premièrement, en ce qui concerne notre politique de désinformation sur la COVID-19, nous avons clairement indiqué qu'en fonction de la propension à nuire et du type d'informations trompeuses, nous pouvons appliquer des avertissements à un Tweet. Ces avertissements informent les internautes que les informations contenues dans le Tweet sont en contradiction avec les recommandations des experts en santé publique. Bien que le contenu faux ou trompeur puisse prendre de nombreuses formes différentes, nous prendrons des mesures en fonction de trois grandes catégories de concepts :

- Informations trompeuses - déclarations ou affirmations dont le caractère faux ou trompeur a été confirmé par des experts en la matière, tels que les autorités de santé publique.
- Affirmations contestées - déclarations ou affirmations dont l'exactitude, la véracité ou la crédibilité sont contestées ou encore inconnues à ce jour.
- Affirmations non vérifiées - informations (qui peuvent être vraies ou fausses) qui ne sont pas confirmées au moment où elles sont partagées.



<b>Misleading Information</b>	Label	Removal
<b>Disputed Claim</b>	Label	Warning
<b>Unverified Claim</b>	No action	No action*
	Moderate	Severe
<b>Propensity for Harm</b>		

\*Nous continuerons à introduire de nouveaux libellés pour fournir du contexte autour des différents types d'affirmations et de rumeurs non vérifiées, selon les besoins.

Deuxièmement, dans le cadre de notre politique en matière d'intégrité civique, nous libellons ou supprimons les informations fausses ou trompeuses destinées à saper la confiance du public dans une élection ou un autre processus civique, en fonction des affirmations contestées/non vérifiables et de la propension à nuire.

Enfin, en février 2020, nous avons annoncé notre politique dédiée à la lutte contre la propagation de médias manipulés et synthétiques sur Twitter - tels que les deep fakes - qui ont la capacité de perturber la conversation publique. Nous pensons, en effet, que toute personne doit être en mesure de trouver des informations fiables sur Twitter. Ainsi, les gens doivent pouvoir obtenir les informations et le contexte nécessaires pour comprendre si le contenu qu'ils voient sur Twitter est réel ou fabriqué. Par conséquent, nous pouvons libeller les Tweets qui incluent des médias (vidéos, audio et images) qui ont été modifiés ou fabriqués de manière trompeuse. En outre, il est interdit de partager sur Twitter des médias altérés de



manière trompeuse, de façon à induire en erreur ou à tromper les gens quant à l'authenticité du média, lorsque des menaces pour la sécurité physique ou d'autres dommages graves peuvent en résulter. Nous utilisons les critères ci-dessous lorsque nous examinons les Tweets et les médias pour les libeller ou les supprimer en vertu de cette politique, dans le cadre de notre travail continu pour faire respecter nos règles et assurer une conversation sereine et sûre sur Twitter.

Is the media significantly and deceptively altered or fabricated?	Is the media shared in a deceptive manner?	Is the content likely to impact public safety or cause serious harm?	
✓	✗	✗	Content <b>may</b> be labeled
✓	✗	✓	Content is <b>likely</b> to be labeled, or <b>may</b> be removed.
✓	✓	✗	Content is <b>likely</b> to be labeled.
✓	✓	✓	Content is <b>very likely</b> to be removed.

### 1.3.6 S'agissant des informations communiquées aux différents utilisateurs concernant l'instruction d'un contenu signalé, l'opérateur précisera :

- quels sont les utilisateurs informés (l'auteur du signalement, l'auteur du contenu signalé, les utilisateurs ayant interagi avec le contenu signalé, l'ensemble des utilisateurs) ;
- pour chacun, quels sont les stades de la procédure qui donnent lieu à une information spécifique adressée aux utilisateurs (envoi d'un signalement, mise en instruction du signalement, issue de l'instruction, raisons précises de la décision prise à cette issue).

La transparence est au cœur du travail que nous faisons chez Twitter. Cela inclut la nécessité de communiquer de manière transparente au sujet de l'application de nos règles. Tout d'abord, lorsque nous recevons un signalement de la part d'une personne, nous accusons réception de ce signalement. Ensuite, en fonction de la méthode de signalement choisie (via l'application Twitter ou les formulaires de notre [Centre d'assistance](#)), nous assurons un suivi par courrier électronique ou par le biais d'une notification Twitter pour informer l'auteur du signalement de la suite donnée à son signalement. En cas de violation de nos règles, nous communiquerons avec l'auteur du contenu pour l'informer des mesures prises à son encontre, en précisant la règle et la politique spécifiques qui ont été enfreintes. Enfin, tout contenu ayant fait l'objet d'une action sur Twitter affiche un message d'information sur la plateforme à l'intention de toutes les personnes, en fonction du type d'action qui a été prise (contenu supprimé pour avoir enfreint les règles de Twitter, contenu masqué localement sur la base des lois locales, libellés, etc.).

### 1.3.7 L'opérateur indiquera au Conseil si des voies de recours internes existent, si elles sont portées à la connaissance des utilisateurs et si oui, desquels. Il précisera notamment :

- quelles sont les décisions de la plateforme susceptibles d'être contestées ;
- quels sont les utilisateurs en droit d'exercer un recours, dans quel délai la possibilité de recours leur est présentée à l'issue de la décision et par quelle voie ;
- quelles sont les équipes en charge de l'instruction de ces recours ;
- quels sont les critères de recevabilité des recours ;
- les conditions dans lesquelles s'effectue l'instruction du recours, en particulier si l'utilisateur à l'origine du recours est en capacité de faire valoir ses explications aux équipes de modération (voire, le cas échéant, aux autres utilisateurs impliqués) ;



- **le délai moyen de traitement d'un recours ;**
- **si les informations sont fournies systématiquement aux utilisateurs s'agissant des recours internes ou s'il existe des exceptions, prévues ou de fait.**

Les systèmes de Twitter sont construits sur la base du concept de "justice procédurale", en reconnaissance du fait que les décisions tant automatisées qu'humaines peuvent être imparfaites et que les personnes utilisant notre service peuvent vouloir exprimer leur désaccord avec une mesure d'application particulière. Twitter propose des voies de recours sous la forme d'appels qui permettent de rétablir des comptes sur la plateforme. Dans le contexte particulier où nous développons de nouveaux moyens de réduire la désinformation sur la plateforme, toutes les actions, telles que les libellés, ne sont pas encore contestables, bien que nous ayons l'intention de développer ces possibilités pour toutes les personnes qui utilisent Twitter.

Afin de maintenir un environnement sûr pour les personnes sur Twitter, nous pouvons envoyer un "défi technique" à un compte en lui demandant de confirmer son identité, libeller un Tweet, libeller un compte, restreindre un compte, demander à un utilisateur de supprimer un Tweet, suspendre temporairement un compte ou suspendre définitivement un compte. Chacune de ces interventions s'accompagne d'une variété de canaux de travail, puisque les restrictions ou défis lancés à des comptes peuvent impliquer de passer un captcha, ou de fournir un numéro de téléphone, tandis que des violations mineures par des systèmes automatisés permettent à un compte de se débloquent par lui-même en appel. Les délais d'appel augmentent en fonction de la gravité de l'infraction. Lorsqu'un compte adopte un comportement abusif, comme l'envoi de menaces à d'autres personnes ou usurpe l'identité d'autres comptes, nous pouvons le suspendre temporairement ou, dans certains cas, définitivement, bien que toutes ces décisions soient susceptibles d'appel. Il existe plusieurs raisons courantes de suspension. La majorité des comptes que nous suspendons le sont parce qu'ils sont des spams ou des faux, et qu'ils présentent des risques de sécurité pour Twitter et nos utilisateurs et perturbent la sérénité de la conversation. En cas d'erreur de notre part, nous travaillerons avec la personne concernée pour faire en sorte que son compte ne soit plus suspendu et nous lui enverrons une notification de notre erreur. Si nous soupçonnons qu'un compte a été piraté ou compromis, nous pouvons le suspendre ou en restreindre l'accès jusqu'à ce qu'il puisse être sécurisé et rendu à son propriétaire, afin de réduire les activités potentiellement malveillantes causées par la compromission.

Selon le type de violation et de problème, les personnes peuvent faire appel et, dans certains cas, annuler la suspension de leur propre compte. Les restrictions sur les comptes sont dues à des activités telles que la création de faux engagements sur des contenus ou le fait de suivre des comptes en masse. Ces actions suspectes nécessitent des informations supplémentaires de la part de l'utilisateur afin qu'il puisse continuer à utiliser son compte Twitter. L'utilisateur aura alors deux possibilités : soit utiliser Twitter de manière temporaire et limitée, soit suivre nos instructions pour vérifier son numéro de téléphone ou son adresse électronique. Enfin, en cas de suspension définitive d'un compte pour violation de nos règles, l'utilisateur peut faire appel, ce qui déclenchera automatiquement un deuxième examen interne par nos équipes. Nous pouvons alors être en mesure de lever la suspension du compte en fonction des résultats de ce deuxième examen. À ce stade, les équipes examinent si l'infraction était grave, si les décisions concernant le compte étaient clairement impartiales et conformes à la politique, ou s'il y a eu des erreurs, ou s'il existe des facteurs atténuants supplémentaires dans le cadre de l'appel. En conséquence, le compte peut être rétabli ou recevoir un dernier avertissement en cas de poursuite de ce comportement, ce qui entraînera une autre suspension permanente et la perte de l'accès à la plateforme.

### **1.3.8 L'opérateur indiquera au Conseil si ont été développés :**

- **des mécanismes d'autoévaluation des procédures et des mesures prises à l'égard d'un contenu signalé et les critères sur lesquels cette évaluation est basée ;**





- **des mécanismes d'évaluation ou enquêtes de satisfaction de ces procédures et mesures à l'attention des utilisateurs. Le cas échéant, il informera le Conseil de leurs résultats et de la façon dont il les a prises en compte.**

Twitter est une organisation en constante évolution, et nous améliorons constamment notre approche et nos processus afin d'être les plus transparents possible envers les personnes qui utilisent Twitter, au service de conversations sereines sur la plateforme. Nous recueillons actuellement des commentaires par le biais d'enquêtes, d'initiatives de recherche auprès de nos utilisateurs, de Tweets, de la presse, de réponses à Twitter Support (@TwitterSupport), ainsi que par le biais de notre procédure d'appel. En outre, nous engageons des experts sur les meilleurs moyens de combattre la désinformation sur la plateforme, et nous réévaluons continuellement nos politiques pour atténuer les préjudices liés aux conversations sur la plateforme. En interne, nous effectuons également des audits sur les signalements pour vérifier si les agents ont traité le dossier avec précision à des fins "d'assurance qualité". À cette fin, l'assurance qualité (AQ) est une mesure essentielle de notre travail quotidien pour nous aider à garantir que nous fournissons un service cohérent au niveau de qualité souhaité à nos principales parties prenantes, tant en externe qu'en interne. La procédure d'assurance qualité (AQ) nous aide à identifier les domaines où il est possible de dispenser des formations et de déceler des défauts potentiels dans notre flux de travail ou nos politiques. Un certain nombre d'agents spécialisés des équipes d'assistance mondiales de Twitter effectuent des contrôles d'assurance qualité des signalements pour s'assurer que le contenu est traité de manière appropriée. Les mises à jour concernant les événements d'actualité importants ou les changements de politique sont partagées avec tous les agents en temps réel, afin de donner des conseils et de faciliter une prise de décision équilibrée et informée. Des séances hebdomadaires de politique et de calibrage sont proposées aux équipes respectives afin d'accroître la compréhension collective et de répondre à toute question rencontrée par les agents dans leur travail quotidien.

#### **1.4 L'opérateur fournira des informations chiffrées sur la mise en œuvre du dispositif de signalement de fausses informations en 2020, notamment :**

- **le nombre de contenus signalés comme fausses informations ;**
- **le nombre de procédures de traitement mises en œuvre ;**
- **le nombre de contenus considérés comme étant de fausses informations à l'issue de la procédure de traitement ;**
- **le délai moyen de traitement d'un signalement ;**
- **le nombre de recours reçus par l'opérateur à l'issue de la procédure de traitement ;**
- **le budget alloué au traitement des signalements par les utilisateurs depuis l'entrée en vigueur de la loi du 22 décembre 2018 et l'évolution de ce budget entre 2019 et 2020.**

**L'opérateur est invité à décrire et analyser la tendance observée pour les indicateurs précités.**

**Il indiquera comment ces chiffres, ou toute autre donnée relative à la circulation effective de fausses informations sur son service, sont portés à la connaissance des utilisateurs.**

Il est interdit d'utiliser Twitter dans le but de manipuler ou d'interférer dans les élections. En plus du canal de signalement requis par la loi française pour les contenus qui diffusent de "fausses informations susceptibles d'altérer la sincérité d'un scrutin ou de troubler l'ordre public", Twitter propose également un canal de signalement distinct pour les contenus qui diffusent de fausses informations sur le vote ou l'inscription sur les listes électorales. Cette politique comprend trois catégories de comportements et de contenus trompeurs liés aux élections. Premièrement, un individu ne peut pas partager des informations fausses ou trompeuses sur la manière de participer à une élection ou à un autre événement civique. Cela inclut, sans s'y limiter, des informations trompeuses sur la façon de voter ou de s'inscrire sur les listes électorales, les conditions requises pour voter, y compris les conditions d'identification, et la date ou l'heure officielle et annoncée d'une élection. Deuxièmement, un individu ne peut pas partager des informations fausses ou trompeuses, destinées à intimider ou dissuader les électeurs de participer à une élection. Cela inclut - sans s'y limiter - les affirmations trompeuses selon lesquelles les bureaux de vote



sont fermés, que le scrutin est terminé ou toute autre information trompeuse relative à des votes non comptabilisés. Nous n'autorisons pas non plus les allégations trompeuses concernant la police ou les activités des forces de l'ordre liées aux lieux de vote ou aux élections, les longues files d'attente, les problèmes d'équipement, les procédures ou techniques de vote qui pourraient dissuader les électeurs de participer à une élection, et les menaces concernant les lieux de vote. Enfin, nous n'autorisons pas non plus la création de faux comptes qui présentent faussement leur affiliation, ni le partage de contenu qui présente faussement son affiliation à un candidat, un élu, un parti politique, une autorité électorale ou une entité gouvernementale. Nous avons encore étendu nos capacités d'application pour les élections à travers le monde en créant une fonction de signalement dédiée au sein de notre produit pour permettre aux utilisateurs de signaler plus facilement les contenus qui nuisent au processus d'inscription sur les listes électorales ou d'engagement dans le processus électoral. En 2020, nous avons reçu 168 709 signalements uniques via le canal de signalement spécifiquement dédié à la France au sujet de "fausses informations susceptibles d'altérer la sincérité d'un scrutin ou de troubler l'ordre public". Veuillez noter qu'en raison d'un changement dans les métriques internes que nous utilisons pour suivre le nombre de signalements uniques reçus de la part de nos utilisateurs au sein de chaque canal de signalement, la comparaison avec les chiffres de 2019 n'est pas pertinente. Le chiffre transmis ici pour 2020 est le meilleur reflet du nombre de signalements uniques d'utilisateurs que Twitter a reçus via le canal de signalement exigé par la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information.

### **1.5 Cas d'usage : l'opérateur partagera également avec le Conseil les procédures qu'il suit dans le cas d'un contenu contenant des éléments reconnus comme faux par des sources fiables d'information et publié sur le service.**

**Le Conseil ne souhaite pas ici que l'opérateur se prononce sur un cas précis mais qu'il indique quelle est sa méthode de travail sur ce type de cas : il souhaite comprendre les grandes étapes suivies par l'opérateur pour traiter ce type de cas, en précisant à chaque fois les critères présidant aux choix qui peuvent être faits, la nature de l'intervention humaine et les données prises en compte (signalement interne, signalement utilisateur, viralité du contenu, etc.).**

**Un arbre de décision général pourra être fourni ainsi que des statistiques agrégées sur la performance de ces étapes : taux de contenus détectés en interne avec taux de faux positifs et de faux négatifs, recours à des partenaires pour la qualification des contenus comme relavant de la fausse information ou qualification de ces contenus par des équipes internes, temps moyen constaté entre la diffusion d'un contenu et la première action prise par l'opérateur, actions prises par l'opérateur et impact de ces actions prises sur la circulation du contenu.**

Afin de déterminer si un contenu atteint notre seuil de préjudice, nous nous appuyons sur des partenaires de confiance et des experts dans un domaine particulier pour prendre des décisions. Lorsque les faits ou les contextes changent autour d'un événement ou d'une question particulière, nous évaluons d'abord si le contenu partagé avance une affirmation de fait, afin de s'assurer que nous protégeons la capacité d'exprimer un large éventail de points de vue et d'opinions. Nous évaluons tous les éléments du contenu partagé, y compris les liens pertinents et les médias inclus. Ensuite, nous déterminons si le contenu présenté est manifestement faux ou présenté de manière trompeuse. Il s'agit ici notamment d'évaluer si le contenu a été présenté d'une manière trompeuse ou altérée qui pourrait conduire à induire le public en erreur.

S'il répond à ces deux critères, nous évaluons alors les préjudices potentiels des allégations fausses ou trompeuses avancées dans le contenu en questions. Pour les contenus qui atteignent notre seuil de préjudice grave, nous les retirerons de la plateforme. Pour les contenus évalués comme présentant un préjudice modéré, nous appliquerons une série de mesures correctives, notamment l'apposition de libellés fournissant des informations fiables sur le sujet ou un lien vers nos politiques. L'application de ces



libellés limite la visibilité et les possibilités d'interaction avec ces Tweets présentant des préjudices modérés.

## **Autres questions**

**1.6 Le Conseil souhaite savoir si l'opérateur a engagé des discussions avec d'autres opérateurs soumis au titre III de la loi du 22 décembre 2018 afin d'harmoniser leurs dispositifs de signalement des fausses informations respectifs et, le cas échéant, quelles ont été les mesures prises dans le cadre de ces discussions.**

**L'opérateur pourra porter à la connaissance du Conseil les initiatives qu'il porte visant à renforcer la coopération avec d'autres plateformes (mise en commun et partage de données, remontée d'informations sensibles...).**

La sérénité de la conversation publique est la priorité absolue de Twitter. Pour atteindre cet objectif, les signalements de la part des personnes qui utilisent notre service restent très importants et il est essentiel pour nous, dans ce processus, d'y répondre le mieux possible, le plus rapidement possible et de manière aussi transparente que possible. Au fil des années, nous avons travaillé pour que notre mécanisme de signalement soit facile à utiliser et à comprendre pour tous.

Chaque produit étant différent, nous pensons qu'il est important que le mécanisme de signalement soit aussi adapté que possible au type de contenu et à la conception de la plateforme où il est mis en œuvre. Twitter a donné la priorité à un mécanisme de signalement facilement accessible, simple et didactique pour les internautes. Le mécanisme de signalement est directement accessible au niveau du contenu ou du compte, marqué par l'icône d'un drapeau et clairement nommé "Signaler ce Tweet".

Pour les raisons mentionnées ci-dessus, nous n'avons pas eu de discussions sur l'harmonisation des mécanismes de signalement avec d'autres opérateurs.

**1.7 L'opérateur indiquera dans quelle mesure il a tenu compte des recommandations du Conseil énoncées dans son bilan relatif à l'exercice 2019 à compter de sa publication en juillet 2020.**

En ce qui concerne les outils de signalement, le CSA a recommandé aux plateformes de travailler sur l'accessibilité, la clarté et l'ergonomie de leurs outils. Nous avons donc utilisé les recommandations du CSA pour examiner et évaluer nos propres outils de signalement existants en matière de fausses informations à l'aune des ces recommandations. Nous avons constaté que nous nous conformions à ces recommandations. En effet, Twitter propose un mécanisme de signalement directement accessible au niveau du Tweet en un clic (coin supérieur droit). Le bouton du mécanisme de signalement est clair : "Signaler ce Tweet" et identifié par un drapeau. Le signalement peut être réalisé en 5 étapes.

**1.8 L'opérateur est invité à indiquer au Conseil l'impact de la crise sanitaire sur son dispositif de signalement de fausses informations et, notamment, les adaptations qu'il a pu être amené à lui apporter en conséquence.**

Les effets de la COVID-19 sur les opérations de Twitter ont déjà dépassé tous les événements que nous avons connus à ce jour, et il est possible qu'au fur et à mesure que la pandémie se développe, nous observons des tensions supplémentaires sur notre service. Depuis mars 2020, la majorité de nos centres d'opérations mondiaux ont été temporairement fermés en raison du confinement et de problèmes sanitaires connexes, ce qui a considérablement réduit notre capacité d'examen humain. Nous avons maintenu des ressources dédiées à l'examen et à la prise de mesures d'application contre les contenus les plus susceptibles de causer un préjudice grave (par exemple, la pédopornographie et le terrorisme),



en plus de donner la priorité aux signalements pour lesquels nous étions en mesure de prédire une forte probabilité de violation de nos Règles. En raison de cette hiérarchisation, nous avons constaté des ralentissements et des allongements de délais de traitement dans d'autres domaines, et nous avons continué à faire évoluer notre approche.

La combinaison de ce nouvel environnement de travail et de l'augmentation de la charge de notre service a créé des contraintes uniques sur nos opérations, obligeant nos équipes d'ingénieurs à collaborer plus étroitement que jamais pour répondre aux nouvelles demandes et pour planifier l'avenir de notre service. De nos équipes d'ingénierie informatique, réseau et produit à nos équipes d'infrastructure et de centre de données, nous nous sommes collectivement mobilisés pour nous assurer d'être en mesure de rester à la fois productifs et en sécurité, le tout sous le poids des nouveaux niveaux de trafic que nous observons sur notre service.



## **2. Transparence des algorithmes**

**2.1 Le Conseil invite les opérateurs à signaler dans leurs réponses les éléments ayant fait l'objet de changements structurants par rapport à l'exercice 2019.**

**L'opérateur précisera en particulier si la tenue des élections municipales en France ainsi que la crise sanitaire ont changé ses procédures habituelles en 2020. Le Conseil souhaite notamment savoir si des modifications ont conduit à une intervention humaine moindre et comment cela a pu affecter la performance globale des systèmes de recommandation et de modération utilisés en 2020 par rapport à l'exercice 2019.**

Alors que le monde entier faisait face - et continue de faire face - à une urgence de santé publique sans précédent, nous avons fait preuve de transparence et d'ouverture quant aux mesures d'urgence que nous avons mises en place dès le début de la crise pour servir la conversation publique en ce moment critique. Cette approche s'est appuyée sur le travail précieux de nos partenaires de confiance, notamment les autorités de santé publique, les organisations non-gouvernementales et les gouvernements du monde entier.

Par conséquent, nos mesures visant à protéger la conversation publique ont été axées sur le maintien de notre service, la poursuite de l'application de nos règles et la protection de la plateforme dans le contexte de la transition de nos équipes vers le télétravail.

Nous avons dû faire pivoter sur le plan opérationnel nos principaux efforts pour assurer la sécurité des personnes sur Twitter. Par conséquent, nous avons augmenté notre utilisation du machine learning et de l'automatisation pour prendre un large éventail d'actions sur les contenus potentiellement abusifs et trompeurs, y compris les fausses informations. Voici quelques-unes des actions et moyens supplémentaires que nous avons mis en place pour protéger la conversation publique :

- Nous avons construit des systèmes qui permettent à nos équipes de continuer à faire respecter nos règles à distance, dans le monde entier.
- Nous avons renforcé l'assistance aux employés et le soutien au bien-être de tous ceux qui participent à ce travail essentiel, et avons veillé à ce que la confidentialité et la sécurité des personnes restent une priorité absolue.
- Nous avons mis en place un système mondial de triage de la gravité du contenu afin de donner la priorité aux violations potentielles des règles qui présentent le plus grand risque de préjudice et ce, afin de réduire la charge qui pèsent sur les victimes et les personnes qui signalent des contenus.
- Nous effectuons des contrôles quotidiens de qualité de nos processus d'application sur les contenus afin de nous assurer que nous sommes toujours en mesure de répondre à cette épidémie mondiale qui évolue rapidement.
- Nous nous engageons auprès de nos partenaires dans le monde entier pour que les voies de signalement restent toujours disponibles et que les cas les plus urgents puissent être portés à notre attention dans les meilleurs délais.

### ***Cadre général de réponse***

**2.2 L'opérateur est invité à commenter les définitions de la loyauté, de l'explicabilité et de l'équité indiquées au début du questionnaire. Le Conseil souhaite en particulier savoir si l'opérateur les utilise (ou des notions équivalentes) notamment dans le cadre des mesures qu'il met en œuvre afin de lutter contre la manipulation de l'information, et les éventuelles limites identifiées à leur application. L'opérateur est invité à formuler d'éventuelles propositions afin d'affiner ces définitions.**



Chez Twitter, nous voulons créer la meilleure expérience possible pour les millions de personnes qui utilisent quotidiennement notre service dans le monde entier. Pour ce faire, nous évaluons continuellement nos systèmes afin de comprendre si les systèmes de machine learning que nous avons en production, et qui ont un impact sur des millions de personnes chaque jour, produisent des informations et des recommandations qui sont transparentes, responsables et évitent autant que possible les biais.

Cependant, nous aimerions attirer l'attention du Conseil sur le fait que le concept "d'équité" n'est pas un concept que nous utilisons actuellement dans le cadre de notre travail, car il implique un jugement moral. De la même manière, l'Organisation internationale de normalisation (ISO) et l'Association de normalisation de l'Institut des ingénieurs électriciens et électroniciens (IEEE) évitent ce concept "d'équité" et recommandent d'utiliser le mot "biais". Plus précisément, l'équité est un concept très subjectif et personnel qui peut avoir une signification différente pour chaque personne. Il n'est pas non plus défini de la même manière à travers les communautés dans le monde et, comme Twitter est au service de nombreuses cultures, ce qui est "équitable" dans une culture peut ne pas être reconnu comme tel ailleurs. Ainsi, nous préférons mesurer des concepts qui sont définis moins subjectivement et nous concentrer sur des mesures quantifiables.

Enfin, pour nous assurer que notre travail est solidement ancré dans ces principes et concepts, nous avons créé l'équipe META (Machine Learning Ethics, Transparency and Accountability) afin de garantir que nos valeurs fondamentales et les concepts qui les définissent - notamment la lutte contre les biais et en faveur des droits de l'Homme - restent au coeur des systèmes de machine learning de Twitter. META est une équipe pluridisciplinaire chargée d'étudier l'équité et la transparence des systèmes automatisés de prise de décision algorithmique de Twitter. Sa mission consiste à s'assurer que les systèmes de machine learning de Twitter sont transparents, responsables et dépourvus au maximum de biais. L'équipe utilise des méthodologies et des cadres d'investigation bien établis pour comprendre et atténuer la partialité de nos produits alimentés par le machine learning.

**2.3 Lorsque l'opérateur est invité à décrire les systèmes reposant sur un ou plusieurs algorithmes qu'il utilise, les éléments suivants sont attendus pour chacun de ces systèmes afin de permettre au Conseil de mesurer le niveau de transparence offert aux utilisateurs :**

- sur quelles fonctionnalités du service opère-t-il ?
- quel est son objectif principal (tâche ou classe de problèmes à résoudre) ?
- repose-t-il sur un ou plusieurs algorithmes d'apprentissage automatique (*machine learning*) ? Ce ou ces derniers est-il ou sont-ils utilisé(s) avec d'autres types d'algorithmes (ex. algorithmes déterministes) ?
- lorsqu'il repose sur un ou plusieurs algorithmes d'apprentissage automatique, quelle est la nature de(s) jeu(x) de données d'apprentissage utilisé(s) ?
- quelles sont les données prises en entrée de ce système ?
- quelles sont les données de sortie de ce système ?
- quelles conditions mises en œuvre par l'opérateur garantissent effectivement que ses services et leurs algorithmes respectent les principes de loyauté et d'équité, et qu'ils soient explicables ?
- le cas échéant, comment ce système est-il personnalisé en fonction de l'utilisateur ?
- le cas échéant, quelles sont modalités de recours à une intervention humaine sur le système à partir de sa mise en œuvre ? Ce recours est-il le fait de personnes en interne, de personnes employées par des prestataires, ou bien les deux ? Combien de personnes sont affectées à ces tâches ? Quelle est leur formation, ainsi que le lieu depuis lequel elles travaillent sur les systèmes affectant les utilisateurs basés en France ?
- les fausses informations font-elles l'objet d'un traitement particulier par ce système et,, si oui, de quelle nature ?



**Les opérateurs pourront s'ils le souhaitent recourir à des représentations graphiques afin de répondre aux questions. Ces représentations devront faire figurer clairement l'ensemble des éléments de réponse demandés aux questions concernées.**

### ***Informations fournies au régulateur pour qu'il évalue les mesures de transparence des algorithmes mises en œuvre par l'opérateur***

#### **2.4 Il est demandé à l'opérateur de décrire les systèmes de recommandation de contenus qu'il utilise. Pour chacun de ces systèmes, il fournira les éléments demandés au point 2.3.**

Nous voulons que Twitter offre une expérience utile et pertinente à toutes les personnes qui utilisent notre service. Avec des centaines de millions de Tweets chaque jour sur Twitter, nous avons investi massivement dans la construction de systèmes qui organisent le contenu sur Twitter afin de montrer aux personnes utilisant le service les informations les plus pertinentes en premier. Comme des millions de personnes utilisent Twitter chaque jour dans des dizaines de langues et d'innombrables contextes culturels, nous nous appuyons sur des algorithmes de machine learning pour nous aider à organiser le contenu par pertinence.

Tout d'abord, le fil d'actualité Twitter affiche un flux de Tweets provenant de comptes que les gens ont choisi de suivre. Les utilisateurs peuvent voir des suggestions de contenu basées sur une variété de signaux. Ils peuvent répondre, retweeter ou aimer un Tweet à partir de leur fil. Les Tweets affichés en premier sont ceux auxquels les gens sont susceptibles de s'intéresser le plus, et nous les choisissons en fonction des comptes avec lesquels les gens interagissent le plus et des Tweets et conversations auxquels ils participent. Toutefois, si une personne souhaite voir Twitter sans qu'aucun algorithme ne soit appliqué, elle dispose d'une commande accessible en un clic afin d'afficher son fil d'actualité dans l'ordre chronologique inverse. Cela permet aux utilisateurs de Twitter de voir très concrètement comment les algorithmes affectent ce qu'ils voient, et offre une plus grande transparence sur la technologie que nous utilisons pour classer les Tweets.

Les utilisateurs peuvent également voir un résumé des Tweets les plus intéressants qu'ils n'ont peut-être pas vus, intitulé "Au cas où vous l'auriez manqué". En outre, lorsque nous identifions un Tweet, un compte à suivre ou tout autre contenu populaire ou pertinent, nous pouvons l'ajouter au fil d'actualité des internautes. Cela signifie que les internautes verront parfois des Tweets de comptes qu'ils ne suivent pas. Nous sélectionnons chaque Tweet à l'aide de divers signaux, notamment sa popularité et la manière dont les autres membres du réseau de la personne concernée interagissent avec lui. Notre objectif est de montrer aux gens, sur leur fil d'actualité, le contenu qui les intéresse le plus et qui contribue à la conversation de manière significative (par exemple, un contenu pertinent, crédible et sûr).

Nous essayons de déterminer le contenu que les gens pourraient apprécier en fonction d'un certain nombre de facteurs. Si le paramètre "Autoriser l'utilisation des endroits où vous voyez du contenu Twitter sur le Web" est activé dans les paramètres "Personnalisation et données", nous pouvons tenir compte de leurs visites sur d'autres sites Web qui intègrent du contenu de Twitter (comme les fils d'actualité intégrés). Par exemple, si une personne visite régulièrement des sites web consacrés à l'observation d'oiseaux, nous pouvons lui suggérer des comptes qui Tweetent fréquemment sur ce sujet, ou lui montrer des publicités pour des jumelles ou des mangeoires à oiseaux, par exemple.

Pour protéger la vie privée des personnes, nous n'associons jamais cet historique de navigation web à un nom, une adresse électronique, un numéro de téléphone ou un identifiant Twitter, et nous le supprimons, l'anonymisons ou l'agrégeons au bout de 30 jours au maximum, comme l'explique notre politique de confidentialité. Nous nous engageons à offrir aux gens des choix significatifs en matière de confidentialité. Si une personne ne souhaite pas que la personnalisation (y compris les publicités) soit



basée sur ses visites sur des sites web tiers qui intègrent du contenu Twitter, il existe plusieurs moyens de désactiver cette fonctionnalité :

- Les utilisateurs qui s'inscrivent à Twitter sur le web peuvent décocher la case "Twitter pour le web".
- Après l'inscription ou si les utilisateurs utilisent Twitter en étant déconnectés, ils peuvent désactiver le paramètre "Autoriser l'utilisation des endroits où vous voyez du contenu Twitter sur le web" dans vos paramètres "Personnalisation et données".

Outre le fil d'actualité, Twitter propose un onglet Notifications qui permet aux utilisateurs de voir qui a aimé, retweeté et répondu à leurs Tweets, et qui les a mentionnés ou suivis. Nous donnons aux personnes sur Twitter des contrôles supplémentaires sur le contenu qui apparaît dans cet onglet Notifications, car les notifications peuvent contenir du contenu qu'une personne n'a pas choisi de recevoir sur Twitter, comme des mentions ou des réponses de quelqu'un que la personne ne suit pas. Par défaut, nous filtrons les notifications en fonction de leur qualité et excluons les notifications concernant des Tweets dupliqués ou potentiellement liés à du spam. Nous offrons également aux utilisateurs de la plateforme des contrôles granulaires sur les types de comptes spécifiques dont ils ne souhaitent pas recevoir de notifications, notamment les nouveaux comptes, les comptes que l'utilisateur ne suit pas et les comptes sans téléphone ou adresse électronique confirmés.

Les réponses à un Tweet donné sont appelées "conversations". Twitter s'efforce de montrer aux gens le contenu qui les intéressera le plus et qui contribuera de manière significative à la conversation. Pour cette raison, les réponses, regroupées par sous-conversations, peuvent ne pas être dans l'ordre chronologique. Par exemple, lorsque nous classons une réponse tout en haut, nous prenons en compte des facteurs comme le fait que l'auteur du Tweet original ait répondu ou que la réponse provienne d'une personne que la personne suit.

En ce qui concerne les suggestions de comptes sur Twitter, elles sont basées sur de nombreux facteurs, y compris les contacts du carnet d'adresses des personnes (si elles ont choisi de les télécharger) ainsi que des modèles provenant de leur historique de suivi. Ces suggestions sont générées par des algorithmes, ce qui signifie que les gens peuvent ou non connaître les comptes ou les trouver pertinents. Les suggestions de comptes de Twitter sont donc basées sur des algorithmes qui font des suggestions personnalisées. Les personnes peuvent voir des suggestions basées sur des critères tels que les suivants :

- Si une personne a téléchargé ses contacts sur Twitter et que l'adresse électronique ou le numéro de téléphone d'une autre personne figure dans ses contacts, nous pouvons suggérer à cette personne de la suivre.
- Twitter peut également faire des suggestions en fonction de l'emplacement, par exemple la ville ou le pays où se trouve une personne.
- Nous pouvons faire des suggestions en fonction de l'activité des personnes sur Twitter, comme les Tweets, les comptes suivis, les comptes et les Tweets que les personnes consultent ou avec lesquels elles interagissent.
- Si les paramètres d'une personne le permettent, nous pouvons faire des suggestions personnalisées en fonction de ses visites sur des sites Web tiers qui intègrent du contenu Twitter.
- Twitter peut inclure des comptes promus comme suggestions de comptes à suivre.

D'autre part, afin de préserver l'intégrité de notre plateforme et de protéger les conversations contre toute manipulation, Twitter utilise également des outils et des technologies pour déceler et minimiser la visibilité de certains types de comportements abusifs et trompeurs.

Twitter utilise une série de signaux comportementaux pour déterminer la manière dont les Tweets sont organisés et présentés dans le fil d'actualité, les conversations et la barre de recherche en fonction de la pertinence. Twitter s'appuie sur des signaux comportementaux, tels que la façon dont les comptes se





comportent et réagissent les uns aux autres, pour identifier les contenus qui nuisent à une conversation publique sereine, comme le spam et les abus. À moins que nous n'ayons déterminé qu'un Tweet viole les politiques de Twitter, il restera sur la plateforme. Lorsque nous avons identifié un Tweet comme pouvant nuire à une conversation sereine (par exemple, comme potentiellement abusif), il ne pourra être consulté que si une personne clique sur "Afficher plus de réponses" ou choisit de tout voir dans ses paramètres de recherche.

Voici quelques exemples de signaux comportementaux que nous utilisons, en combinaison les uns avec les autres et avec une série d'autres signaux, pour aider à déceler ce type de contenu : un compte sans adresse électronique confirmée, l'inscription simultanée de plusieurs comptes, des comptes qui Tweetent et mentionnent de manière répétée des comptes qui ne les suivent pas, ou un comportement qui pourrait indiquer une attaque coordonnée. Twitter examine également comment les comptes sont liés à ceux qui violent nos règles et comment ils interagissent entre eux. La précision des algorithmes développés à partir de ces signaux comportementaux continue de s'améliorer au fil du temps.

Ces signaux comportementaux sont un facteur important dans la façon dont Twitter organise et présente le contenu dans des espaces tels que la conversation et la barre de recherche. Notre objectif principal est de faire en sorte que les contenus pertinents et les Tweets contribuant à une conversation sereine apparaissent en premier dans les conversations et les recherches. Comme notre service fonctionne dans des dizaines de langues et des centaines de contextes culturels à travers le monde, nous avons constaté que le comportement est un signal fort qui nous aide à identifier les acteurs malveillants sur notre plateforme. Le classement comportemental utilisé par Twitter ne tient absolument pas compte des opinions politiques ou de l'idéologie. Il se concentre uniquement sur le comportement de tous les comptes. Twitter s'efforce en permanence d'améliorer ses modèles de classement basés sur le comportement afin d'en accroître l'étendue et la précision au fil du temps. Nous utilisons des milliers de signaux comportementaux dans nos modèles de classement basés sur le comportement, ce qui garantit qu'aucun signal ne détermine les résultats du classement et protège contre les tentatives malveillantes de manipulation de nos systèmes de classement.

Les utilisateurs peuvent [télécharger leurs données](#) à tout moment sur Twitter, y compris les intérêts que Twitter a déduits de leur compte ou de leur appareil de connexion. Ces centres d'intérêt permettent d'améliorer l'expérience de Twitter, par exemple, en affichant un meilleur contenu, notamment des publicités, des notifications et des Tweets recommandés dans le fil d'actualité et dans la barre de recherche.

## **2.5 L'opérateur décrira également au Conseil les systèmes de modération de contenus qu'il utilise pour permettre notamment de lutter contre les fausses informations.**

**Pour chacun de ces systèmes, il fournira les éléments demandés au point 2.3 et répondra à la question suivante :**

- **Quelle est la performance de ce système (notamment, taux de faux positifs et de faux négatifs) ?**

Twitter utilise des outils automatisés pour déceler de manière proactive certains types de contenus d'informations trompeuses. Ce contenu est examiné par des membres de nos équipes afin de déterminer s'il enfreint nos politiques. Notre objectif est d'utiliser à terme un examen à la fois automatisé et humain pour traiter le contenu qui enfreint nos règles en matière de désinformation. Le machine learning et le traitement automatisé du langage prennent du temps pour être efficaces. C'est pourquoi nous commençons par le contenu en langue anglaise et utiliserons par la suite le même processus pour l'étendre à d'autres langues et contextes culturels au fil du temps.



## **2.6 Lorsque l'opérateur utilise un ou plusieurs algorithmes d'apprentissage automatique, il est invité à détailler sa politique de lutte contre les biais pouvant exister dans les données d'apprentissage.**

Le machine learning est une technologie en pleine évolution et nous en apprenons chaque jour davantage sur cet outil puissant. Nous pensons que les entreprises doivent être conscientes de la façon dont des biais involontaires peuvent être introduits par des systèmes de décision algorithmique comme le machine learning. Les politiques ainsi que la formation interne pour aider à lutter contre les biais doivent être une pratique standard dans l'industrie. Nous sommes conscients que la lutte contre les biais est un effort à long terme et qu'il est impossible d'éliminer tous ces biais dans les systèmes qui apprennent d'un monde malheureusement toujours biaisé.

Les biais peuvent être introduits plus largement au cours du processus de développement d'un produit, et non seulement via les données utilisées pour entraîner un modèle. Dans un premier temps, nous pensons, toutefois, que la mesure de la représentativité des échantillons choisis dans les ensembles de données d'apprentissage est un excellent moyen de lutter contre les biais.

Par conséquent, notre approche chez Twitter consiste à évaluer nos systèmes pour comprendre si les systèmes de machine learning que nous avons en développement peuvent augmenter ou élargir les biais qui sont expérimentés en dehors de nos produits et services. Nous évaluons les biais avant de publier de nouveaux algorithmes, mais cela ne signifie pas que nos systèmes seront toujours parfaits en matière de lutte contre les biais et de représentation des informations, car nous évoluons dans un monde en constante évolution. Nous continuons donc à apprendre de manière publique et ouverte lorsque notre conception présente des défaillances.

Twitter dispose d'une équipe dédiée chargée de répondre à ce besoin. L'équipe META (Machine learning Ethics, Transparency and Accountability) a pour mission d'aborder les questions de partialité, d'explicabilité et de nombreux autres problèmes émergents que l'industrie et le monde universitaire découvrent en relation avec l'utilisation du machine learning. META est une équipe dédiée et appliquée d'ingénierie, de science des données, de recherche et de produits qui est intégrée à notre organisation d'infrastructure machine learning. L'équipe utilise des méthodologies et des cadres d'investigation bien établis pour comprendre et atténuer les biais dans nos produits alimentés par le machine learning.

### ***Informations fournies aux utilisateurs***

## **2.7 L'opérateur est invité à indiquer comment il décrit à ses utilisateurs les systèmes de recommandation qu'il utilise.**

**Pour chacun de ces systèmes, il précisera comment il leur apporte des informations sur les éléments décrits au point 2.3 et répondra aux questions suivantes :**

- **quel(s) réglage(s) peu(ven)t être effectué(s) par l'utilisateur pour en changer les paramètres ? Combien d'utilisateurs y ont eu recours en 2020 et à quelle fréquence ?**
- **comment ces réglages sont-ils rendus facilement accessibles et compréhensibles ?**
- **comment les utilisateurs sont-ils informés des mises à jour de ces systèmes de recommandation affectant significativement la manière dont les contenus leur sont recommandés ?**
- **de quelle manière les utilisateurs peuvent-ils demander des informations supplémentaires sur ces systèmes de recommandation ?**
- **comment l'utilisateur bénéficie-t-il d'une explication locale sur les résultats de ces systèmes de recommandation qui l'affectent ?**



**Le Conseil sera attentif à toute précision visant à expliquer comment l'opérateur modifie le niveau et la nature de ces informations en fonction de ses utilisateurs (mineurs, personnes en situation de handicap, développeurs, etc.).**

Twitter offre des informations facilement accessibles en français via son propre [site web](#) afin que tout le monde puisse comprendre comment le contenu est recommandé sur la plateforme. Il s'agit notamment des facteurs pris en considération pour nos systèmes de recommandation de comptes et autres algorithmes de recommandation. En outre, de la même manière que nous fournissons des informations sur nos algorithmes de recommandation, nous fournissons également des informations sur les moyens de mieux contrôler son propre compte et son expérience, par exemple :

- La découvrabilité de l'adresse électronique ou du numéro de téléphone. Les utilisateurs peuvent contrôler si Twitter suggère leur compte à une autre personne dont l'adresse électronique ou le numéro de téléphone figure dans leurs contacts en réglant leurs paramètres de confidentialité.
- Suggestions de comptes basées sur des contacts importés. Lorsqu'une personne télécharge ses contacts sur Twitter, ces derniers peuvent apparaître comme des comptes suggérés à suivre. Inversement, leur compte peut apparaître comme une suggestion à suivre pour les autres. Les utilisateurs peuvent contrôler l'utilisation par Twitter de leur carnet d'adresses pour suggérer leur compte à d'autres personnes en réglant les paramètres de confidentialité qui permettent aux autres de les trouver par leur adresse électronique ou leur numéro de téléphone. Les utilisateurs peuvent également supprimer les contacts qu'ils ont précédemment téléchargés. Lorsqu'elles suppriment des contacts précédemment téléchargés, nous ne ferons plus de suggestions basées sur ces contacts, mais nous pourrons faire des suggestions similaires en fonction des personnes qu'elles suivent déjà et d'autres facteurs.

**2.8 L'opérateur indiquera également la manière dont il décrit à ses utilisateurs les systèmes de modération qu'il utilise notamment dans le cadre de sa lutte contre la diffusion des fausses informations.**

**Pour chacun de ces systèmes, il précisera comment il leur apporte des informations sur les éléments décrits au point 2.3 et répondra aux questions suivantes :**

- **quelle est la performance de ce système (notamment, taux de faux positifs et de faux négatifs) ?**
- **quel(s) réglage(s) peu(ven)t être effectué(s) par l'utilisateur pour en changer les paramètres ?**
- **comment ces réglages sont-ils rendus facilement accessibles et compréhensibles ?**
- **comment les utilisateurs sont-ils informés des mises à jour de ces systèmes de modération affectant significativement la manière dont les contenus sont modérés ?**
- **de quelle manière les utilisateurs peuvent-ils demander des informations supplémentaires sur ces systèmes de modération ?**
- **comment l'utilisateur bénéficie-t-il d'une explication locale des résultats des systèmes de modération qui l'affectent ?**

Les Règles de Twitter ont pour but de garantir que chacun puisse participer à la conversation publique librement et en toute sécurité, à condition de respecter les Conditions d'utilisation de Twitter. Ces Règles - ainsi que notre gamme d'options pour les faire respecter - sont facilement accessibles à tous sur notre [site web](#). Ces Règles, disponibles en français, sont claires et précises et permettent de définir ce qui est autorisé et interdit sur la plateforme, notamment autour de la désinformation sur la COVID-19, de l'intégrité civique et des médias synthétiques et manipulés. Nos Règles sont constamment mises à jour afin de mieux répondre aux nouveaux modèles de comportements problématiques en ligne et potentiellement abusifs.

Twitter est attaché à la transparence et au libre échange d'informations. Nous offrons des informations facilement accessibles via notre propre site web pour que les gens puissent comprendre comment le



contenu est recommandé dans les différentes zones de la plateforme (fil d'actualité, conversations, barre de recherche, etc.).

Nous pensons également qu'il est plus important que jamais d'éclairer toutes nos parties prenantes sur nos propres pratiques, notamment l'application des règles de Twitter et notre travail continu pour perturber les opérations de désinformation soutenues par des États à l'échelle mondiale. Le public et les décideurs politiques veulent être mieux informés de nos actions et nous reconnaissons ces appels à une plus grande transparence.

C'est pourquoi, depuis 2012, nous avons investi du temps, des efforts et des ressources pour faire évoluer notre rapport initial sur l'application de nos règles en un [Centre de transparence](#) plus complet, couvrant un éventail plus large de nos efforts de transparence. Nous incluons maintenant des sections couvrant les demandes d'information et de suppression de contenus (y compris les données par pays et les taux de conformité), les signalement en matière de droits d'auteur et des marques, les spams, l'application des Règles de Twitter, la manipulation de la plateforme et les opérations de désinformation soutenues par des États. Aujourd'hui, plus de la moitié des Tweets sur lesquels nous intervenons sont mis en évidence de manière proactive par notre technologie pour un examen humain.

### ***Systemes spécifiques à la lutte contre la manipulation de l'information***

**2.9 L'opérateur décrira tout système qu'il a développé ou utilise spécifiquement dans le cadre de la lutte contre la manipulation de l'information, notamment contre les campagnes coordonnées de désinformation. Il indiquera s'il informe les utilisateurs de l'existence de ces systèmes.**

**Pour chacun de ces derniers, il fournira les éléments demandés au point 2.3 et indiquera la performance de ce système (notamment, taux de faux positifs et de faux négatifs).**

Twitter utilise une série d'outils de détection pour enquêter sur les acteurs malveillants, les comportements et contenus problématiques sur la plateforme, ainsi qu'un certain nombre d'outils spécialisés contre les campagnes de désinformation coordonnées. Lorsque nous découvrons une campagne de désinformation coordonnée que nous pouvons attribuer de manière fiable à des acteurs soutenus par un État, nous publions les détails de tous les comptes sur notre [site de transparence des opérations de désinformation](#), base de données leader dans l'industrie en la matière. Nous ne fournissons pas de détails sur les systèmes que nous utilisons pour enquêter et attribuer les tactiques, techniques et procédures utilisées par des acteurs spécifiques de la menace pour des raisons de sécurité et pour nous permettre de continuer à déceler les campagnes coordonnées de désinformation.

**2.10 L'opérateur indiquera dans quelle mesure il a tenu compte des recommandations du Conseil énoncées dans son bilan relatif à l'exercice 2019 à compter de sa publication en juillet 2020.**

En 2019, les recommandations du CSA s'orientaient vers la nécessité d'offrir un accès facile aux utilisateurs quant aux informations sur les types d'algorithmes qui aident à organiser le contenu pour eux, ainsi que d'offrir un plus grand contrôle sur ces algorithmes. Conformément à ces recommandations, la philosophie de Twitter a toujours été de responsabiliser les individus et de leur donner autant de contrôle que possible sur le type de contenu qu'ils voient sur la plateforme. Les utilisateurs ont toujours la possibilité de désactiver en un clic les algorithmes de classement dans leur fil d'actualité et de voir les Tweets dans un ordre chronologique inverse, sans influence algorithmique. Nous nous engageons à améliorer notre produit dans ce sens et nous accueillerons volontiers tout commentaire à ce propos. Par ailleurs, toutes les informations concernant notre utilisation des algorithmes sont facilement accessibles via notre site web et nos pages du Centre d'assistance, le tout en de termes simples et clairs, disponibles en plusieurs langues, dont le français.



### 3. Promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle

#### *Identification des sources des contenus*

**3.1 L'opérateur indiquera au Conseil si les sources des contenus issus d'entreprises et d'agences de presse et de Services de communication audiovisuelle (ci-après SCA) sont identifiées et visibles sur son service et, le cas échéant, s'il s'agit d'un dispositif qui leur est spécifique.**

**Il est invité à expliquer les moyens mis en œuvre pour procéder à cette identification, à savoir :**

- **qui la réalise (l'opérateur, les entreprises, agences de presse et SCA ou un tiers) et, dans l'hypothèse où l'opérateur ne la réalise pas directement, s'il vérifie et s'assure de la fiabilité de l'identification des sources (avant ou après publication) ;**
- **par quels moyens, sur la base de quelles sources et selon quels critères l'identification des sources et/ou la vérification de la fiabilité de l'identification est réalisée par l'opérateur et comment ces critères sont portés à la connaissance de l'utilisateur ;**
- **comment l'identification des sources est portée à la connaissance de l'utilisateur sur le service et si des éléments complémentaires d'information et de contexte lui sont apportés ;**
- **les moyens humains, technologiques et financiers déployés par l'opérateur afin d'assurer l'identification, la visibilité et la mise en avant des informations provenant de ces sources dans les résultats des moteurs de recherche, des fils d'actualité et autres espaces de diffusion de contenus opérant par classement algorithmique, y compris, le cas échéant, au sein de groupes d'utilisateurs fermés ;**
- **si ces mêmes moyens sont susceptibles d'évoluer en fonction de l'actualité et, le cas échéant, dans quelle mesure. L'opérateur pourra donner des exemples de cas où l'actualité l'a conduit à faire évoluer ces moyens.**

De par sa nature publique, largement distribuée et en temps réel, Twitter est une plateforme importante pour les journalistes, à la fois pour partager librement leur contenu et pour offrir accès aux gens à des informations publiques en temps réel. De nombreux médias et éditeurs parmi les plus fiables au monde utilisent régulièrement Twitter comme plateforme de diffusion de contenu. Ainsi, Twitter ne recherche pas le contenu des éditeurs de presse et ne l'apporte pas sur la plateforme. Lorsqu'un éditeur de presse (via son propre compte) télécharge son contenu sur Twitter, il a l'intention d'utiliser Twitter comme un canal de distribution et de marketing pour attirer un public par le biais de ses abonnés et un public plus large par l'engagement avec d'autres personnes sur la plateforme.

Un grand nombre de ces organisations, ainsi que certains journalistes et médias, sont certifiés. Le badge bleu "certifié" sur Twitter permet aux gens de savoir qu'un compte d'intérêt public est authentique. Le badge apparaît à côté du nom sur le profil d'un compte et à côté du nom du compte dans les résultats de recherche. Il est toujours de la même couleur et placé au même endroit, quelles que soient les personnalisations de couleurs du profil ou du thème. Fin 2020, après avoir consulté le public pour recueillir ses commentaires, nous avons annoncé que nous relancerions les demandes publiques de certification en 2021 par le biais d'un nouveau processus de demande en libre-service qui sera disponible sur la page "Paramètres" du compte sur le web et dans l'application mobile Twitter. Le processus consistera à demander aux postulants de sélectionner une catégorie - dont une dédiée aux organismes de presse et aux journalistes - pour leur statut certifié et à confirmer leur identité via des liens et d'autres documents justificatifs. Nous prévoyons d'utiliser des processus d'examen automatisés et humains pour nous assurer que nous examinons les demandes de manière réfléchie et en temps utile. Nous prévoyons également de donner aux gens la possibilité de partager des informations démographiques après avoir rempli la nouvelle demande de certification afin que nous puissions mieux



mesurer et améliorer l'équité de notre processus. Nous nous sommes engagés à partager plus de détails sur le processus de demande dans le courant de l'année 2021.

Enfin, en 2020, nous avons pris des mesures supplémentaires pour fournir aux gens davantage de contexte sur le type de comptes qu'ils voient et avec lesquels ils s'engagent sur Twitter. Après avoir interdit toute publicité de la part de média soutenu par un État et toute publicité politique sur Twitter en 2019, nous avons élargi les types de comptes politiques que nous libellons en août 2020. Cela inclut les comptes appartenant à des entités médiatiques affiliées à un État, leurs rédacteurs en chef et/ou leurs cadres dirigeants. Les médias affiliés à un État sont définis comme des organes de presse où l'État exerce un contrôle sur le contenu éditorial par le biais de ressources financières, de pressions politiques directes ou indirectes, et/ou du contrôle de la production et de la distribution du média. Contrairement aux médias indépendants, les médias affiliés à un État utilisent fréquemment leur couverture de l'actualité comme un moyen de promouvoir un programme politique. Nous pensons que les gens ont le droit de savoir quand un média est affilié directement ou indirectement à un acteur étatique. Les organisations médiatiques financées par l'État et jouissant d'une indépendance éditoriale, comme la BBC au Royaume-Uni, France Télévisions en France ou NPR aux États-Unis par exemple, ne seront pas libellées. Dans le cadre de l'élaboration de ce processus, nous avons consulté plusieurs groupes d'experts, notamment les membres du groupe consultatif sur le numérique et les droits de l'Homme au sein du Conseil de confiance et de sécurité de Twitter. Nous n'amplifions pas non plus les comptes de médias affiliés à un État ou leurs Tweets par le biais de nos systèmes de recommandation, notamment sur le fil d'actualité, les notifications et la barre de recherche. Enfin, pour offrir un contexte encore plus large au public, le fait de cliquer sur ces libellés de médias affiliés à un État renvoie les gens à un article expliquant la politique et les renvoie au rapport de transparence de Twitter pour des informations supplémentaires.

### ***Prise en compte des démarches de labellisation***

**3.2 Les moyens mis en œuvre par l'opérateur pour tenir compte des démarches de labellisation et de bonnes pratiques, notamment celles réalisées en tout ou partie par les entreprises et agences de presse et les SCA (ex. : Trust Project, Journalism Trust Initiative), seront également portés à la connaissance du Conseil.**

**L'opérateur décrira les moyens mis en œuvre (ex. : démarches réalisées par ses soins ou par des tiers, intervention directe de ces tiers sur la plateforme ou non, financement, partenariats, mise en avant sur le service (et comment), informations aux utilisateurs, etc.). Il précisera, le cas échéant, sur quels critères sont sélectionnées les démarches de labellisation concernées.**

**L'opérateur précisera si ces démarches de labellisation et de bonnes pratiques constituent un des moyens d'identification des sources et de mise en avant de contenus issus d'entreprises et d'agences de presse et de SCA et de quelle manière, et/ou un des moyens de choisir d'éventuels partenaires œuvrant au *fact-checking*.**

Alors que le monde était confronté à une crise sanitaire sans précédent, nous avons intensifié nos efforts pour protéger la conversation publique, élever les sources d'informations fiables et établir des partenariats avec les gouvernements et les organisations à but non lucratif.

En janvier 2020, avant la qualification officielle du virus et en partenariat avec les agences nationales de santé publique et l'OMS, nous avons lancé une fonctionnalité dédiée dite "d'encart" au sein de la barre de recherche afin que lorsqu'un utilisateur recherche des termes liés à la COVID-19, il soit confronté à un contenu crédible et faisant autorité, tout en haut de son expérience de recherche. Nous surveillons en



permanence la conversation sur le service pour nous assurer que tous les mots clés, y compris les fautes d'orthographe, génèrent des résultats de recherche de qualité.

Plus tard en 2020, nous avons soutenu des informations fiables et faisant autorité sur le thème des vaccins, notamment sur la COVID-19. Notre rôle est de veiller à ce que les gens disposent des informations crédibles nécessaires pour prendre des décisions éclairées. Aujourd'hui, il est en effet plus important que jamais que les gens soient en mesure de prendre des décisions éclairées concernant leur santé et celle de leur famille. Nous comprenons l'importance des vaccins dans la prévention des maladies, et nous reconnaissons le rôle joué par Twitter dans la diffusion d'informations fiables en matière de santé publique. C'est pourquoi, dans le cadre de nos efforts pour protéger la santé du public, nous nous sommes associés à Santé publique France (@SantePubliqueFr) pour faire en sorte que les personnes recherchant des informations sur les vaccins - y compris contre la COVID-19 - sur Twitter puissent facilement accéder à des informations fiables et précises dans leur langue.

Au total, en 2020, le hashtag le plus utilisé était #COVID19, qui a été Tweeté près de 400 millions de fois. Cela montre la valeur et l'importance de la mise en avant d'informations fiables par le biais d'encart au sein de la barre de recherche. #Stayhome a été le troisième hashtag le plus utilisé de l'année, ce qui montre que les gens partagent une plus grande sensibilisation et des préoccupations pour la santé publique, ainsi que la promotion de plus de moyens d'être actif et de se divertir à la maison.

Enfin, dans le contexte des élections présidentielles et de mi-mandat de 2020 aux États-Unis, et dans le but d'aider les électeurs à identifier leurs candidats, nous avons déployé des libellés électoraux qui contiennent des informations supplémentaires sur les comptes Twitter de certains candidats politiques. Le libellé apparaît sur le profil du compte Twitter du candidat et sur les Tweets envoyés et retweetés par le compte du candidat. Les libellés contiennent des informations sur le poste pour lequel le candidat se présente, l'État dans lequel il se trouve et (le cas échéant) le numéro du district. En outre, ces libellés comportent une petite icône représentant une urne pour signaler le statut du compte en tant que candidat à un poste gouvernemental. Afin de soutenir nos efforts d'identification et de confirmation des candidats et de leurs comptes Twitter de campagne, nous nous sommes associés à Ballotpedia, une organisation politiquement neutre et enregistrée en tant qu'organisation à but non lucratif. Ballotpedia est une encyclopédie en ligne sur la politique et les élections américaines dont l'objectif est d'informer le public en fournissant des informations précises et objectives à tous les niveaux de gouvernement. Dans le contexte de ces élections et en relation avec nos efforts visant à fournir plus de contexte et d'informations aux personnes sur Twitter ainsi qu'à mettre en avant les sources d'information faisant autorité, nous avons également :

- Développé un "Hub" électoral avec des pages d'événements liés, des Moments dédiés aux messages d'intérêt public, des Listes de sujets et un module "Qui suivre" avec des candidats localisés ;
- Géré plus de 573 Moments et événements sélectionnés en anglais et en espagnol pour fournir un contexte et des expériences en direct, pour lesquels 75 millions d'utilisateurs uniques ont visité au moins un de nos Moments sur les élections américaines dans le monde entier en 2020 et 106 millions d'impressions ont été générées pour la page d'événements sur le jour des élections et les résultats des élections ;
- Lancé une nouvelle expérience liée aux tendances aux États-Unis, avec uniquement des sujets contextualisés dans l'onglet "Pour vous";
- Utilisé des Moments contextuels dans la barre de recherche, les libellés, et le fil d'actualité pour fournir des informations fiables contre des affirmations trompeuses qui ont été présentées à 8,5 millions d'utilisateurs par jour pendant la semaine des élections ;
- Créé des incitations à voter et des ressources sur l'inscription des électeurs et le vote anticipé, diffusées de manière passive et active sur Twitter via le fil d'actualité (95,5 millions d'impressions en 2 jours), la barre de recherche, les tendances et les notifications;



- Fourni des centaines de formations à nos partenaires de contenu pour les éduquer et les encourager à créer un contenu électoral fiable et de haute qualité.

### **Dispositifs de vérification des faits (« fact-checking »)**

**3.3 Le Conseil s'interroge sur les dispositifs de *fact-checking* utilisés par l'opérateur, qu'ils soient confiés à des tiers ou mis en œuvre par lui-même. Il souhaite connaître les partenariats et/ou actions (en interne et externe) existants en la matière, notamment aux fins de lutter contre les hypertrucages ou *deep fakes*.**

**Il souhaite connaître en particulier :**

- **les critères retenus par l'opérateur pour sélectionner les *fact-checkers* (notamment : fiabilité, nationalité et langue de travail sur une zone géographique donnée) ;**
- **les critères, sources et outils qu'il utilise, d'une part, s'il procède en interne au *fact-checking* et, d'autre part, si une décision de modération est prise en interne à l'issue de cette vérification ;**
- **si tous les types de contenus (informatifs, publicitaires, écrits, vidéos, sons) sont susceptibles de faire l'objet d'une vérification ou, dans le cas contraire, pourquoi ;**
- **si les contenus sur la plateforme pouvant être soumis aux *fact-checkers* sont laissés à leur libre appréciation ou sélectionnés (le cas échéant, selon quels critères et quelle méthode) ;**
- **l'ensemble des moyens technologiques mis à la disposition des *fact-checkers* dans leur travail de vérification de l'information (notamment : outil de suggestion des contenus à vérifier et, le cas échéant, selon quels critères ; mise à disposition d'une API) ;**
- **s'il existe un dispositif d'urgence en cas d'actualité particulière. L'opérateur pourra donner des exemples de cas d'actualité et précisera, le cas échéant, si la crise sanitaire en 2020 a eu un impact sur ce dispositif ;**
- **le budget consacré aux partenariats et/ou actions concourant à la vérification de l'information sur ses services (notamment les contreparties financières versées aux *fact-checkers*, le cas échéant).**

Les personnes qui utilisent Twitter nous ont clairement dit que nous ne devrions pas déterminer la véracité des Tweets. Au contraire, les internautes sont de plus en plus nombreux à penser que nous devrions fournir plus de contexte pour les aider à se faire leur propre opinion dans les cas où la substance d'un Tweet est contestée. Nous nous concentrons donc sur la fourniture d'un contexte et non sur la vérification des faits ("fact-checking").

En février 2020, c'est l'approche que nous avons choisi d'adopter lors de la formulation de notre politique dédiée à la lutte contre la diffusion de médias manipulés et synthétiques tels que les deep fakes sur Twitter qui ont la capacité de perturber la conversation publique. Nous pensons, en effet, que toute personne doit être en mesure de trouver des informations fiables sur Twitter. Ainsi, les gens doivent pouvoir obtenir les informations et le contexte nécessaires pour comprendre si le contenu qu'ils voient sur Twitter est réel ou fabriqué. Par conséquent, nous pouvons libeller les Tweets qui incluent des médias (vidéos, audio et images) qui ont été modifiés ou fabriqués de manière trompeuse. En outre, il est interdit de partager sur Twitter des médias faussement modifiés d'une manière qui induit en erreur ou trompe les gens quant à l'authenticité du média, lorsque des menaces pour la sécurité physique ou d'autres dommages graves peuvent en résulter.

Twitter n'étant ni un agrégateur d'informations ni un éditeur, il ne propose pas d'outil de vérification des faits aux éditeurs. Cependant, de par sa nature publique, largement distribuée et en temps réel, Twitter est une plateforme importante pour les journalistes, à la fois pour partager librement leur contenu et pour offrir accès aux gens à des informations publiques en temps réel. De nombreux médias et éditeurs parmi les plus fiables au monde utilisent régulièrement Twitter comme plateforme de diffusion de contenu.





Twitter ne recherche pas le contenu des éditeurs de presse et ne l'apporte pas sur la plateforme. Lorsqu'un éditeur de presse (via son propre compte) télécharge son contenu sur Twitter, il a l'intention d'utiliser Twitter comme un canal de distribution et de marketing pour attirer un public par le biais de ses followers et un public plus large par l'engagement avec d'autres personnes sur la plateforme. Ainsi, les organisations de vérification des faits sur Twitter partagent largement leur contenu, recherchent des sujets et répondent à des questions ou à du contenu.

Toutefois, bien que nous ne vérifions pas nous-mêmes les faits, nous nous sommes associés à des organisations fiables de vérification des faits pendant les périodes de risque aigu, qui sont intégrées à notre plateforme en tant que signaleurs de confiance. Ils soutiennent la plateforme en signalant les Tweets qui sont en violation avec nos politiques concernant la manipulation de la plateforme. Un partenaire important dans ce domaine est le EU DisinfoLab, qui a travaillé activement avec nous en période électorale. Le EU DisinfoLab est une organisation non-gouvernementale basée à Bruxelles dont la mission est de lutter contre la désinformation par une méthodologie innovante et un soutien scientifique à la communauté de la contre-désinformation.

**3.4 S'il recourt à des *fact-checkers* tiers, l'opérateur est invité à indiquer au Conseil :**

- la liste des structures de *fact-checking* partenaires de la plateforme ;
- la liste et le nombre de contenus suggérés par la plateforme aux tiers pour vérification et la liste et le nombre de contenus proposés à la vérification par les tiers,
- la liste et le nombre de contenus ayant fait l'objet d'une intervention des *fact-checkers* tiers sur la plateforme (décliné par types d'interventions possibles),
- la nature des liens juridiques unissant l'opérateur aux *fact-checkers* tiers,
- la nature des contreparties qui leur sont offertes.

N/A

**3.5 Si aucun dispositif interne ni externe œuvrant au *fact-checking* n'est mis en œuvre, l'opérateur en précisera les raisons.**

Comme mentionné dans la réponse 3.3, les utilisateurs de Twitter nous ont dit que nous ne devons pas déterminer la véracité des Tweets. Au contraire, ils sont de plus en plus nombreux à penser que nous devrions fournir un contexte supplémentaire pour les aider à se faire leur propre opinion dans les cas où la substance d'un Tweet est contestée. Nous nous concentrons donc sur la fourniture d'un contexte et non sur la vérification des faits ("fact-checking").

**3.6 L'opérateur décrira les procédures et les moyens technologiques et humains permettant de traiter les contenus dont le *fact-checking* (interne ou externe) a montré qu'ils contenaient ou étaient susceptibles de contenir une fausse information, y compris au sein des groupes d'utilisateurs fermés.**

**A cet égard, il est invité à préciser :**

- la chaîne de décision amenant l'opérateur à agir sur un contenu ou un compte ;
- les mesures prises le cas échéant (ex. : déréférencement, réduction de la visibilité des contenus, limitation de certaines fonctionnalités, messages apposés aux contenus) ;
- la gradation de ces mesures et les critères d'application de chacune d'entre elles ;
- si les mesures et leur gradation sont susceptibles de varier selon le type et/ou à la nature du contenu vérifié et, le cas échéant, pourquoi ;
- si l'intervention des *fact-checkers* est directement visible sur le service ;
- si des voies de recours internes sont offertes aux auteurs des contenus concernés et, le cas échéant, si ces recours sont traités par l'opérateur ou par le tiers partenaire.



N/A

**3.7 Le Conseil souhaite également connaître les modalités de communication de l'opérateur, ou les moyens qu'il met en œuvre pour permettre aux *fact-checkers* de communiquer, auprès :**

- des utilisateurs ayant publié un tel contenu, l'ayant partagé ou ayant interagi avec lui (ex. : commentaire, like) ;
- de l'ensemble des utilisateurs en rendant visible, sur le service, les interventions sur un contenu ou un compte liées au *fact checking*.

N/A

### ***Autres questions***

**3.8 Le Conseil souhaite connaître les éventuels freins (notamment technologiques, économiques, et/ou juridiques) rencontrés par l'opérateur dans l'identification des sources des contenus issus des entreprises et des agences de presse et SCA, la mise en avant de leurs contenus, la participation à la vérification de l'information ou la collaboration avec les *fact-checkers* tiers.**

**L'opérateur pourra également mentionner les limites qu'il voit aux approches de labellisation et de bonnes pratiques visées au point 3.2.**

N/A

**3.9 L'opérateur indiquera dans quelle mesure il a tenu compte des recommandations du Conseil énoncées dans son bilan relatif à l'exercice 2019 à compter de sa publication en juillet 2020.**

Comme mentionné tout au long de cette section, les utilisateurs de Twitter nous ont dit que nous ne devrions pas déterminer la véracité des Tweets. Nous voulons nous concentrer sur la mise en contexte, et non sur la vérification des faits. C'est pourquoi nous nous sommes efforcés de fournir davantage de contexte pour aider les gens à se faire leur propre opinion dans les cas où la substance d'un Tweet est contestée. Nous nous engageons à poursuivre ce travail important de manière ouverte, en travaillant avec un large éventail d'acteurs et en recueillant leurs commentaires sur nos politiques et les fonctionnalités de nos produits.

**3.10 L'opérateur est invité à indiquer au Conseil l'impact de la crise sanitaire sur la promotion des contenus issus d'entreprises SCA et les mesures de *fact-checking* et notamment, les adaptations qu'il a pu être amené à lui apporter en conséquence.**

N/A



## 4. Lutte contre les comptes propageant massivement de fausses informations

### *Identification et définition des pratiques de diffusion massive de fausses informations*

**4.1 Le Conseil souhaite savoir comment l'opérateur appréhende la notion de comptes propageant massivement de fausses informations, telle que mentionnée à l'article 11 de la loi du 22 décembre 2018.**

**L'opérateur déclinera les différentes catégories de pratiques de diffusion massive de fausses informations qu'il identifie sur son service, en en donnant leur définition, leur seuil éventuel de déclenchement et en les caractérisant, au égard notamment de leur origine (type d'acteur), de leurs moyens de diffusion (exclusivement algorithmiques et/ou humains) et de la nature de l'objectif poursuivi par leurs émetteurs (exclusivement économique et/ou politique, ou autre).**

**Il indiquera, pour chacune de ces catégories, s'il s'agit d'une pratique interdite dans les règles d'utilisation de la plateforme.**

En septembre 2020, nous avons mis à jour notre politique en matière de manipulation de la plateforme et de spam. Il est interdit d'utiliser les services de Twitter d'une manière destinée à amplifier ou supprimer artificiellement des informations ou à adopter un comportement qui manipule ou perturbe l'expérience des gens sur Twitter. Nous définissons la manipulation de la plate-forme comme l'utilisation de Twitter pour se livrer à une activité massive, agressive ou trompeuse qui induit les autres en erreur et/ou perturbe leur expérience. La manipulation de la plateforme peut revêtir de nombreuses formes et nos règles sont destinées à traiter un large éventail de comportements interdits, notamment :

- le spam à motivation commerciale, qui vise généralement à attirer du trafic ou de l'attention d'une conversation sur Twitter vers des comptes, des sites Web, des produits, des services ou des initiatives ;
- les engagements inauthentiques, qui tentent de faire apparaître des comptes ou des contenus plus populaires ou actifs qu'ils ne le sont ;
- une activité coordonnée, qui tente d'influencer artificiellement les conversations par l'utilisation de comptes multiples, de faux comptes, de l'automatisation et/ou de scripts ; et
- une activité coordonnée nuisible qui encourage ou promeut un comportement qui viole les Règles de Twitter.

Twitter continue de développer les outils et systèmes de détection nécessaires pour lutter contre l'automatisation malveillante sur notre service. Par exemple, Twitter donne la priorité à l'identification de l'activité suspecte d'un compte, telle qu'un volume exceptionnellement élevé de Tweets avec le même hashtag ou mentionnant le même @handle sans réponse du compte visé. Lorsque nous identifions une telle activité, nous pouvons demander à une personne utilisant le service de confirmer le contrôle humain du compte ou son identité. Twitter a également augmenté son utilisation de défis techniques destinés à identifier les comptes automatisés, tels que les reCAPTCHA, qui demandent aux individus d'identifier des parties d'une image ou de taper des mots affichés à l'écran, ainsi que les demandes de réinitialisation de mot de passe qui protègent les comptes potentiellement compromis.

Au cours du premier semestre 2020, nous avons émis 135 676 973 défis techniques de ce type, soit une augmentation de +54 % par rapport à la période de référence précédente. En plus de cela, nous nous concentrons intensivement sur la dissuasion de la création de faux comptes lors de l'inscription plutôt que de simplement répondre au comportement d'un compte après son inscription.



L'accent que nous mettons simultanément sur la lutte contre l'automatisation malveillante et les abus provenant de notre API a conduit à la suspension de centaines de milliers d'applications au cours de depuis 2019. Cela a été réalisé en utilisant des mesures à la fois automatisées et proactives.

En outre, notre processus amélioré d'onboarding des développeurs est conçu pour empêcher les acteurs malveillants d'accéder à notre plateforme d'API. Ainsi, depuis le lancement du programme en juillet 2018, environ 800 000 cas d'utilisation ont été examinés et vérifiés pour la conformité aux politiques avant que le développeur ne se voie accorder ou refuser son accès.

Notre objectif avec ces changements est de réduire la charge de signalement qui pèse sur les personnes qui utilisent Twitter, notamment en matière de spams et d'automatismes malveillants. Ces améliorations technologiques ont entraîné une réduction correspondante du nombre de signalements de spam par les utilisateurs de Twitter. Les signalements pour du spam sont des signalements de personnes qui utilisent Twitter après avoir reçu une interaction, telle qu'un suivi, une mention ou un message privé, d'un compte suspecté de spam. Au cours du premier semestre de 2020, 5 440 691 signalements de spam ont été soumis par des utilisateurs de Twitter après avoir reçu une interaction avec un compte suspecté de spam.

Nous avons également supprimé les comptes verrouillés du décompte des followers afin que ces chiffres reflètent plus fidèlement la portée et la popularité réelles d'un compte. Les comptes sont verrouillés lorsque nos systèmes détectent une activité inhabituelle et imposent un changement de mot de passe ou un autre défi technique. Si le défi n'a pas été relevé ou si le mot de passe n'a pas été modifié, le compte est immédiatement verrouillé, ce qui l'empêche d'envoyer des Tweets, des retweets ou d'aimer les publications d'autres personnes. La suppression des comptes verrouillés du nombre de followers permet de garantir la fiabilité des informations contenues dans les profils sur Twitter.

Les tactiques de manipulation, de spam et d'abus d'API employées par les acteurs malveillants évoluent constamment, et nous nous engageons à améliorer en permanence nos mesures de réponse et de prévention en travaillant quotidiennement à l'élaboration d'approches améliorées en matière de politique, de détection et d'application.

#### **4.2 L'opérateur est invité à fournir :**

- **la liste ou, à défaut, des exemples d'opérations de propagation de fausses informations soutenues par des entités étatiques ou institutionnelles observées et identifiées sur son (ses) service(s) et les mesures prises à leur égard ; il indiquera le cas échéant, s'il coopère avec les autorités ainsi que d'autres opérateurs dans le but de mieux détecter et de prévenir ces phénomènes ;**
- **des exemples de pratiques individuelles ou à l'échelle d'un petit groupe d'utilisateurs de comportements trompeurs et abusifs massifs (ex. : multidiffusion, publication à répétition) observées et identifiées sur sa (ses) plateforme(s), ainsi que les moyens mis en œuvre pour y lutter.**

Conformément à nos principes de transparence et afin d'améliorer la compréhension par le public des campagnes d'influence inauthentiques, Twitter met à la disposition du public des [archives de Tweets et de médias](#) qui, selon nous, résultent d'opérations d'information liées à des États sur notre service.

Nous pensons qu'il incombe à Twitter de protéger l'intégrité de la conversation publique, notamment en divulguant en temps utile des informations sur les tentatives de manipulation de Twitter par des entités étrangères ou nationales liées à un État pour influencer les élections et d'autres conversations civiques. Nous pensons que le public et la communauté des chercheurs sont mieux informés par la transparence.



En octobre 2018, nous avons lancé la première archive dans l'industrie des opérations potentielles de désinformation étrangère que nous avons vues sur Twitter. Notre conviction fondamentale est que ces comptes doivent être rendus publics et consultables afin que le public, les gouvernements et les chercheurs puissent enquêter, apprendre et renforcer les capacités d'éducation aux médias pour l'avenir.

Depuis 2018, nous avons continué à documenter et à développer ces archives avec plus de 85 000 divulgations de comptes liés à des opérations de désinformation soutenues par des États, et l'année 2020 n'a pas été différente. Au total, en 2020, nous avons ajouté 54 254 comptes à notre archive, doublant notre total de 2019 de 26 627.

### Mars

En mars 2020, nous avons publié un ensemble de données de 70 comptes attribués à des opérations de désinformation liées à un État au Ghana et au Nigeria. En collaboration avec les forces de l'ordre, des pairs du secteur, des journalistes et des chercheurs experts, nous avons suspendu un petit réseau de comptes Tweetant en grande partie en anglais et qui se présentaient comme basés aux États-Unis. Ces 70 comptes supprimés, opérant depuis le Ghana et le Nigeria et que nous avons pu associer de manière fiable à la Russie, tentaient de semer la discorde en engageant des conversations sur des questions sociales, comme la race et les droits civils.

### Avril

En avril, nous avons publié une liste de 20 348 comptes attribués à des opérations de désinformation liées à un État et provenant de Serbie, du Honduras, d'Égypte, d'Indonésie et d'un acteur affilié à l'Arabie saoudite. Nous avons supprimé 2 541 comptes dans un réseau basé en Égypte, connu sous le nom de réseau El Fagr. Ce groupe de médias a créé des comptes inauthentiques pour amplifier les messages critiquant l'Iran, le Qatar et la Turquie. Les informations que nous avons obtenues de l'extérieur indiquent qu'il recevait des instructions du gouvernement égyptien. Nous avons découvert que de nombreux comptes inauthentiques accédaient à Twitter à partir d'une seule plage d'adresses IP au Honduras et retweetaient massivement le compte du président. Nous avons supprimé 3 104 comptes lorsqu'il est apparu qu'un membre du personnel avait créé les faux comptes au nom du gouvernement. À la suite d'une enquête découlant d'un rapport de @Bellingcat sur une opération d'information en Indonésie visant le mouvement pour l'indépendance de la Papouasie occidentale, nous avons supprimé 795 faux comptes qui diffusaient du contenu provenant de sites d'information suspects et qui promouvaient du contenu pro-gouvernemental. Enfin, un réseau de comptes associés à l'Arabie saoudite et opérant à partir de plusieurs pays, dont l'Arabie saoudite, l'Égypte et les Émirats arabes unis, diffusait du contenu faisant l'éloge des dirigeants saoudiens et critiquant le Qatar et l'activité de la Turquie au Yémen. Au total, 5 350 comptes ont été supprimés.

### Juin

En juin, nous avons divulgué 32 242 comptes liés à trois opérations distinctes que nous avons attribuées respectivement à la République populaire de Chine (RPC), à la Russie et à la Turquie. Chaque compte et chaque élément de contenu associé à ces opérations a été définitivement supprimé du service. En outre, nous avons partagé les données pertinentes de cette divulgation avec deux partenaires de recherche de premier plan : l'Australian Strategic Policy Institute (ASPI) et le Stanford Internet Observatory (SIO). En RPC, notre divulgation concernait deux ensembles de comptes interconnectés. Tout d'abord, 23 750 comptes constituaient le cœur du réseau, c'est-à-dire le réseau central très engagé. Deuxièmement, environ 150 000 comptes étaient conçus pour stimuler ce contenu, c'est-à-dire les amplificateurs. Suite aux commentaires des chercheurs sur nos divulgations antérieures, selon lesquels nous devons mieux affiner le processus de divulgation pour permettre une enquête efficace sur l'activité principale, nous n'avons pas inclus les 150 000 comptes d'amplificateurs dans les archives publiques. En Russie, aidés



en partie par le partage d'informations utiles de la part de chercheurs externes et de nos pairs, nous avons enquêté sur les comptes associés à Current Policy, un site web de médias qui se livre à une propagande politique soutenue par le gouvernement russe. Un réseau de comptes liés à cette opération médiatique a été suspendu pour violation de notre politique de manipulation de la plateforme, en particulier pour publication croisée et amplification de contenu de manière inauthentique et coordonnée à des fins politiques. Les activités comprenaient la promotion du parti Russie Unie et l'attaque de dissidents politiques. Nous avons divulgué les 1 152 comptes et médias associés dans nos archives publiques. Enfin, en Turquie, le réseau de comptes avait recours à une activité inauthentique coordonnée, qui visait principalement les publics nationaux en Turquie. Sur la base de notre analyse des indicateurs techniques du réseau et des comportements des comptes, la collection de comptes faux et compromis était utilisée pour amplifier les récits politiques favorables à l'AKP, et démontrait un fort soutien au Président Erdogan. Nous avons divulgué 7 340 comptes dans nos archives.

## Octobre

En octobre, nous avons publié un ensemble de données de 1 594 comptes relatifs à des opérations d'information indépendantes que nous avons attribuées à l'Iran, à l'Arabie saoudite, à Cuba, à la Thaïlande et à la Russie. En Iran, sur la base d'informations partagées par des pairs du secteur, nous avons identifié un réseau de comptes principalement compromis sur Twitter opérant depuis l'Iran, qui a amplifié artificiellement les conversations sur des sujets politiquement sensibles, notamment Black Lives Matter (BLM), le meurtre de George Floyd et d'autres questions de justice raciale et sociale aux États-Unis. Dans les cas où nous avons identifié qu'un compte avait été compromis dans le but d'adopter un comportement violent, nous avons cherché à rétablir le compte auprès de son titulaire initial. Nous avons pu le faire avec succès dans la plupart des cas. En tout, nous avons suspendu un total de 104 comptes liés à cette campagne. En Arabie saoudite, nous avons enquêté et supprimé 33 comptes originaires de ce pays et ayant des liens avec le gouvernement saoudien. Ces comptes ont été créés pour usurper l'identité de personnalités politiques qatariques clés et pour promouvoir des récits sur la politique qatarique qui sont géostratégiquement favorables aux autorités saoudiennes. À Cuba, nous avons suspendu définitivement 526 faux comptes gérés par des organisations de jeunesse liées au gouvernement cubain, notamment l'Unión de Jóvenes Comunistas (UJC) et la Federación Estudiantil Universitaria (FEU). En Thaïlande, notre enquête a mis au jour un réseau de comptes participant à des opérations d'information que nous avons pu relier de manière fiable à l'Armée royale thaïlandaise (RTA). Ces comptes amplifiaient des contenus pro-RTA et pro-gouvernementaux et ciblaient des personnalités de l'opposition politique. Nous avons divulgué 926 comptes et continué à prendre des mesures contre les activités à petite échelle associées à ce réseau, au fur et à mesure que nous les avons identifiées. En Russie, en collaboration avec des pairs de l'industrie et le FBI, nous avons été informés de l'existence d'un certain nombre de comptes qui avaient des liens potentiels avec une fausse agence de presse appelée PeaceData. Ces comptes ont été immédiatement supprimés du service et nous avons marqué tous les liens vers le site comme non sûrs dans l'ensemble du service.

L'objectif de ces divulgations est de continuer à faire comprendre au public les moyens par lesquels les acteurs étatiques tentent d'abuser et de saper les conversations démocratiques ouvertes. Pour poursuivre cette mission, nous avons organisé cet été une [conférence en ligne](#) avec la Fondation Carnegie qui a réuni des experts du secteur pour discuter de la poursuite de notre collaboration et des moyens d'approfondir la recherche sur ces questions essentielles. Nous continuerons à faciliter ces forums en 2021.

Nous pensons qu'il est de notre responsabilité de protéger l'intégrité de la conversation publique et de demander des comptes aux pouvoirs publics. Nous continuerons à mettre à jour nos archives afin que le public, les journalistes et la communauté des chercheurs puissent accéder à ces réseaux et les analyser.

Chacun peut consulter les mises à jour régulières directement depuis le compte @TwitterSafety.



## ***Moyens de lutte contre la diffusion massive de fausses informations***

**4.3 Pour chacune des catégories de pratiques de diffusion massive de fausses informations identifiées en 4.1, l'opérateur indiquera les mesures et moyens (humains et technologiques) qu'il a pris afin de :**

- **détecter de telles pratiques et les comptes qui y ont recours,**
- **faire obstacle à ces pratiques et à ces comptes s'agissant :**
  - **de la procédure d'intervention à leur rencontre : mesures prises (avertissement, restriction des droits, mise en quarantaine, déréférencement, suppression du contenu, du compte, etc.), gradation de ces mesures et critères déclenchant l'application de chacune d'elle ;**
  - **de la lutte contre l'accélération et la viralité des contenus contenant de fausses informations.**

Sur la base de notre politique en matière de manipulation de la plateforme et de spam, tout le monde peut signaler des comptes ou des Tweets via notre canal de signalement dédié ou via notre [formulaire de signalement de spam](#). Ces signalements sont utilisés dans leur ensemble pour nous aider à affiner nos systèmes d'application et à identifier les tendances et les modèles de comportement nouveaux et émergents. Nous utilisons une combinaison de technologie et d'examen humain pour déceler une série de comportements dans les domaines suivants :

### 1. Comptes et identité

Les personnes ne peuvent pas induire en erreur les autres sur Twitter en exploitant de faux comptes. Cela inclut l'utilisation d'informations de compte trompeuses pour se livrer à du spamming, à des comportements abusifs ou perturbateurs. Voici quelques-uns des facteurs que nous prenons en considération :

- l'utilisation de photos de profil stockées ou volées, notamment celles représentant d'autres personnes ;
- l'utilisation de biographies de profil volées ou copiées ;
- l'utilisation d'informations de profil intentionnellement trompeuses, notamment l'emplacement du profil.

Les personnes ne peuvent pas amplifier ou perturber artificiellement les conversations par l'utilisation de plusieurs comptes ou en se coordonnant avec d'autres pour violer les Règles de Twitter. Cela inclut :

- les comptes qui se chevauchent - l'exploitation de plusieurs comptes dont les cas d'utilisation se chevauchent, tels que des personnes identiques ou similaires ou des contenus substantiellement similaires ;
- les comptes interagissant mutuellement - exploitation de plusieurs comptes qui interagissent les uns avec les autres afin de gonfler ou de manipuler la proéminence de Tweets ou de comptes spécifiques ;
- la coordination - création de plusieurs comptes pour publier du contenu dupliqué ou créer un faux engagement, notamment :
  - la publication de Tweets ou de hashtags identiques ou substantiellement similaires à partir de plusieurs comptes que vous exploitez ;
  - l'engagement (Retweets, Likes, mentions, votes Twitter Poll) de manière répétée avec les mêmes Tweets ou comptes à partir de plusieurs comptes que vous exploitez ;
  - la coordination avec d'autres personnes ou les rémunérer pour qu'elles se livrent à un engagement ou à une amplification artificielle, même si les personnes concernées n'utilisent qu'un seul compte ;



- la coordination avec d'autres personnes pour commettre ou promouvoir des violations des Règles de Twitter, y compris des violations de notre politique de comportement abusif.

## 2. Engagement et mesures

Les personnes ne peuvent pas gonfler artificiellement leur nombre de followers ou leur engagement, ni ceux d'autres personnes. Cela inclut :

- la vente ou l'achat de Tweet ou l'inflation des métriques de compte - la vente ou l'achat de followers ou d'engagements (Retweets, Likes, mentions, votes dans les sondages Twitter) ;
- les applications - l'utilisation ou la promotion de services ou d'applications de tiers qui prétendent ajouter des adeptes ou des engagements à des Tweets ;
- l'inflation réciproque - échanger ou coordonner l'échange de followers ou d'engagements dans les Tweets (y compris, mais sans s'y limiter, la participation à des "trains de followers", des "decks" et des comportements de "Retweet pour Retweet") ; et
- transfert ou vente de comptes - vente, achat, échange ou offre de vente, d'achat ou d'échange de comptes Twitter, de noms d'utilisateur ou d'accès temporaire à des comptes Twitter.

## 3. Utilisation abusive des fonctionnalités des produits Twitter

Il est interdit d'utiliser abusivement les fonctionnalités des produits Twitter pour perturber l'expérience des autres. Cela inclut :

### Tweets et Messages privés

- l'envoi massif de réponses, de mentions ou de messages privés non sollicités, agressifs ou à fort volume ;
- la publication et la suppression répétées du même contenu ;
- la publication répétée de Tweets identiques ou presque identiques, ou l'envoi répété de Messages directs identiques ; et
- la publication répétée de Tweets ou l'envoi répété de messages privés consistant en des liens partagés sans commentaire, de sorte que cela constitue l'essentiel de votre activité de Tweet/message privé.

### Suivi

- "follow churn" (suivre puis abandonner un grand nombre de comptes dans le but de gonfler son propre nombre d'abonnés) ;
- suivre sans discernement - suivre et/ou supprimer un grand nombre de comptes sans rapport entre eux dans un court laps de temps, en particulier par des moyens automatisés ;
- dupliquer les followers d'un autre compte, notamment par des moyens automatisés.

### Engagement

- créer de l'engagement de manière agressive ou automatique avec des Tweets pour attirer le trafic ou l'attention sur des comptes, des sites web, des produits, des services ou des initiatives.
- l'ajout agressif d'utilisateurs à des Listes ou à des Moments.

### Hashtags

- l'utilisation d'un hashtag tendance ou populaire dans l'intention de subvertir ou de manipuler une conversation ou d'attirer le trafic ou l'attention sur des comptes, des sites web, des produits, des services ou des initiatives ; et
- l'utilisation d'un nombre excessif de hashtags sans rapport avec le sujet dans un seul Tweet ou dans plusieurs Tweets.





## URLs

- publier ou établir un lien vers un contenu malveillant destiné à endommager ou à perturber le navigateur (malware) ou l'ordinateur d'une autre personne ou à compromettre la vie privée d'une personne (phishing) ;
- la publication de liens trompeurs ou mensongers, par exemple des liens d'affiliation et des liens de détournement de clics.

Comme nous l'avons mentionné au point 1.3.2, nous utilisons une combinaison d'apprentissage automatique et d'examen humain pour appliquer ces règles. Nos systèmes sont capables de présenter du contenu aux modérateurs humains qui utilisent un contexte important pour prendre des décisions sur toutes les violations potentielles des Règles de Twitter.

#### **4.4 Le Conseil souhaite disposer de données chiffrées sur les mesures de lutte prises par l'opérateur contre les comptes propageant massivement de fausses informations, notamment, pour chacune des catégories de pratiques identifiées en 4.1 :**

- le nombre de comptes propageant massivement de fausses informations détectés sur le service, et leur proportion dans le nombre total de comptes sur le service ;
- les origines de cette détection (moyens technologiques, signalements par des tiers, signalements en interne) et leur répartition ;
- parmi les comptes détectés, le nombre de ceux ayant fait l'objet d'une instruction par l'opérateur et de ceux ayant fait l'objet de mesures (en déclinant par types de mesures) ;
- les moyens financiers et humains mis en œuvre par l'opérateur pour lutter contre les comptes propageant massivement de fausses informations ;
- dans le cas où ces comptes auraient fait l'objet de recettes, la rémunération tirée de ces comptes et celle reversée à leur titulaire en 2020, ainsi que les mesures éventuellement prises pour limiter cette rémunération pendant et après la phase d'instruction d'un compte détecté ou signalé comme propageant massivement de fausses informations.

Sur l'ensemble de l'année 2020, nous avons identifié et pris des mesures à l'égard de 54 254 comptes liés à des opérations de désinformation provenant de divers pays et qui tentaient de diffuser de fausses informations sur Twitter. Notre propre technologie interne a joué un rôle essentiel dans la détection et l'action sur ces réseaux de comptes. Cependant, nous avons aussi grandement bénéficié de notre étroite collaboration avec les forces de l'ordre, les pairs du secteur, les journalistes et les chercheurs experts pour partager des informations, comprendre le contexte local et s'attaquer aux nouvelles tactiques émergentes utilisées par les acteurs malveillants.

#### ***Informations aux utilisateurs sur la diffusion massive de fausses informations***

#### **4.5 L'opérateur décrira les dispositifs de suivi et de statistiques des mesures de détection et de traitement des pratiques et comptes propageant massivement de fausses informations, et notamment :**

- si ces dispositifs sont à usage interne ou s'ils sont accessibles aux utilisateurs, et selon quels forme et niveau de détail ;
- les informations suivies et mesurées : définition(s) décomptes propageant massivement de fausses informations ; nombre de comptes détectés, signalés, traités (par des moyens humains d'une part, et des outils technologiques d'autre part) ; caractéristiques des comptes et des contenus concernés ; actions prises ; résultats ;
- les modalités de consultation et de téléchargement de ces informations par l'utilisateur.



La transparence est au cœur du fonctionnement de Twitter. Notre [rapport de transparence](#) mis à jour fournit des détails sur les opérations de désinformation et sur le nombre de comptes qui sont contestés par des systèmes automatisés en tant que spam potentiel.

Pour permettre aux universitaires et au public de mieux comprendre les opérations de désinformation dans le monde, et pour permettre à des tiers indépendants d'examiner ces tactiques sur notre plateforme, nous publions des archives complètes des opérations d'information soutenues par des États sur Twitter. Nous avons mis à disposition tous les comptes et le contenu connexe associés à des opérations de désinformation potentielles que nous avons trouvées sur notre service depuis 2016 et continuons à publier de nouvelles données à mesure que nous détectons de nouvelles activités.

**4.6 L'opérateur indiquera si le service dispose d'un espace d'information renseignant les utilisateurs sur les pratiques liées à l'utilisation de comptes propageant massivement de fausses informations susceptibles d'entraîner une intervention de sa part et, le cas échéant :**

- **l'information des utilisateurs sur les risques encourus notamment dans les cas de création de comptes dans des volumes anormaux, de partage de contenus à des fréquences anormales et d'utilisation de renseignements faux, volés ou trompeurs ;**
- **la manière dont cette information est fournie à l'utilisateur : quelle forme prend-elle, où se situe-t-elle, est-elle claire et accessible ? est-elle notifiée à l'utilisateur (ou nécessite-t-elle une recherche de sa part) et, le cas échéant, dans quelle(s) situation(s) ?**

Dans notre Centre d'assistance, nous fournissons une description détaillée des comportements interdits au regard de nos politiques en matière de manipulation de la plateforme et de spam. Une section intitulée "Que se passe-t-il si vous violez cette politique ?" indique également quelles sont les conséquences des violations susmentionnées. Comme pour toutes les règles de Twitter, ces explications ont été mises à jour et affinées au fil des ans par un langage simple et clair et par des exemples de ces comportements.

### **Autres questions**

**4.7 L'opérateur décrira les mesures et moyens mis en œuvre pour détecter les pratiques de deepfakes, les comptes y ayant recours et lutter contre ce phénomène.**

**Il indiquera le nombre de deepfakes propagés à des fins de diffuser de fausses informations sur son service dont il a eu connaissance.**

En février 2020, nous avons annoncé notre politique contre les médias synthétiques et manipulés. Nous avons été clairs sur le fait qu'il est interdit de partager de manière trompeuse sur Twitter des médias synthétiques ou manipulés qui sont susceptibles de causer des dommages. Dans le cadre de cette règle, une fois qu'un contenu potentiellement en infraction est porté à l'attention de nos équipes par des signalements d'utilisateurs, nous utilisons les critères suivants pour examiner les Tweets et les médias à libeller ou à supprimer dans le cadre de cette règle :

1. Le média est-il synthétique ou manipulé ?

Pour déterminer si le médias a été modifié ou fabriqué de manière significative et trompeuse, nous prenons en compte certains facteurs, notamment :

- Si le contenu a été substantiellement édité d'une manière qui modifie fondamentalement sa composition, sa séquence, son timing ou son cadrage ;
- Toute information visuelle ou auditive (telle que de nouvelles images vidéo, du son superposé ou des sous-titres modifiés) qui a été ajoutée ou supprimée ;



- si le média représentant une personne réelle a été fabriqué ou simulé.
2. Le média est-il partagé de manière trompeuse ?

Nous examinerons également si le contexte dans lequel le média est partagé pourrait entraîner une confusion ou un malentendu ou s'il suggère une intention délibérée de tromper les gens sur la nature ou l'origine du contenu, par exemple, en prétendant faussement qu'il représente la réalité.

Nous évaluons également le contexte fourni avec les médias, par exemple :

- le texte du Tweet accompagnant ou faisant partie du média
  - les métadonnées associées au média
  - des informations sur le profil de la personne qui partage le média
  - les sites web liés au profil de la personne partageant le média ou dans le Tweet partageant le média.
3. Le contenu est-il susceptible d'avoir un impact sur la sécurité publique ou de causer un préjudice grave ?

Les Tweets qui partagent des médias synthétiques et manipulés peuvent être supprimés en vertu de cette politique s'ils sont susceptibles de causer un préjudice. Parmi les préjudices spécifiques que nous prenons en compte :

- les menaces pour la sécurité physique d'une personne ou d'un groupe
- le risque de violence de masse ou de troubles civils généralisés
- les menaces pour la vie privée ou la capacité d'une personne ou d'un groupe à s'exprimer librement ou à participer à des événements civiques, par exemple : harcèlement ou attention non désirée et obsessionnelle ; contenu ciblé comprenant des tropes, des épithètes ou des éléments visant à réduire quelqu'un au silence ; suppression ou intimidation d'électeurs.

Identifier les médias synthétiques et manipulés est un vrai défi et nous nous engageons à maintenir nos efforts et nos investissements dans ce domaine. La mise à jour de nos règles de manière publique et à travers une participation démocratique restera au cœur de notre approche.

**4.8 L'opérateur décrira les pratiques dont il a eu connaissance consistant à générer un trafic ou un engagement artificiel massif pour propager fortement un contenu en augmentant sa viralité (achat/vente de faux abonnés, fraude aux clics...) ainsi que les mesures mises en œuvre à leur encontre.**

**Il indiquera, parmi les contenus connus comme contenant de fausses informations, combien ont bénéficié de ces pratiques.**

Notre politique en matière de manipulation de la plateforme et de spam indique clairement qu'il est interdit de gonfler artificiellement le nombre de followers ou d'engagements, que ce soit les siens ou ceux d'autres personnes. Cela inclut, sans s'y limiter, des comportements tels que la vente ou l'achat de tweets ou de comptes gonflés, la vente ou l'achat de followers ou d'engagements (retweets, likes, mentions, votes sur des sondages Twitter) ou l'utilisation ou la promotion de services ou d'applications tiers qui prétendent ajouter des followers ou des engagements aux Tweets.

**4.9 L'opérateur indiquera dans quelle mesure il a tenu compte des recommandations du Conseil énoncées dans son bilan relatif à l'exercice 2019 à compter de sa publication en juillet 2020.**



Dans ses recommandations, le CSA a d'abord indiqué qu'il souhaitait en savoir plus sur les outils et les processus de traitement des comptes qui diffusent massivement des fausses informations. Nous nous sommes donc efforcés de fournir le plus d'informations possible sur cette question importante liée aux opérations de désinformation soutenues par des États, tout en gardant à l'esprit le souci de la sécurité de la plateforme et des personnes. Un deuxième axe de travail défini par le CSA était fondé sur l'information de la part des plateformes quant aux comptes qui mènent des opérations de désinformation malveillantes. Par conséquent, pour permettre aux universitaires et au public de mieux comprendre les opérations désinformation dans le monde, et pour permettre un examen indépendant par des tiers de ces tactiques sur notre plateforme, nous avons continué à divulguer en 2020 des archives complètes des opérations d'information soutenues par des États sur Twitter. Nous avons mis à disposition tous les comptes et contenus associés à des opérations de désinformation potentielles que nous avons trouvées sur notre service et nous continuerons à publier de nouvelles données à mesure que nous détectons de nouvelles activités.

#### **4.10 L'opérateur indiquera, dans le cadre de sa lutte globale contre les comptes propageant massivement de fausses informations, les nouvelles mesures initiées en 2020, notamment celles prises dans le cadre de la crise sanitaire mondiale.**

Alors que le monde entier faisait face - et continue de faire face - à une urgence de santé publique sans précédent, nous avons fait preuve de transparence et d'ouverture quant aux nouveaux défis auxquels nous avons été confrontés pour servir la conversation publique en cette période critique. Cette approche s'est appuyée sur le travail précieux de nos partenaires de confiance, notamment les autorités de santé publique, les organisations non gouvernementales et les gouvernements du monde entier.

Par conséquent, nos mesures visant à protéger la conversation publique se sont concentrées sur deux piliers majeurs : 1) l'adaptation de nos politiques pour faire face aux menaces émergentes, notamment en ce qui concerne la COVID-19, les médias synthétiques et manipulés et les élections ; 2) le maintien de notre service, la poursuite de l'application de nos règles et la protection de la plateforme dans le contexte de la transition de nos équipes vers le télétravail.



## 5. Mesures de lutte contre les fausses informations en matière de communications commerciales et de promotion des contenus d'information se rattachant à un débat d'intérêt général

### 5.a. Focus : lutte contre la diffusion de fausses informations en lien avec des communications commerciales

#### *Catégorisation des communications commerciales*

**5.1 L'opérateur est invité à commenter les catégories d'annonces publicitaires, de contenus sponsorisés et de contenus en partenariat telles qu'établies au début du questionnaire.**

**Il indiquera si cette catégorisation lui semble pertinente, notamment au regard du modèle de sa plateforme et dans le cadre particulier de la lutte contre les fausses informations. Il pourra identifier les limites éventuelles de ces catégories et formuler des propositions pour que les catégories retenues couvrent l'ensemble de l'activité publicitaire sur son service.**

Les solutions Twitter Ads sont conçues pour aider les annonceurs à atteindre leurs objectifs de marketing, qu'il s'agisse d'accroître la notoriété de leur marque, de susciter la considération de la part des consommateurs ou de générer des conversions (par exemple, des clics). Twitter Ads propose plus de 20 options de produits pour mettre en valeur le contenu et les médias sur Twitter dans le format le plus adapté et pertinent. Bien qu'ils puissent être utilisés de manière interchangeable, la plupart de ces produits correspondent directement à un objectif de campagne et à un but marketing précis afin d'en maximiser les résultats. C'est pourquoi Twitter utilise une classification légèrement différente de celle décrite par le CSA, adaptée aux spécificités des solutions Twitter Ads. Vous trouverez ci-dessous une liste des solutions publicitaires de Twitter :

- Tweets sponsorisés (texte, image, GIF)
- Vidéos (Vidéo sponsorisée, Amplify Pre-roll, Amplify Sponsorship, First View, Vidéo live sponsorisée)
- Cartes (Apps, Carrousels, Message privé, Conversation)
- Branding (emojis, rappels, Moments sponsorisés, Tendances sponsorisées, Spotlight)

Veillez noter que sur Twitter, le contenu "sponsorisé" fait référence aux produits publicitaires Twitter Ads. Seule la solution "Amplify Publisher Program" relèverait de la catégorie "contenu en partenariat", telle que définie par le Conseil. Par conséquent, nous pensons que toute catégorisation des communications commerciales doit tenir compte de la nature et de la taille spécifiques des services et des solutions publicitaires de chaque plateforme.

#### *Annonces publicitaires et contenus sponsorisés*

**5.2 L'opérateur décrira les modalités de vente des espaces publicitaires sur le service : le placement est-il régi par des mécanismes programmatiques ? Repose-t-il sur un modèle fermé propre à la plateforme ou sur un modèle d'enchères en temps réel (*real-time bidding*) ? Le cas échéant, quels sont les intermédiaires utilisés ?**

Twitter utilise un système d'enchères dans lequel les annonceurs placent des offres les uns contre les autres pour diffuser des publicités à un public cible. Toutefois, à la différence d'une enchère traditionnelle, les enchères des publicités Twitter tiennent compte non seulement du prix que l'annonceur est prêt à payer, mais aussi de la qualité de la publicité. Chaque fois qu'une publicité remporte une enchère, elle est diffusée une fois à une personne sur Twitter, générant ainsi une "impression".



Les publicités entrent automatiquement dans le système d'enchères une fois la campagne lancée, et les enchères auxquelles les publicités participent sont déterminées par les critères de ciblage de la campagne. Par exemple, si un annonceur cible l'intérêt "basketball", ses annonces seront en concurrence au sein des enchères pour être diffusées auprès des personnes sur Twitter qui s'intéressent au basketball. Notre algorithme d'enchères comparera en temps réel la publicité en question à d'autres publicités qui ciblent également cet intérêt, et la publicité qui remporte l'enchère sera montrée à une personne. Ensuite, si la personne qui voit l'annonce effectue une action dite "facturable" sur la campagne (par exemple, elle clique sur l'annonce), Twitter facturera l'annonceur pour cette action en utilisant un modèle d'enchères dit de "second prix" (l'annonceur ayant formulé la meilleure offre remporte l'enchère, mais paie le prix proposé par le deuxième annonceur le plus offrant).

L'algorithme d'enchères tient également compte de la qualité de la publicité. Chaque publicité se voit attribuer un "score de qualité", qui est utilisé conjointement avec l'offre pour participer aux enchères. Le score de qualité est déterminé en prenant en compte des facteurs tels que :

- La résonance : les internautes s'intéressent-ils à l'annonce ? Est-ce qu'ils cliquent, aiment et/ou retweetent ?
- La pertinence : le contenu du Tweet correspond-il aux intérêts de l'audience ciblée ?
- La caractéristique récente : les Tweets sont-ils récents ? Twitter est l'actualité du moment, les enchères récompensent donc le contenu récent.

Pour remporter les enchères et maximiser le succès d'une campagne, les annonces doivent avoir à la fois un bon score de qualité et une bonne enchère. Cette combinaison constitue le "score publicitaire", qui détermine qui remporte l'enchère et diffuse les annonces sur notre plateforme.

**5.3 L'opérateur précisera les moyens, technologiques ou autres, permettant aux annonces publicitaires et contenus sponsorisés de gagner en visibilité sur la plateforme (possibilité pour l'annonceur de viser un public en particulier, de choisir le format du contenu, avec quelles contreparties ; impact des interactions des utilisateurs avec le contenu ; etc.).**

Les publicités Twitter sont des produits publicitaires payants que les marques utilisent pour amplifier leurs messages, cibler des clients potentiels et développer leur base d'abonnés. Les contenus sponsorisés par les services publicitaires de Twitter se distinguent clairement des contenus organiques et sont marqués d'une icône "Sponsorisé". Les internautes peuvent interagir avec la plupart des contenus promus de la même manière que les contenus organiques. Les solutions Twitter Ads offrent aux annonceurs différents formats publicitaires (allant des Tweets, comptes et tendances sponsorisés aux vidéos et Cards sponsorisées disponibles dans le fil d'actualité ou dans la barre de recherche). Les annonceurs ont également la possibilité de cibler l'audience souhaitée en fonction de critères démographiques spécifiques (lieu, langue, type d'appareil, âge et sexe) et de types d'audience (centres d'intérêt, types d'abonnés, conversations, événements, mots-clés, etc.) D'autre part, les gens peuvent voir les détails des raisons pour lesquelles ils ont reçu une annonce spécifique en cliquant sur l'annonce. Les informations partagées renvoient également l'utilisateur à ses propres paramètres afin de contrôler s'il reçoit ou non des publicités basées sur ses intérêts déduits par Twitter en utilisant le paramètre "Personnaliser les publicités". Les [politiques de Twitter en matière de publicité](#) sont facilement accessibles au public et sont adaptées à la langue et aux restrictions juridiques locales.

**5.4 Dans la mesure où le dispositif différencierait du dispositif de signalement général, le Conseil souhaite que l'opérateur lui indique si l'utilisateur peut signaler une annonce publicitaire ou un contenu sponsorisé pour « fausse information » (ou équivalent).**

**Le cas échéant, l'opérateur indiquera :**

- les modalités de signalement (faut-il avoir un compte pour signaler le contenu, quelle est la démarche pour signaler ces contenus...) ;



- dans quelles conditions et comment l'opérateur interagit avec les utilisateurs concernés par ce signalement (auteur du contenu et auteur du signalement ; information concernant l'avancée du traitement du signalement) ;
- le nombre de signalements reçus et traités relatifs à des annonces publicitaires et contenus sponsorisés et, plus particulièrement, aux contenus promus d'information se rattachant à un débat d'intérêt général, en indiquant, quand cela est possible, le nombre de signalements pour des raisons de « fausses informations » ;
- les moyens (technologiques, humains et financiers) mis en œuvre pour traiter ces signalements.

Lorsque les annonceurs choisissent de promouvoir leur contenu avec Twitter Ads, leur compte et leur contenu sont soumis à une procédure d'approbation. Cette procédure d'approbation est conçue pour garantir la qualité et la sécurité de la plateforme Twitter Ads.

Bien que les annonceurs soient responsables du contenu disponible sur leur profil, de leurs Tweets et de toutes les sponsorisations qu'ils effectuent sur la plateforme Twitter Ads, nous examinons toutes les violations signalées de notre politique relative aux publicités Twitter, des règles Twitter et des conditions d'utilisation et prenons des mesures en conséquence. À cette fin, Twitter offre aux gens deux options pour signaler les publicités. Premièrement, Twitter dispose d'un [formulaire de signalement des publicités](#), accessible à tous (et non aux seuls titulaires de comptes). Deuxièmement, nous autorisons les signalements de publicités par les personnes qui se trouvent sur Twitter directement au niveau de la publicité elle-même et nous prenons des mesures si les publicités sont en violation des politiques de Twitter en matière de publicités. Dans le cadre de cette méthode de signalement, il existe 5 motifs de signalement existants : "Contenu sexuel pour adultes", "Produits et services illégaux", "Obscénité", "Publicité politique" et "Autre". Parmi "Autre", Twitter interdit les publicités qui contiennent des contenus inappropriés à l'échelle mondiale, notamment les contenus déformés, les contenus synthétiques trompeurs ou manipulés et, entre autres, les sujets ou événements sensibles (par exemple, les décès, les catastrophes naturelles/industrielles, les attaques violentes, les troubles civils, etc.) L'appropriation de hashtags sensibles liés à des événements est également interdite.

En outre, nous avons décidé de restreindre la publicité contenant une référence implicite ou explicite à la COVID-19. Plus précisément, la publicité contenant une référence implicite ou explicite à la COVID-19 est autorisée lorsqu'elle fait référence à des ajustements en termes de pratiques et/ou de modèles commerciaux en réponse à la COVID-19 et au soutien aux clients et aux employés lié à la COVID-19, avec les restrictions suivantes :

- les références de mauvais goût à la COVID-19 sont interdites
- le contenu ne doit pas être sensationnaliste ou susceptible d'inciter à la panique
- les prix des produits liés à la COVID-19 ne doivent pas être gonflés
- la promotion de certains produits liés à la COVID-19 peut être interdite
  - Nous interdisons actuellement la publicité pour les masques médicaux et les désinfectants hydroalcooliques pour les mains (avec ou sans référence à la COVID). D'autres produits peuvent être ajoutés à cette liste et l'application peut être rétroactive.
- la mention de vaccins, de traitements et de kits de test est autorisée, uniquement sous forme d'information, par les éditeurs de presse qui ont été exemptés en vertu de notre politique relative aux publicités politiques.

Toutefois, les messages d'intérêt public relatifs à la COVID-19 émanant de gouvernements et d'entités supranationales (par exemple, l'Organisation mondiale de la santé), ainsi que de partenaires de confiance, sont autorisés.

**5.5 L'opérateur précisera les mesures mises en œuvre afin de détecter les annonces publicitaires et contenus sponsorisés porteurs de fausses informations, notamment :**



- **l'existence d'un système de vérification des annonceurs, interne ou externe ;**
- **les mesures de détection et de lutte contre la publicité en faveur de sites internet spécialisés dans les fausses informations. L'opérateur indiquera s'il dispose d'informations sur ces sites, notamment sur leur origine et leur typologie, ainsi que sur la durée d'apparition d'une publicité pour ces sites sur le service ;**
- **les moyens de coopération avec d'autres opérateurs ou des organismes externes afin de détecter et de signaler les annonces publicitaires et contenus sponsorisés porteurs de fausses informations et s'il s'appuie-t-il sur des normes ou des organismes externes afin d'identifier et de signaler ces contenus**

La politique publicitaire de Twitter s'articule autour de plusieurs principes clés : assurer la sécurité des personnes, promouvoir des contenus honnêtes et les cibler de manière responsable, interdire la diffusion de spams, de contenus préjudiciables ou d'autres contenus perturbateurs, fixer des normes éditoriales élevées pour les contenus Twitter Ads créés, et rester informé des processus Twitter Ads qui soutiennent ces politiques.

Dans le cas des publicités commerciales, avant qu'un compte puisse faire de la publicité sur Twitter, il doit répondre à des critères spécifiques pour être éligible. Par exemple, les publicités doivent inclure une URL fonctionnelle dans leur biographie, elles doivent représenter la marque et le produit sponsorisé et ne peuvent pas inciter les gens à ouvrir un contenu en incluant un langage exagéré ou sensationnaliste ou des appels à l'action trompeurs. En outre, les comptes nouvellement créés sont soumis à un examen pendant un certain temps avant de pouvoir commencer à faire de la publicité avec les solutions Twitter.

Lorsque les annonceurs sur Twitter choisissent de promouvoir leur contenu avec les solutions Twitter, leur contenu peut être soumis à un processus d'examen. Cette procédure vise à garantir la qualité et la sécurité de la plateforme publicitaire Twitter. Les publicités Twitter peuvent être examinées avant d'être utilisées dans les campagnes. Elles sont soumises à approbation de manière automatique, sur la base du statut publicitaire d'un compte, de son utilisation historique de Twitter et d'autres facteurs évolutifs. L'examen prend généralement en considération la façon dont un compte utilise Twitter, son profil, son contenu et le ciblage inclus dans toute campagne publicitaire active ou en projet. Ce processus permet à Twitter de vérifier que les annonceurs se conforment à nos politiques publicitaires. Toutes les politiques publicitaires font l'objet de révisions et de mises à jour régulières afin d'être conformes à la législation en vigueur et aux évolutions des produits. Nous investissons massivement dans ces domaines afin de maintenir notre produit et nos processus à jour.

#### **5.6 L'opérateur fournira les chiffres suivants :**

- **le volume d'annonces publicitaires et contenus sponsorisés porteurs de fausses informations ainsi que leur nombre d'impressions ;**
- **les revenus générés par les publicités porteuses de fausses informations avant leur retrait. Il décrira la méthode et les indicateurs employés pour obtenir le montant de ces revenus. Le cas échéant, il distinguera la part de ces revenus alloués à l'opérateur et aux utilisateurs ayant posté les contenus auxquels ont été accolées ces publicités. Il indiquera le nombre d'impressions de ces publicités ;**
- **le nombre d'utilisateurs ayant été exposés aux publicités porteuses de fausses informations avant leur détection et leur retrait ;**
- **le nombre d'entre eux ayant cliqué sur ces publicités.**

Calculer le volume de contenu sponsorisé contenant des fausses informations est un processus complexe. En effet, bien que nous disposions du nombre de Tweets sponsorisés qui ont violé des politiques spécifiques telles que notre politique de publicité lié au COVID-19, une violation de cette politique n'implique pas nécessairement de la désinformation (par exemple, la promotion de certains produits liés à la COVID-19 tels que les masques chirurgicaux est interdite sur Twitter). Cela s'explique





par le fait que nous fixons des normes très élevées en matière de publicité et que nous appliquons des politiques très strictes.

Ainsi, à l'échelle mondiale, du 1er mars 2020 au 31 décembre 2020, 18 287 Tweets sponsorisés ont enfreint notre politique de publicité pour COVID-19. Mais cela ne signifie pas pour autant que 18 287 Tweets sponsorisés ont véhiculé des fausses informations sur Twitter.

Ainsi, nous avons également examiné le nombre total de Tweets sponsorisés en 2020 qui ont été libellés comme comportant de la désinformation par nos modérateurs ou nos systèmes. Pour 2020, dans le monde, nous n'avons trouvé que 9 Tweets sponsorisés uniques comportant un libellé de désinformation et aucun d'entre eux ne ciblait la France. En outre, notre dernière série de tests s'est déroulée du 1er mars 2020 au 12 août 2020. Nous avons examiné un échantillon aléatoire de 153 publicités et n'en avons trouvé aucune qui contenait des fausses informations. Nous estimons qu'environ 95 % du contenu contrevenant durant cette période a été détecté par nos systèmes automatisés. Ce résultat est important car il montre que même si des publicités auraient pu contenir des fausses informations, elles auraient très probablement été supprimées automatiquement par nos systèmes avant même de générer des impressions sur la plateforme.

### **5.7 L'opérateur détaillera les mesures de sécurité des marques mises en place et expliquera comment elles ont été conçues.**

**Il détaillera notamment les éléments suivants :**

- **liste des mesures;**
- **moyens humains et automatisés dédiés à l'exécution de ces mesures ;**
- **appréciation de l'impact de ces mesures ;**
- **existence d'une information des annonceurs lorsque leurs contenus se sont trouvés accolés à des fausses informations et, le cas échéant, de mécanismes de compensation ou de recours ;**
- **modalités de participation de l'opérateur dans tout mécanisme de coopération avec les marques et l'interprofession publicitaire.**

**L'opérateur pourra donner des exemples afin d'illustrer les mesures mises en œuvre.**

L'objectif de Twitter est de servir la conversation publique et nous nous engageons à créer un environnement serein où chacun, y compris les annonceurs, peut participer et s'engager en toute sécurité. C'est pourquoi nous avons mis en place nos politiques de sécurité des marques ("Brand Safety"), qui reposent sur les principes fondamentaux des Règles de Twitter, afin de promouvoir une expérience publicitaire sûre pour toutes les personnes qui utilisent notre service et pour les marques. Nos politiques de sécurité des marques informent le contexte dans lequel nous diffusons les publicités et incluent, sans s'y limiter, les produits et services restreints et illégaux. Nous savons qu'il y a beaucoup de travail à faire pour y parvenir, et nous investissons activement dans ce domaine et travaillons avec des partenaires clés du secteur pour faire progresser la sécurité des marques en tant que composante centrale des solutions de publicité et de mesure sur Twitter.

En décembre 2020, nous avons annoncé nos principaux axes de travail pour les mois à venir en termes de sécurité des marques. Premièrement, nous doublons les solutions de mesure de la sécurité des marques via des tiers. Nous avons sélectionné DoubleVerify (DV) et Integral Ad Science (IAS) comme partenaires privilégiés de Twitter pour fournir des rapports indépendants sur le contexte dans lequel les publicités apparaissent sur Twitter. Nous considérons qu'il s'agit là d'une opportunité de créer des solutions qui permettront aux annonceurs de mieux comprendre les types de contenus qui apparaissent à côté de leurs publicités, les aidant ainsi à prendre des décisions éclairées pour atteindre leurs objectifs marketing. Ces solutions viendront compléter nos solutions existantes de mesure de la visibilité par des tiers avec DV et IAS.



Deuxièmement, nous avons partagé notre engagement à suivre le processus d'accréditation pour les quatre services d'accréditation proposés par le Media Ratings Council (MRC) : mesure de la visibilité, filtrage sophistiqué du trafic non valide, mesure des audiences, et sécurité des marques. Nous donnerons la priorité à l'audit de sécurité des marques et avons déjà commencé le processus de proposition pour cette accréditation. Nous pensons que ces quatre accréditations sont essentielles pour démontrer notre engagement durable envers la transparence et nous sommes heureux d'être en mesure d'étendre et de soutenir ce processus de collaboration. Nous sommes impatients de travailler avec le MRC et de nous aligner davantage sur la portée, la séquence et le calendrier de cette collaboration.

Un troisième domaine de travail est le partenariat continu que nous établissons avec les leaders du secteur. Twitter est un membre actif et fondateur de la Global Alliance for Responsible Media (GARM). Parallèlement à nos efforts continus pour mieux aligner nos solutions de sécurité des marques sur les principes de la GARM, nous soutenons la DV et l'IAS dans leurs efforts pour adapter leurs cadres de catégorisation actuels afin de mieux correspondre aux catégories de contenu monétisé nuisible identifiées par la GARM. Nous continuerons à nous associer aux leaders de l'industrie pour définir et affiner le travail nécessaire pour que les plateformes restent des lieux sûrs pour les annonceurs.

Enfin, nous cherchons à utiliser le pouvoir de la recherche pour en savoir plus sur la sécurité des marques. L'environnement personnalisé de Twitter, basé sur les flux, rend la contiguïté des contenus intrinsèquement unique lorsqu'il s'agit d'examiner les questions de sécurité des marques. Il est essentiel de comprendre où les associations se produisent, où elles ne se produisent pas, et comment ces associations affectent les marques alors que nous continuons à investir dans ce domaine. Les premières recherches effectuées par les équipes de science des données de Twitter indiquent que sur notre plateforme, il n'a pas encore été constaté que la contiguïté entre les publicités et le contenu controversé affecte la favorabilité des marques. Nous travaillons actuellement sur des recherches supplémentaires afin d'acquérir davantage de connaissances et d'informations sur l'impact de la contiguïté sur les publicités Twitter. Nous sommes impatients d'intégrer ces résultats dans notre travail en cours avec DV et l'IAS et de les utiliser pour déterminer comment nous construisons et perfectionnons nos produits publicitaires.

#### **5.8 L'opérateur fournira les chiffres suivants :**

- **le volume d'annonces publicitaires et contenus sponsorisés accolés à des contenus porteurs de fausses informations ainsi que leur nombre d'impression ;**
- **les revenus générés par les annonces publicitaires et contenus sponsorisés accolés à des contenus porteurs de fausses informations avant l'éventuel retrait de ces dernières. Il décrira la méthode et les indicateurs employés pour obtenir le montant de ces revenus. Le cas échéant, il distinguera la part de ces revenus alloués à l'opérateur et aux utilisateurs ayant posté les contenus porteurs de fausses informations.**

Pour un certain nombre de raisons, notamment d'ordre technique, il est actuellement impossible pour Twitter d'assurer le suivi du volume, du nombre d'impressions et du montant des revenus générés par les publicités liées à des contenus identifiés comme des fausses informations. Cependant, comme mentionné au point 5.7, l'objectif de Twitter est de servir la conversation publique et nous nous engageons à créer un environnement sain où chacun, y compris les annonceurs, peut participer et s'engager en toute sécurité. Par conséquent, nous nous engageons à intensifier nos efforts en matière de sécurité des marques afin de mieux appréhender le problème et de promouvoir une expérience publicitaire sûre pour toutes les personnes qui utilisent notre service et pour les marques.

#### **5.9 L'opérateur indiquera au Conseil s'il dispose d'une base de données relative aux communications commerciales diffusées sur le service ; il détaillera :**

- **si cette base est publique,**



- les types de données qu'elle contient,
- si celles-ci peuvent être filtrées spécifiquement pour la France,
- si l'utilisateur peut les consulter et les télécharger,
- si cette base a évolué depuis 2019.

Chez Twitter, nous avons lancé le [Centre de transparence des publicités](#) en juin 2018 en tant que base de données offrant une plus grande transparence et un meilleur aperçu des publicités politiques et d'intérêt général sur la plateforme. Avec l'interdiction officielle des publicités politiques sur Twitter en novembre 2019 et la mise en place d'un processus obligatoire de certification et d'examen des publicités d'intérêt général, le Centre de transparence a perdu sa raison d'être initiale.

Pour remplacer le Centre et maintenir les mêmes niveaux de transparence, Twitter fournit désormais une version archivée des données de toutes les publicités politiques diffusées entre le 24 mai 2018 et le 22 novembre 2019 et des publicités d'intérêt général diffusées entre le 8 août 2018 et le 22 novembre 2019. Si les gens souhaitent en savoir plus sur toutes les publicités qui leur sont diffusées, ils peuvent télécharger ces informations en cliquant sur "Votre compte" dans l'onglet "Paramètres et confidentialité".

### ***Contenus des utilisateurs réalisés en partenariat avec des tiers***

#### **5.10 L'opérateur a-t-il décelé des contenus porteurs de fausses informations ayant bénéficié d'un partenariat rémunéré entre l'utilisateur ayant posté le contenu et un tiers (placement de produit, contenu de marque, etc.) ? Le cas échéant, a-t-il pris des mesures spécifiques et si oui, lesquelles ?**

Les éditeurs qui partagent du contenu vidéo sur Twitter sont un élément essentiel de la conversation qui se déroule chaque jour sur notre plateforme. Le programme d'édition Amplify a été créé pour permettre aux éditeurs de monétiser facilement du contenu vidéo sur Twitter, et aux annonceurs de s'aligner sur du contenu sûr pour les marques. Tous les contenus des éditeurs ne sont pas adaptés à la monétisation par le biais du programme d'édition Amplify ; les directives décrites ci-dessous permettent d'aider les éditeurs à comprendre quels contenus ne sont pas éligibles à la monétisation.

Toutes les vidéos et tous les Tweets associés monétisés par le biais du programme d'éditeurs Amplify doivent respecter un ensemble de directives. Nous nous réservons le droit, à notre discrétion, de ne pas afficher de publicités, et donc de supprimer la monétisation, sur un contenu que nous estimons ne pas convenir aux annonceurs.

Lorsque nous examinons le contenu monétisé par les éditeurs, nous prenons en compte tous les éléments d'un Tweet monétisé (contenu vidéo, texte du Tweet, audio, images, texte superposé, contexte). Les règles de Twitter - y compris nos politiques en matière d'intégrité civique et de fausses informations liées à la COVID-19 - s'appliquent pleinement au programme des éditeurs Amplify et les catégories suivantes ne sont pas éligibles à la monétisation :

- Contenu sexuel pour adultes
- Produits et services restreints et illégaux
- Obscénité et langage offensant
- Spam, contenu trompeur et malveillant.

Si une vidéo monétisée contient du contenu qui enfreint de manière flagrante les règles de Twitter ou l'une des directives décrites ci-dessus, le compte de l'éditeur peut être soumis à la politique dite de "strike" de Twitter. Chaque "strike" sera traitée comme suit :

- Strike 1 : la monétisation est supprimée de la vidéo et l'éditeur reçoit un avertissement dans le Media Studio.



- Strike 2 : la monétisation est supprimée de la vidéo et l'éditeur ne pourra pas monétiser de vidéos pendant une semaine.
- Strike 3 : l'éditeur est retiré du programme d'éditeurs Amplify et ne pourra pas monétiser de vidéos pendant six mois.

### **5.11 Existe-t-il des moyens permettant de lutter contre ce phénomène ?**

Nous reconnaissons que la monétisation de la désinformation est un problème complexe. C'est pourquoi nous avons pris des mesures proactives et apporté un certain nombre de changements, notamment au niveau de nos politiques de sécurité renforcées, de meilleurs outils et ressources pour déceler et stopper les activités malveillantes, des normes publicitaires et un processus d'approbation plus stricts, ainsi qu'une transparence accrue pour favoriser la compréhension du public dans tous ces domaines.

### **5.b. Information des utilisateurs sur la nature, l'origine et les modalités de diffusion des contenus, et l'identité des personnes versant des rémunérations en contrepartie de la promotion des contenus d'information se rattachant à un débat d'intérêt général**

#### *Informations apportées aux utilisateurs sur les annonces publicitaires et contenus sponsorisés en général*

**5.12 L'opérateur précisera les informations fournies à l'utilisateur sur les annonces publicitaires et contenus sponsorisés présents sur son service, en décrivant les spécificités des deux catégories le cas échéant :**

- sont-ils clairement distingués des autres contenus et, si oui, comment ?
- la plateforme permet-elle à l'utilisateur d'identifier les critères pris en compte dans l'affichage des annonces publicitaires et contenus sponsorisés, notamment, si celles-ci sont personnalisées en fonction de son profil ?
- l'origine de l'annonce publicitaire ou du contenu sponsorisé est-elle clairement identifiée et affichée (ainsi que celle de l'annonceur, si elle diffère) ?
- le service fournit-il des indications sur les modalités de diffusion des annonces publicitaires et contenus sponsorisés (existence de contreparties financières, nombre de vues, type de population ciblée, manière dont le contenu a été généré (automatisée ou non), autres) ?
- l'opérateur décrit-il ses politiques publicitaires dans des documents publics disponibles en français, notamment en matière de lutte contre la désinformation sur son service ? Comment l'utilisateur peut-il accéder à ces descriptions ?
- où et comment la plateforme appelle-t-elle les utilisateurs à la vigilance sur les annonces publicitaires et contenus sponsorisés ayant fait l'objet de signalements mais qui n'auraient pas été retirés ?
- l'utilisateur a-t-il la possibilité de masquer des annonces publicitaires et contenus sponsorisés ? Peut-il paramétrer ses préférences publicitaires ? Si oui, comment est-il informé de cette possibilité, comment a-t-il accès à ces paramètres, que recouvre ce paramétrage en pratique, quelles informations lui sont fournies ?

Les contenus sponsorisés par les services publicitaires de Twitter se distinguent clairement des contenus organiques et sont marqués d'une icône "Sponsorisé". Les utilisateurs peuvent interagir avec la plupart des contenus promus de la même manière que les contenus organiques.

Ils peuvent voir les raisons pour lesquelles une publicité spécifique leur a été proposée en cliquant sur celle-ci. Les informations partagées renvoient également l'utilisateur à ses propres paramètres afin de contrôler s'il reçoit ou non des publicités basées sur ses intérêts déduits par Twitter en utilisant le paramètre "Personnaliser les publicités".



Les [politiques de Twitter en matière de publicité](#) sont accessibles au public et sont adaptées à la fois à la langue et aux restrictions juridiques locales.

Twitter autorise les signalements de publicités directement au niveau de la publicité elle-même et prend des mesures si la publicité est en violation des politiques publicitaires de Twitter.

### ***Contenus promus d'information se rattachant à un débat d'intérêt général***

#### **5.13 L'opérateur indiquera au Conseil :**

- **la définition des contenus promus d'information se rattachant à un débat d'intérêt général adoptée par l'opérateur ;**
- **la manière dont il identifie les contenus sponsorisés relevant de cette catégorie**
- **leur nombre sur la plateforme en 2020 ;**
- **l'existence de tout système de vérification particulier pour les annonceurs faisant la promotion de contenus d'information se rattachant à un débat d'intérêt général et, le cas échéant, le nombre d'annonceurs ayant fait l'objet de cette procédure spécifique ;**
- **si ces contenus font l'objet de restrictions ou règles particulières (ex : restriction de ciblage) ;**
- **si la procédure de signalement des contenus d'informations se rattachant à un débat d'intérêt général diffère de celle des autres communications commerciales ;**
- **les différences de traitement de ce type de communications commerciales avec le reste des communications commerciales.**

**Il peut préciser s'il s'est rapproché d'autres acteurs (opérateurs, société civile, ONG, *think tank*, institutions...) pour œuvrer à une plus grande harmonisation des approches retenues en la matière.**

Chez Twitter, nous pensons que la portée des messages politiques se doit d'être gagnée, et non achetée. C'est pourquoi, en novembre 2019, nous avons décidé d'interdire au niveau mondial la promotion de contenus politiques. Nous définissons le contenu politique comme un contenu qui fait référence à un candidat, un parti politique, un fonctionnaire élu ou nommé, une élection, un référendum, une mesure de vote, une législation, une réglementation, une directive ou un résultat judiciaire. Les publicités qui contiennent des références à un contenu politique, y compris les appels aux votes, les sollicitations de soutien financier et les plaidoyers pour ou contre l'un des types de contenu politique énumérés ci-dessus, sont interdites en vertu de cette politique. Nous n'autorisons pas non plus les publicités de tout type émanant de candidats, de partis politiques ou de fonctionnaires élus ou nommés.

En outre, Twitter restreint la promotion - et exige la certification des annonceurs - pour les publicités qui éduquent, sensibilisent et/ou appellent les gens à agir dans le cadre de l'engagement civique, de la croissance économique, du changement climatique ou des causes d'équité sociale. Nous avons pris cette décision sur la base des deux convictions suivantes. Premièrement, la publicité ne doit pas être utilisée pour influencer les résultats politiques, judiciaires, législatifs ou réglementaires. Cependant, la publicité basée sur une cause peut faciliter la conversation publique autour de sujets importants. Deuxièmement, la publicité qui utilise le micro-ciblage présente des défis entièrement nouveaux pour le discours civique qui ne sont pas encore entièrement compris.

Par conséquent, en plus de se conformer à toutes les autres politiques relatives aux publicités Twitter, y compris notre politique en matière de contenu politique, tous les annonceurs souhaitant diffuser des publicités d'intérêt général doivent suivre notre processus de certification des annonceurs et respecter les restrictions suivantes. Premièrement, le ciblage est restreint et limité au ciblage géographique, par mot-clé et par centre d'intérêt. Aucun autre type de ciblage n'est autorisé, y compris les audiences



personnalisées. Le ciblage géographique ne peut se faire au maximum qu'au niveau d'une province ou d'une région. Le ciblage au niveau du code postal n'est pas autorisé. Le ciblage par mots-clés et par centres d'intérêt ne peut pas inclure de termes associés à des contenus politiques, à des annonceurs interdits ou à des tendances ou affiliations politiques (par exemple, "conservateur", "libéral", "élections politiques", etc.) Deuxièmement, les annonces ne peuvent pas être diffusées au nom de personnes ou d'entités auxquelles il est interdit de faire de la publicité en vertu de notre politique en matière de contenu politique, ni faire spécifiquement référence à ces personnes ou entités. Troisièmement, la page de destination d'une annonce ne doit pas être directement liée à des personnes ou des entités dont la publicité est interdite en vertu de notre politique en matière de contenu politique. Enfin, les organisations à but lucratif qui cherchent à diffuser des publicités qui éduquent, sensibilisent et/ou appellent les gens à agir en rapport avec ces causes doivent également se conformer aux restrictions supplémentaires suivantes : les publicités ne doivent pas avoir pour objectif principal d'obtenir des résultats politiques, judiciaires, législatifs ou réglementaires ; et les publicités doivent être liées aux valeurs, principes et/ou convictions publiquement déclarés de l'organisation.

**5.14 L'opérateur indiquera au Conseil s'il dispose d'une base de données relative aux contenus promus d'information se rattachant à un débat d'intérêt général diffusés sur le service.**

**Il décrira cette base en fournissant les informations telles que listées à la question 5.9.**

Nous avons lancé notre [Centre de transparence des publicités](#) en juin 2018 pour fournir un meilleur aperçu des publicités politiques et d'intérêt général sur la plateforme. Avec l'interdiction officielle des publicités politiques sur Twitter en novembre 2019 et la mise en place d'un processus obligatoire de certification et d'examen des publicités d'intérêt général, le Centre de transparence n'offre plus la valeur initialement prévue. Pour remplacer le Centre et maintenir les mêmes niveaux de transparence, Twitter fournit désormais une version archivée des données de toutes les publicités politiques diffusées entre le 24 mai 2018 et le 22 novembre 2019 et des publicités d'intérêt général diffusées entre le 8 août 2018 et le 22 novembre 2019. Si les gens souhaitent en savoir plus sur toutes les publicités qui leur sont diffusées, ils peuvent télécharger ces informations en cliquant sur "Votre compte" dans l'onglet "Paramètres et confidentialité".

### **Autres questions**

**5.15 L'opérateur précisera si des changements ou de nouvelles mesures ont été mis en place en 2020 et s'il a adopté des initiatives particulières relatives aux contenus sponsorisés dans le cadre de la crise sanitaire de la COVID-19.**

En plus de nos politiques en matière de publicité politique et d'intérêt général, nous avons formulé des politiques de publicité spécifiques à la COVID-19, conformément à notre engagement d'assurer la sécurité des personnes, à promouvoir un contenu honnête et à interdire la distribution de spams, de contenus nuisibles ou d'autres contenus perturbateurs. Par conséquent, nous avons décidé de restreindre les publicités contenant une référence implicite ou explicite à la COVID-19. Plus précisément, la publicité contenant une référence implicite ou explicite à la COVID-19 est autorisée lorsqu'elle fait référence à des ajustements en termes de pratiques et/ou de modèles commerciaux en réponse à la COVID-19 et au soutien aux clients et aux employés lié à la COVID-19, avec les restrictions suivantes :

- les références de mauvais goût à COVID-19 sont interdites
- le contenu ne doit pas être sensationnaliste ou susceptible d'inciter à la panique
- les prix des produits liés à COVID-19 ne doivent pas être gonflés
- la promotion de certains produits liés à COVID-19 peut être interdite



- Nous interdisons actuellement la publicité pour les masques médicaux et les désinfectants alcoolisés pour les mains (avec ou sans référence à COVID). D'autres produits peuvent être ajoutés à cette liste et l'application peut être rétroactive.
- la mention de vaccins, de traitements et de kits de test est autorisée, uniquement sous forme d'information, par les éditeurs de presse qui ont été exemptés en vertu de notre politique relative aux publicités politiques.

**5.16 L'opérateur indiquera dans quelle mesure il a tenu compte des recommandations du Conseil énoncées dans son bilan relatif à l'exercice 2019 à compter de sa publication en juillet 2020.**

Pour cette section du questionnaire, le Conseil a formulé plusieurs recommandations à l'intention des plateformes, notamment : la labellisation claire des contenus sponsorisés (par rapport aux contenus organiques) ; la mise à disposition de politiques publicitaires claires en français (notamment en ce qui concerne les publicités trompeuses ou inappropriées) ; l'inclusion d'une section sur la manipulation de l'information ; le renforcement des mesures et des efforts en matière de sécurité des marques ; et la mise à disposition d'un accès facile à une base de données de publicités pour les utilisateurs. Nous avons examiné attentivement ces recommandations et avons constaté que Twitter était en conformité avec la plupart d'entre elles. Toutefois, en ce qui concerne le dernier point relatif à l'accès des utilisateurs à une base de données de publicités, nous tenons à souligner que, bien que nous ne puissions pas fournir une base de données de toutes les publicités passées et présentes qui ont été ou sont actuellement diffusées sur Twitter, nous fournissons une grande transparence via nos archives concernant les publicités politiques et d'intérêt général sur la plateforme par le biais de notre [Centre de transparence des publicités](#). Avec la fin de ce Centre de transparence liée à notre interdiction des publicités politiques et aux restrictions des publicités d'intérêt général sur Twitter en novembre 2019, nous fournissons désormais une version archivée des données de toutes les publicités politiques diffusées entre le 24 mai 2018 et le 22 novembre 2019 et des publicités d'intérêt général diffusées entre le 8 août 2018 et le 22 novembre 2019.



## **6. Favoriser l'éducation aux médias et à l'information**

### ***Actions d'éducation aux médias et à l'information***

#### **6.1 L'opérateur est invité à déclarer au Conseil s'il a développé ou mis en place :**

- **des outils permettant aux utilisateurs d'identifier la fiabilité des informations et sources d'informations qui sont proposées sur le service ;**
- **des actions ciblées sur le service afin de sensibiliser les utilisateurs à une utilisation maîtrisée de celui-ci (ex : encart d'information au moment de la publication d'un message, rappel des règles de savoir-vivre en ligne, etc.), qu'elles soient menées par l'opérateur seul ou en partenariat ;**
- **plus généralement, des actions qu'il aurait menées, accompagnées ou promues en matière d'éducation aux médias et à l'information dans le domaine de la lutte contre la désinformation en ligne, et leur nombre (en distinguant en fonction de la nature de son implication) ;**
- **des soutiens à des initiatives indépendantes émanant de journalistes et de chercheurs destinées à mieux comprendre et mesurer le phénomène de la désinformation.**

**L'opérateur décrira chaque outil, action ou soutien en renseignant le tableau ci-après. Il s'attachera notamment à indiquer : sa nature, son objectif, en quoi il/elle consiste, sa date de lancement, sa mise en valeur auprès du public et des utilisateurs, le ou les public(s) visé(s) et l'impact de son action sur ces derniers (en termes quantitatifs – le nombre de personnes effectivement touchées – et qualitatifs – l'impact de son action sur le comportement du public visé), les partenaires éventuels, la nature de sa propre implication et le montant des dépenses qu'il a engagées et/ou des recettes qu'il a tirées pour chacune de ces opérations.**

**Il pourra fournir toute présentation illustrée de ces différentes mesures.**

Chez Twitter, nous sommes conscients de l'importance de l'éducation aux médias et à l'information ainsi que des compétences numériques, en particulier auprès du jeune public. Notre mission est de donner aux gens les moyens de développer les compétences qui les aideront à analyser de manière critique les contenus pour en vérifier la crédibilité et à se poser les bonnes questions sur les informations qu'ils voient, utilisent et partagent en ligne.

En cette période de pandémie mondiale, le besoin en termes de compétences en matière d'éducation aux médias et à l'information n'a jamais été aussi aigu. C'est pourquoi nous avons adapté certaines caractéristiques de notre produit afin de créer certaines incitations vertueuses. Ainsi, lorsqu'une personne retweete un article qu'elle n'a pas ouvert sur Twitter, il lui est demandé si elle souhaite l'ouvrir avant de le partager. Cette expérience s'est avérée particulièrement efficace pour inciter les gens à lire davantage et à envoyer des messages sur Twitter de manière plus informée. Nous avons, en effet, observé que :

- les gens ouvrent les articles 40 % plus souvent après avoir vu cette incitation
- le nombre de personnes ouvrant les articles avant de les retweeter a augmenté de 33 %

En parallèle, nous nous sommes associés à des organisations internationales de premier plan pour lancer des campagnes de sensibilisation dans le domaine de l'éducation aux médias et à l'information. Par exemple, depuis le début de la crise de la COVID-19, nous avons mené deux campagnes d'éducation aux médias en partenariat avec l'UNESCO et la Commission européenne en utilisant le hashtag #ThinkBeforeSharing.







This week we celebrate our partnership with @UNESCO on media and information literacy.

In these turbulent times, with more content than ever at our fingertips, it's increasingly important to #ThinkBeforeSharing.

Check out @MILClicks for more.



Nous avons travaillé avec l'UNESCO pour célébrer la semaine mondiale de l'éducation aux médias et à l'information (Global Media and Information Literacy week) du 24 au 31 octobre 2020, dont le thème était *Resisting Disinfodemic: Media & Information Literacy for everyone & by everyone*. Nous avons notamment pris la parole lors de sessions de haut niveau, rejoint des panels d'intervenants et de juges et partagé notre émoji #ThinkBeforeSharing.

Cependant, le travail de Twitter sur l'éducation aux médias a commencé bien avant ce que l'on a appelé "l'infodémie de COVID-19". Twitter et l'UNESCO travaillent ensemble depuis 2018 pour faire avancer la conversation autour de l'éducation aux médias et à l'information. À l'occasion de la semaine mondiale de l'éducation aux médias et à l'information, Twitter, en collaboration avec l'UNESCO, a produit la ressource "Enseigner et apprendre avec Twitter". Il s'agit d'un manuel facile à lire pour les éducateurs de tous bords qui souhaitent tirer parti des avantages de Twitter en tant qu'outil d'apprentissage en classe et à la maison, tout en recevant des conseils directement de l'UNESCO sur les meilleures pratiques en matière d'éducation aux médias et à l'information. Il illustre la manière dont Twitter peut être utilisé pour l'apprentissage créatif autour de la sensibilisation aux médias et à l'information, ainsi que sur la manière de gérer la cyberintimidation, de contrôler son empreinte numérique et de se protéger en ligne. Cela est également lié aux préoccupations concernant la désinformation et à nos efforts pour mettre en avant les informations de qualité sur la plateforme. Dans un contexte comme celui que nous connaissons actuellement, où les enfants et les jeunes doivent parfois suivre leur cours à distance, cette ressource est extrêmement précieuse.

### **Relations avec le monde de la recherche**



Twitter, Inc. | [about.twitter.com](https://about.twitter.com)

**6.2 Il renseignera le Conseil sur les relations qu'il entretient avec le monde de la recherche et les initiatives qu'il a conduites ou soutenues dans ce domaine en France, ou impliquant la recherche française, ainsi que la publicité qu'il aurait faite de ces initiatives.**

**En particulier, il indiquera et décrira les partenariats qu'il aurait conclus dans ce cadre :**

- **visant à mieux cerner certains comportements et phénomènes liés à la diffusion de fausses informations en ligne et précisera les moyens qu'il met à leur disposition (ex. : concernant la propagation de fausses informations, l'opérateur donne-t-il accès à des données, notamment via une API ; si oui, sous quelles conditions cet accès est-il possible ?) ;**
- **visant à vérifier la qualité des systèmes de recommandation et de modération du service qu'il opère en termes d'explicabilité, de loyauté et d'équité et, le cas échéant, proposer des pistes d'amélioration.**

**L'opérateur indiquera et décrira au Conseil les protocoles existants ou en construction avec le monde de la recherche portant sur le partage et l'exploitation des données.**

**Enfin, s'il n'a conclu aucun partenariat avec le monde de la recherche, il indiquera au Conseil les raisons et/ou les difficultés rencontrées.**

Depuis 2006, l'API de Twitter donne aux chercheurs et aux développeurs la possibilité d'exploiter ce qui se passe dans le monde. Voici quelques exemples récents :

- [changement climatique](#)
- [lutte contre les discours de haine](#)
- plusieurs études sur la COVID-19

Twitter croit fermement à l'accès ouvert aux données pour étudier, analyser et contribuer à la conversation publique ; c'est pourquoi nous continuons à maintenir une large Application Programming Interface (API) publique. Les chercheurs utilisent les données de Twitter pour fournir des informations précieuses sur l'évolution des conversations et des interactions en ligne sur et en dehors de Twitter. Nous continuons à fournir des moyens plus accessibles pour mettre les données et les informations à la disposition des chercheurs.

L'API de Twitter constitue une source de données unique pour les universitaires, utilisée chaque jour dans le monde entier dans un large éventail de domaines, de la gestion des catastrophes naturelles aux sciences politiques. Chaque grande conférence sur les sciences sociales est susceptible de présenter plusieurs articles basés entièrement ou en grande partie sur des données Twitter. Notre service est la plus grande source de données de réseaux sociaux en temps réel, et nous mettons ces données à la disposition du public gratuitement par le biais de notre API publique. Aucun autre grand service ne fait cela. Nous mettons à disposition tout le contenu public de Twitter via notre API. Notre [page dédiée](#) offre plus d'informations à ce sujet ici et [tout le monde peut découvrir des usages innovants](#) de notre API par des chercheurs du monde entier qui utilisent les données Twitter. Par exemple, c'est grâce à l'API ouverte de Twitter que le CSA a pu collecter de manière transparente toutes les données nécessaires à son étude de novembre 2020 sur la propagation des fausses informations sur Twitter.

Afin de soutenir nos efforts continus pour protéger la conversation publique et aider les gens à trouver des informations sanitaires faisant autorité autour de la COVID-19, nous avons publié un point d'accès API spécifique à la COVID-19 dans Twitter Developer Labs pour permettre aux développeurs et chercheurs agréés d'étudier la conversation publique sur la COVID-19 en temps réel. Il s'agit d'un ensemble de données unique qui couvre plusieurs dizaines de millions de Tweets par jour et offre un aperçu de l'évolution de la conversation publique mondiale autour d'une crise sans précédent. Rendre cet accès gratuit est l'une des choses les plus uniques et les plus précieuses que Twitter puisse faire alors



que le monde se rassemble pour protéger nos communautés et chercher des réponses aux défis urgents. Les questions sur lesquelles les chercheurs se penchent le plus souvent sont les suivantes : comment se propagent les fausses informations, la façon dont le public perçoit les politiques ou les déclarations officielles, les effets à long terme du virus sur la santé mentale, les communautés vulnérables et le comportement social autour de la COVID-19. Par exemple, le point d'accès COVID-19 a servi de base à une étude précieuse sur les [alertes précoces de l'épidémie de COVID-19 à travers l'Europe](#) à partir des réseaux sociaux et ce, avant même les premières annonces publiques de la part des autorités de santé.

Nous avons récemment lancé une mise à jour conçue pour répondre aux besoins de la communauté des chercheurs universitaires qui effectuent des recherches à l'aide de données Twitter, appelée [Academic Research Product Track](#). Cette mise à jour permettra aux chercheurs d'obtenir un accès gratuit à l'historique complet des conversations publiques, des niveaux d'accès plus élevés et gratuits à la plateforme de développement Twitter, des capacités de filtrage plus précises, de [nouveaux guides techniques et méthodologiques](#), un accès à tous les [points d'accès v2 publiés à ce jour](#).

Enfin, Twitter divulgue au public des ensembles de données sur les [opérations de désinformation](#) afin de faciliter la recherche et de fournir un degré élevé de transparence au public et à la communauté des chercheurs. Ce niveau de fourniture de données reste inégalé dans le secteur. Nous sommes fiers du fait que des milliers de chercheurs ont utilisé ces ensembles de données qui contiennent plus de 30 millions de Tweets individuels et plus d'un téraoctet de médias pour mener leurs propres enquêtes indépendantes. Il s'agit de la plus grande base de données de ce type dans le secteur et nous continuons à fournir des données aux chercheurs du monde entier par le biais de notre API ouverte et de notre point d'accès COVID-19 dédié.

## **Autres questions**

### **6.3 L'opérateur indiquera dans quelle mesure il a tenu compte des recommandations du Conseil énoncées dans son bilan relatif à l'exercice 2019 à compter de sa publication en juillet 2020.**

L'une des recommandations du Conseil consistait à demander aux plateformes de favoriser des partenariats de long-terme avec la communauté académique par le biais du partage et de l'exploitation transparente des données. Twitter s'est toujours engagé à faire progresser la compréhension globale des sujets sociétaux et scientifiques importants, notamment autour de la COVID-19. Par conséquent, au cœur de cette pandémie mondiale, nous avons publié notre point d'accès à l'API spécifiquement dédié à la COVID-19 dans le Twitter Developer Labs afin de permettre aux développeurs et aux chercheurs agréés d'étudier la conversation publique sur la COVID-19 en temps réel. Enfin, comme mentionné dans la réponse 6.2, nous avons travaillé sur plusieurs mises à jour et améliorations importantes de notre accès API ouvert afin de mieux répondre aux besoins de la communauté académique qui effectue des recherches avec des données Twitter.

### **6.4 Il indiquera l'impact de la crise sanitaire sur ses actions en faveur de l'éducation aux médias et à l'information et notamment, les adaptations qu'il a pu être amené à lui apporter en conséquence.**

La pandémie de COVID-19 et les restrictions sanitaires, notamment le télétravail obligatoire, ont eu un impact considérable sur l'ensemble des activités de Twitter dans le monde. En ce qui concerne nos actions en faveur de l'éducation aux médias et à l'information, la pandémie nous a obligés à adapter et à réinventer nos efforts et nos initiatives autour de ce domaine d'activité clé, tant sur le contenu que le format de nos actions. Cependant, nous avons réussi à maintenir des relations étroites avec nos partenaires existants et nos principaux publics cibles par le biais de formations en ligne régulières, de tables rondes et des partenariats clés décrits dans cette section du questionnaire.



## Annexe

Nom de la campagne, de l'outil ou de l'action de soutien (indiquer le lien URL)	Description et objectif (s)	Partenariat (NB : préciser si c'est un partenariat initié par l'opérateur ou une réponse à une sollicitation)	Nature de l'implication de l'opérateur	Public(s) visé(s)	Pays / région (s) visé(s)	Dates de début et de fin	Budg et alloué	Mise en valeur auprès du public	Données sur les audiences touchées (si aucune, préciser les difficultés rencontrées)	Données sur l'impact de la campagne sur le comportement des publics visés et méthodes pour les mesurer (si aucune, préciser les difficultés rencontrées)
Semaine mondiale de l'éducation aux médias et à l'information	Création d'un panel d'intervenants	UNESCO	Partenaire clé	Jeunes	Monde	24 au 31 octobre 2020	N/A	<a href="#">Tweet</a>	N/A	N/A
Campagne d'éducation aux médias	Création de ressources accessibles en ligne	UNESCO	Partenaire clé	Jeunes	Monde	14 août 2020	N/A	<a href="#">Tweet</a>	N/A	N/A
Campagne #ThinkBeforeSharing	Soutien à la communication	Commission européenne	Partenaire clé	Jeunes	UE	21 avril-5 mai	N/A	Promotion de la <a href="#">campagne</a> via @Policy et @Nonprofits	N/A	N/A

