



Questionnaire 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 référent légal de l'opérateur de la plateforme en ligne en France** : Audrey Herblin-Stoop, Directrice des Affaires Publiques
- **Nombre de visiteurs uniques par mois de la plateforme en ligne, calculé sur la base de 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** :
 - o **Disponibilité sur internet (URL du site internet)** : www.twitter.com
 - o **Disponibilité via des applications (systèmes d'exploitation concernés)** : IOs, Android
 - o **Disponibilité via des assistants vocaux** : N/A
 - o **Langue du service sur le territoire français (par défaut)** : Français
 - o **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** :
 - o **Objet** : Servir la conversation Publique
 - o **Population(s) ciblée(s) (tranches d'âges, communautés d'intérêt...)** : Toutes (15 ans et plus)
 - o **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)
 - o **Thématique(s)** : Toutes les thématiques peuvent être discutées sur notre plateforme à condition qu'elles respectent nos Conditions générales d'utilisation.
 - o **Modèle économique du service (types d'activités rémunératrices pour l'opérateur tirées de l'exploitation du service en France)** : Publicité et licences de données.
 - o **Effectifs employés aux activités du service dans le monde et en France** : Twitter emploie 4800 personnes dans le monde. L'entité française, Twitter France SAS, emploie 30 salariés.
 - o **Comptes de résultats pour l'activité du service en France en 2019** : Les résultats financiers pour 2019 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 2018 et ceux-ci sont accessibles au public et joints au questionnaire. Ils montrent les résultats suivants :
 - Chiffre d'affaires €12,695,991
 - Revenu d'exploitation €13,325,006
 - Frais d'exploitation €12,762,897
- **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 de l'histoire. 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 permettant au public de s'exprimer et de discuter en temps réel. Twitter permet aux gens de consommer, de créer, de distribuer et de découvrir des contenus et a démocratisé la création et la distribution de contenus. Periscope est une application mobile qui permet à chacun de diffuser et de regarder des vidéos en direct avec d'autres personnes. Les vidéos en direct de Periscope peuvent également être visionnées via Twitter et sur des navigateurs web de bureau ou mobiles.

1. Dispositif de signalement de fausses informations

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 (et le cas échéant, dans quels termes). Plus précisément, le Conseil souhaite connaître le nombre de motifs de signalement proposés à l'utilisateur au total ainsi que leur arborescence et s'il existe une possibilité de saisir un motif autre que ceux proposés par la plateforme.

Twitter permet de signaler des Tweets, des comptes et des Messages Privés contenant du spam, des contenus inappropriés ou préjudiciables, des usurpations d'identité, des violations du droit d'auteur ou des marques déposées. De plus, les utilisateurs peuvent signaler tout Tweet qu'ils voient sur Twitter, que ce soit des Tweets dans leur fil d'actualité, dans l'onglet Notifications ou dans la recherche Twitter. Le signalement d'un Tweet permet à l'utilisateur d'indiquer quel Tweet spécifique il pense être en violation. Les outils de signalement sont accessibles directement depuis chaque Tweet (<https://help.twitter.com/en/safety-and-security/report-a-tweet>), au niveau du compte et depuis les pages du centre d'aide (<https://help.twitter.com/forms>).

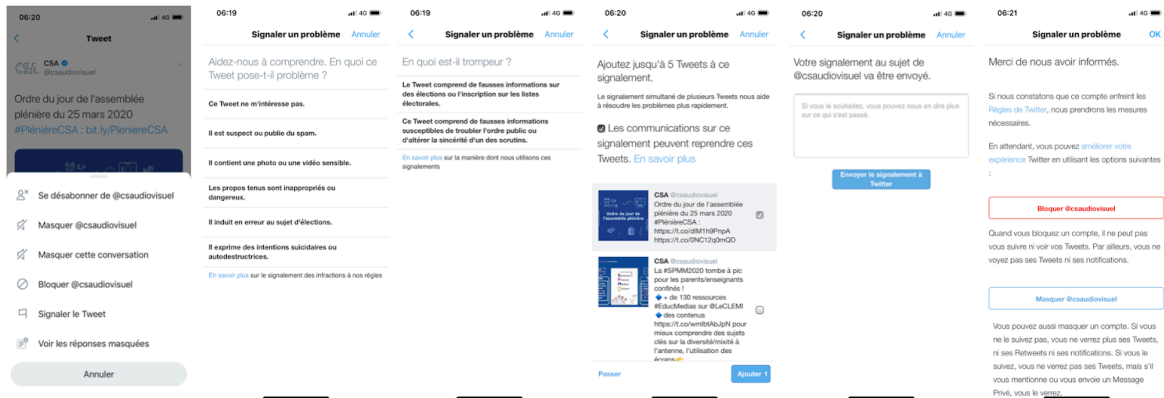
Concernant les fausses informations, Twitter permet deux types de signalement pour les utilisateurs français :

- Des fausses informations sur le vote ou l'inscription au vote, comme décrit ci-dessous.
Les règles de Twitter interdisent de publier un contenu fournissant de fausses informations sur le vote ou l'inscription au vote.
Lorsque ce type de contenu est signalé, nous examinerons le Tweet incriminé. Si nous déterminons que le Tweet est en violation de nos politiques, nous prenons des mesures (allant de l'obligation de retirer le contenu interdit jusqu'à la suspension permanente du compte). L'utilisateur recevra un suivi de notre part si nous avons besoin de plus d'informations, ou lorsque nous prendrons des mesures concernant le Tweet signalé.
- Des fausses informations susceptibles d'altérer la sincérité du vote ou de troubler l'ordre public
La loi française exige que nous fournissions également un moyen pour les utilisateurs de signaler les fausses informations qui pourraient altérer la sincérité du vote ou troubler l'ordre public.

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

- sa présentation visuelle, son affichage par défaut ou non,
- 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.

Twitter permet un accès au signalement directement au sein du Tweet en 1 clic. Le bouton de signalement est clair : "Signaler ce Tweet", et il est identifié par un drapeau. Le signalement peut être complété en 5 étapes, incluant une demande de commentaire supplémentaire et la possibilité d'ajouter des Tweets additionnels au signalement.



1.3. En termes d'accessibilité et d'ergonomie du dispositif, le Conseil souhaite savoir :

- combien d'actions et d'hyperliens nécessite l'envoi du signalement d'une fausse information depuis la page du contenu,
- si la procédure de signalement prévoit une demande de confirmation de l'envoi à l'utilisateur avant la finalisation du signalement.

Après avoir accédé à la procédure de signalement et pour le finaliser, l'utilisateur devra identifier dans une liste la raison de son signalement. Celui-ci pourra fournir des informations supplémentaires sur le problème qu'il signale. Nous pouvons également lui demander de sélectionner des Tweets supplémentaires provenant du compte qu'il rapporte afin d'avoir un meilleur contexte pour évaluer le signalement.

Nous incluons le texte des Tweets signalés dans nos emails de suivi ainsi que dans les notifications à l'utilisateur. Une fois le signalement soumis, nous fournirons des recommandations sur les mesures supplémentaires à prendre pour améliorer l'expérience de l'utilisateur, comme l'utilisation de nos fonctionnalités pour masquer ou bloquer.

1.4. S'agissant du traitement d'un signalement d'une fausse information, l'opérateur adressera au Conseil une description de l'ensemble de la procédure de traitement des signalements, et notamment:

- avant le déclenchement de cette procédure: les outils (techniques, scientifiques, journalistiques, lignes directrices internes, etc.) permettant d'instruire les signalements ; les critères d'ouverture de l'instruction d'un signalement ;
- dès le déclenchement de la procédure :
 - o l'organisation et les étapes de la procédure d'instruction, notamment s'il existe une procédure d'instruction spécifique en période électorale ou en cas d'urgence (ex.: sanitaire),
 - o les moyens humains mis en œuvre dans le cadre de procédure, en précisant le nombre de personnes en charge de cette activité, leur formation, leur localisation ainsi que les modalités d'organisation de leur travail (notamment si elles impliquent une astreinte) et leur statut interne ou externe à la société,
 - o les moyens automatisés mis en œuvre dans le cadre de procédure, en précisant leur nature, leur nombre et leur fonctionnement,
 - o à quelles étapes et par quelle voie l'auteur du signalement et/ou celui du contenu signalé sont tenus informés par l'opérateur des actions que celui-ci a entreprises ou décidées ;
- à la fin de la procédure :



- o la liste des actions que peut mettre en œuvre l'opérateur vis-à-vis des contenus signalés dont il a considéré à l'issue de la procédure qu'ils contenaient de fausses informations, ainsi que des comptes sur lesquels ils ont été publiés,**
- o si les résultats de la procédure sont portés à la connaissance des autres utilisateurs, notamment par l'affichage d'un message sur le contenu signalé,**
- o si la mise en œuvre de la décision de l'opérateur s'effectue par des moyens humains et/ou automatisés,**
- o l'existence de voies de recours pour les auteurs du contenu signalé et du signalement.**

La loi française exige que nous fournissions également un moyen pour les utilisateurs de signaler les fausses informations pouvant altérer la sincérité du vote ou troubler l'ordre public. Ces signalements fournissent des informations utiles pour guider l'application de nos règles relatives à la désinformation en France. Nous les utilisons pour informer sur la manière dont nous défendons notre plateforme contre les manipulations. Mais nous n'examinons pas et ne prenons pas d'action sur ces signalements spécifiques sur une base individuelle.

Nous affirmons que la désinformation est un problème complexe et sociétal qui mérite une réponse sociétale. À cette fin, nous nous engageons à travailler avec la Commission européenne, les gouvernements, la société civile et les institutions universitaires afin de mettre au point des solutions collaboratives et réalisables. Nous avons signé le Code de conduite de l'Union Européenne sur la désinformation, nous avons établi un partenariat avec l'UNESCO sur l'éducation aux médias et nous publions l'ensemble de données sur les campagnes d'ingérence étrangères et ce, afin de promouvoir une compréhension commune des tactiques et des menaces des acteurs malveillants.

Nous ne nous reposons pas non plus uniquement sur les seules interactions des utilisateurs. Nous prenons désormais des mesures proactives pour tenter d'empêcher la propagation de comptes et de Tweets malveillants, afin de garder une longueur d'avance sur les tactiques en constante évolution des acteurs mal intentionnés. Pour atteindre cet objectif, nous avons apporté un certain nombre de changements notamment des politiques de sécurité renforcées, de meilleurs outils et des ressources pour détecter et arrêter les activités malveillantes, des normes de publicité plus strictes et une transparence accrue pour favoriser la compréhension du public sur tous ces sujets. Ces politiques sont présentées plus en détail ci-dessous.

Notre équipe Site Integrity (Intégrité du Site) conduit nos efforts proactifs visant à protéger la conversation publique contre toute manipulation. Ils sont en mesure de tirer parti des compétences de l'ensemble de l'entreprise et de s'associer à des experts techniques et analystes pour répondre rapidement à l'intensification d'activités automatisées ou coordonnées par des personnes, inauthentiques ou malveillantes, sur la plateforme. Le travail de l'équipe nous permet de mieux comprendre la nature des activités malveillantes et de les atténuer plus rapidement.

Pendant les élections, l'équipe Site Integrity enquête, répond aux potentiels cas d'activités coordonnées inauthentiques liées aux élections dans les conversations politiques, et effectue des analyses approfondies des comportements concernés.

Pour compléter ses propres analyses, l'équipe d'analystes de Twitter reçoit et répond également aux signalements provenant de toute l'entreprise mais aussi de tiers externes.

Grâce à l'examen et à la détection en temps réel d'activités automatisées ou humaines, coordonnées, anormales et potentiellement malveillantes, notre équipe s'efforce d'identifier et de traiter toute tentative d'ingérence dans le processus électoral par des acteurs malveillants. Elle est également mieux informée sur l'endroit et la manière de déployer des ressources pour examiner de manière proactive les activités malveillantes potentielles, et ainsi donner des conseils sur les changements de politique, les évolutions produits et les partenariats possibles.



1.5. Le Conseil souhaite disposer d'informations chiffrées sur la mise en œuvre du dispositif de signalement de fausses informations en 2019, notamment :

- o le nombre de contenus signalés comme fausses informations**
- o le nombre de procédures de traitement mises en œuvre,**
- o le nombre de contenus considérés comme étant de fausses informations à l'issue de la procédure de traitement,**
- o le délai moyen de traitement d'un signalement,**
- o le nombre de recours reçus par l'opérateur à l'issue de la procédure de traitement, o**
- le budget alloué au traitement des signalements par les utilisateurs.**

Les personnes qui utilisent Twitter ne peuvent pas se servir de Twitter dans le but de manipuler ou d'interférer dans des élections. En avril 2019, nous avons publié une nouvelle politique concernant l'intégrité des élections, celle-ci régissant trois catégories de comportements et de contenus manipulateurs liés aux élections. Premièrement, un individu ne peut pas partager des informations fausses ou trompeuses sur la façon de participer à une élection ou à tout événement civique. Cela comprend, sans s'y limiter, les informations trompeuses sur la façon de voter ou de s'inscrire pour voter, les conditions de vote (y compris les conditions d'identification) ainsi que les informations officielles telles que la date ou les heures d'une élection. Deuxièmement, une personne ne peut pas communiquer d'informations fausses ou trompeuses, destinées à intimider ou à dissuader les électeurs de participer à une élection. Cela inclut, sans s'y limiter, à des 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 affirmations trompeuses concernant les activités de la police ou des forces de l'ordre liées aux bureaux de vote ou aux élections, les longues files d'attente, les problèmes d'équipement, les procédures ou techniques de vote, les menaces concernant les lieux de vote qui pourraient dissuader les électeurs de participer à une élection. Enfin, nous n'autorisons pas non plus la création de faux comptes qui dénaturent leur affiliation, ni le partage de contenu qui représente faussement leur affiliation à un candidat, un élu, un parti politique, une autorité électorale ou une entité gouvernementale. Nous avons encore étendu nos capacités d'action autour des élections en créant une fonction de signalement dédiée au sein de la plateforme afin de permettre aux utilisateurs de signaler plus facilement les contenus qui compromettent le processus d'inscription au vote, ou d'engagement dans le processus électoral. En 2019, nous avons reçu 49 945 signalements d'utilisateurs par le biais de notre fonctionnalité de signalement sur la désinformation liée aux élections pour l'Union européenne. Nous avons reçu 14 529 signalements par le biais de l'option de signalement de fausses informations spécifique à la France.

1.6. Enfin, le Conseil souhaite savoir si l'opérateur a engagé des discussions avec d'autres opérateurs soumis au titre III de la loi visant à une harmonisation de leur dispositif de signalement des fausses informations respectif et, le cas échéant, quelles ont été les mesures prises dans le cadre de ces discussions.

La sérénité de la conversation publique est la priorité de Twitter. Pour atteindre cet objectif, les signalements des utilisateurs sont toujours très importants pour nous. Il est essentiel à nos yeux, dans ce processus, d'y répondre au mieux, le plus rapidement et de la manière la plus transparente possible. Au fil des ans, nous avons fait en sorte que notre processus de signalement soit facile à utiliser et à comprendre.

Chaque produit étant différent, nous pensons qu'il est important que le processus de signalement soit aussi adapté que possible au type de contenu, qu'au design de la plateforme où il est mis en œuvre. Twitter a mis la priorité sur un processus de signalement facilement accessible, simple et didactique pour les utilisateurs de la plateforme. Celui-ci 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 les autres opérateurs.



2. Transparence des algorithmes

Les indications suivantes s'appliquent aux opérateurs qui auraient mis en œuvre des mesures relatives à la transparence des algorithmes au titre des mesures complémentaires prévues par l'article 11 de la loi du 22 décembre 2018.

2.1. Au préalable, il est nécessaire au Conseil, pour sa bonne appréciation des mesures prises par l'opérateur en la matière, que ce dernier lui fournisse la liste des algorithmes qui régissent l'organisation, la sélection et l'ordonnancement des contenus sur le service, en particulier ceux participant à la lutte contre la manipulation de l'information (par exemple en détectant des contenus ou en freinant leur propagation). Cette liste prendra la forme d'une courte description textuelle comportant pour chaque algorithme :

- son ou ses objectif(s),
- ses grands principes de fonctionnement, en détaillant notamment leur(s) critère(s) (définis succinctement), le poids de ces critères dans l'algorithme considéré ainsi que les données qu'ils utilisent, la façon dont elles sont collectées, leur durée de conservation, ainsi que la longueur de l'historique utilisée par l'algorithme,
- les effets de l'algorithme, soit la manière dont les résultats sont utilisés par la plateforme en lien avec les objectifs précités,
- si cet algorithme est personnalisable par l'utilisateur et/ou s'il est ajusté aux données propres à chaque utilisateur
- enfin, le nombre de changements opérés dans ces algorithmes au cours de l'année d'exercice, notamment lorsqu'ils concernent la lutte contre la manipulation de l'information.

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 postés chaque jour sur Twitter, nous avons beaucoup investi dans la mise en place de systèmes organisant le contenu sur Twitter afin de montrer aux personnes qui se connectent à la plateforme les informations les plus pertinentes pour elles avant tout. Avec 152 millions d'utilisateurs quotidiens utilisant Twitter dans des dizaines de langues et reflétant d'innombrables contextes culturels, nous nous appuyons sur des algorithmes de machine learning pour nous aider à organiser le contenu par pertinence. Si un individu souhaite accéder à son contenu sans qu'aucun algorithme ne soit appliqué, il dispose d'un contrôle unique accessible afin de visualiser son fil d'actualité dans l'ordre anté-chronologique.

Cela permet aux personnes sur Twitter de comprendre comment les algorithmes affectent ce qu'elles voient, et offre une plus grande transparence dans la technologie que nous utilisons pour classer les Tweets.

En plus du fil d'actualité d'accueil, Twitter dispose d'une chronologie de notifications permettant aux gens de voir qui a aimé, Retweeté et répondu à leurs Tweets, ainsi que ceux les ayant mentionnés ou suivis. Nous donnons aux utilisateurs de Twitter des contrôles supplémentaires sur le contenu qui apparaît dans la chronologie de leurs notifications, puisque celles-ci peuvent contenir du contenu qu'une personne sur Twitter n'a pas choisi de recevoir, 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 nous excluons les notifications concernant les Tweets en double ou potentiellement spammeurs. Nous permettons également aux utilisateurs de la plateforme de contrôler de façon granulaire des comptes spécifiques dont elles pourraient ne pas vouloir recevoir de notifications, comprenant les nouveaux comptes, les comptes que la personne ne suit pas et les comptes sans téléphone ou adresse électronique confirmés.

Les réponses à un Tweet sont appelées "conversation". Twitter s'efforce de montrer aux gens les contenus qui, selon nous, les intéresseront le plus et qui contribueront de manière significative à la conversation. C'est pourquoi les réponses, regroupées par sous-conversations, peuvent ne pas être dans l'ordre chronologique. Par exemple, lorsque nous classons une réponse en tête des réponses, nous prenons en compte des facteurs tels que la réponse de l'auteur du Tweet original, ou si la réponse vient de quelqu'un que la personne suit.



Afin de préserver l'intégrité de notre plateforme et de protéger les conversations contre la manipulation, Twitter utilise également des outils et des technologies pour détecter et minimiser la visibilité de certains types de comportements abusifs et manipulateurs.

Twitter utilise une série de signaux comportementaux pour déterminer comment les Tweets sont organisés et présentés dans le fil d'actualité, dans les conversations et les recherches, en fonction de leur pertinence. Twitter s'appuie sur des signaux comportementaux, tels que le comportement et les réactions des comptes entre eux, pour identifier les contenus tels que le spam et les abus qui nuisent à une conversation publique sereine. À moins que nous n'ayons déterminé qu'un Tweet contrevient aux politiques de Twitter, il restera sur la plateforme. Lorsque nous avons identifié un Tweet comme étant potentiellement préjudiciable à une conversation sereine (par exemple, comme potentiellement abusif), il sera disponible uniquement 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 dont nous nous servons, en les associant les uns avec les autres ainsi qu'avec une série d'autres signaux, pour aider à identifier ce type de contenu : un compte sans adresse électronique confirmée, une inscription simultanée pour plusieurs comptes, des comptes qui Tweetent et mentionnent de manière répétée des comptes qui ne les suivent pas, ou encore, 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 les uns avec les autres. La précision des algorithmes développés à partir de ces signaux comportementaux continuera à s'améliorer avec le temps.

Ces signaux comportementaux sont un facteur important dans la façon dont Twitter organise et présente le contenu dans les espaces communs comme la conversation et la recherche. Notre objectif premier est de faire en sorte que les contenus pertinents et les Tweets contribuant à une conversation sereine apparaissent en premier dans ces espaces (conversation et recherche). Comme notre service fonctionne dans des dizaines de langues et des centaines de contextes culturels dans le monde entier, nous avons constaté que le comportement est un signal fort qui nous aide à identifier les acteurs malveillants sur notre plateforme. Les signaux comportementaux utilisés par Twitter ne tiennent aucunement compte des opinions politiques ou des idéologies. Ils se concentrent uniquement sur le comportement de tous les comptes. Twitter s'efforce en permanence d'améliorer ses modèles de classement basés sur le comportement, de sorte que leur portée et leur précision s'améliorent avec le temps. Nous utilisons des milliers de signaux comportementaux dans nos modèles de classement basés sur le comportement, afin de garantir qu'aucun signal seul 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 déduit au sujet du compte ou de l'appareil utilisé. Ces intérêts contribuent à améliorer l'expérience Twitter en affichant, par exemple, un meilleur contenu, notamment des publicités, des notifications et des Tweets recommandés dans le fil d'actualité et l'onglet Explorer. Sont également disponibles les annonceurs Twitter qui ont inclus le compte ou l'appareil actuel dans leur audience personnalisée.

<https://help.twitter.com/fr/managing-your-account/accessing-your-twitter-data>

2.2. L'opérateur décrira l'information qu'il apporte aux utilisateurs sur :

- les données utilisées par les algorithmes de recommandation et de hiérarchisation (notamment : liste des données utilisées ; explication sur la manière dont les données sont utilisées par le(s) algorithme(s), en particulier lorsque certaines données sont utilisées par certains algorithmes au contraire d'autres),**
- les critères utilisés par les algorithmes de recommandation et hiérarchisation des contenus (notamment : informations générales ; liste des critères utilisés ; impact des critères sur le fonctionnement de l'algorithme qui les utilise),**



- les changements opérés dans les algorithmes de recommandation et de hiérarchisation (notamment : types de changements concernés ; liste des changements ; impact de ces changements sur les résultats de(s) algorithmes(s) concernés).

Voir réponse au 2.1.

2.3. En outre, si l'utilisateur peut faire des réglages aux fins de modifier le référencement et la recommandation opérés par les algorithmes du service et de personnaliser tout ou partie des critères utilisés à des fins de sélection et d'ordonnement des contenus par les algorithmes, l'opérateur indiquera si une explication est apportée à l'utilisateur sur les effets produits par les réglages sélectionnés avant qu'ils ne soient effectifs.

Twitter fournit une grande variété d'explications sur l'organisation du contenu dans son centre d'aide, notamment :

<https://help.twitter.com/en/using-twitter/twitter-timeline>

<https://help.twitter.com/en/safety-and-security/control-your-twitter-experience>

2.4. Le Conseil invite l'opérateur à lui indiquer où sont disponibles les informations mentionnées ci-dessus, et notamment si l'utilisateur peut les trouver au sein même du service et si elles figurent dans un espace unique.

Toutes les informations relatives à Twitter sont disponibles à l'adresse www.help.twitter.com
Les informations sont organisées par catégorie ou recherchable.

En juin 2019, nous avons actualisé nos règles avec un langage simple et clair. Nous les avons réorganisées en catégorie : sécurité, vie privée et authenticité. Nous avons également ajouté des renseignements concernant d'autres politiques, notamment l'intégrité des élections, la manipulation de la plateforme et le spam.

Ce faisant, nous sommes passés d'environ 2 500 mots à moins de 600. En 280 caractères ou moins, chaque règle décrit clairement ce qui n'est pas autorisé sur Twitter. Grâce aux nouvelles catégories - sécurité, confidentialité et authenticité - nos règles sont organisées par thème, ce qui permet de trouver plus rapidement les informations recherchées. Nous mettons à jour les pages de nos règles pour fournir plus d'informations (incluant des exemples), des instructions étape par étape sur la façon de nous faire un signalement ainsi que des détails sur ce qui se passe lorsque nous prenons des actions.

Twitter publie également des informations sur ses activités via son blog (en anglais et en français). Les informations sont également partagées par le biais des comptes Twitter : @twitter, @TwitterFrance, @TwitterSafety...

2.5. Par ailleurs, l'opérateur indiquera s'il existe un outil d'interaction en temps réel entre l'utilisateur et l'opérateur aux fins notamment d'obtenir des informations personnalisées et précises sur le fonctionnement des algorithmes de recommandation et de hiérarchisation des contenus. Le Conseil invite la plateforme à lui préciser, dans l'hypothèse où un tel outil existerait:

- la présentation de cet outil ainsi que son fonctionnement (informations que l'utilisateur doit fournir, contact avec une personne humaine ou non, etc.),
- si les réponses fournies sont standardisées ou personnalisées, et comment,
- les délais moyens de réponses d'une part, et de traitement d'autre part, des requêtes reçues
- sa politique en matière de traitement de ces requêtes, notamment quant au fait de ne pas répondre à certaines d'entre elles.

Les utilisateurs de Twitter peuvent trouver des informations sur la personnalisation, la confidentialité et le contrôle de leur expérience grâce à notre centre d'aide. Le compte @TwitterSupport fournit également des conseils et des mises à jour sur l'expérience utilisateur.



2.6. Enfin, le Conseil souhaite savoir comment l'opérateur fait en sorte de rendre l'ensemble de ces informations relatives aux algorithmes aisément compréhensibles par l'utilisateur.

Comme décrit au point 2.4, Twitter a actualisé ses règles avec un langage simple et clair, et les a réorganisées en catégories : sécurité, vie privée et authenticité. Nous avons également ajouté des renseignements concernant d'autres politiques, notamment l'intégrité des élections, la manipulation de la plateforme et le spam.

Ce faisant, nous sommes passés d'environ 2 500 mots à moins de 600. En 280 caractères ou moins, chaque règle décrit clairement ce qui n'est pas autorisé sur Twitter. Twitter travaille constamment à mettre à jour les règles et les informations afin de les rendre compréhensibles par les utilisateurs.

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

Les indications suivantes s'appliquent aux opérateurs qui auraient mis en œuvre des mesures relatives à la promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle au titre des mesures complémentaires prévues par l'article 11 de la loi du 22 décembre 2018.

Twitter, de par son caractère public, largement distribué et en temps réel, est une plate-forme importante qui permet aux journalistes et aux éditeurs de presse de partager librement leur contenu et aux utilisateurs d'accéder aux informations publiques en temps réel. Un grand nombre des médias et éditeurs les plus fiables du monde utilisent régulièrement Twitter comme plateforme de distribution de contenu.

Comme Twitter n'est ni un agrégateur de presse ni un éditeur de contenu, Twitter n'offre pas d'outil de vérification des faits (fact-checking) aux éditeurs. En fait, Twitter n'opère pas de recherche de contenu des éditeurs de presse en vue de leur indexation et mise à disposition 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 l'intermédiaire de ses abonnés et un public plus large par l'engagement avec d'autres utilisateurs sur la plateforme.

Les organisations de fact-checking sur Twitter partagent le contenu à grande échelle, recherchent des sujets et répondent aux questions ou au contenu d'utilisateurs. Certaines 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 des couleurs du profil ou du thème.

Bien que Twitter ne vérifie pas la véracité des contenus partagés, nous avons établi des partenariats avec des organisations fiables de vérification des faits pendant les périodes de risque aigu, qui sont intégré sur notre plateforme en tant que journalistes de confiance. Ils soutiennent la plateforme en signalant les Tweets qui sont en violation avec nos politiques concernant la manipulation de la plateforme et peuvent les faire remonter plus rapidement. Un partenaire important dans ce domaine est le DisinfoLab de l'UE, qui a travaillé activement avec nous lors des dernières élections européennes. Le EU DisinfoLab est une organisation non gouvernementale basée à Bruxelles dont la mission est de lutter contre la désinformation grâce à une méthodologie innovante et un soutien scientifique à la communauté de la contre-désinformation.

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 et agences de presse et SCA, ou un tiers) et, dans l'hypothèse où l'opérateur ne la réalise pas directement, s'il la vérifie (avant ou après publication),
- par quels moyens et sur la base de quelles sources elle est réalisée,
- comment elle est portée à la connaissance de l'utilisateur sur le service,
- les moyens humains et technologiques déployés par l'opérateur afin d'assurer 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.

N/A

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. Outre la description de ces moyens mis en œuvre (ex. : réalisation directe, financement, partenariats, mise en avant sur le service (et comment), informations aux utilisateurs, etc.), l'opérateur précisera, le cas échéant, sur quels critères sont sélectionnées les démarches de labellisation concernées.

L'opérateur est invité à préciser si ces démarches constituent un des moyens d'identification des sources de contenus issus d'entreprises et d'agences de presse et de SCA et le cas échéant, de quelle manière.

N/A

3.3. Le Conseil s'interroge sur les dispositifs mis en œuvre par l'opérateur pour œuvrer au « fact-checking ». Il souhaite connaître les partenariats et/ou actions (en interne et externe) existants en matière de « fact-checking » (notamment aux fins de lutter contre les hypertrucages ou « deepfakes ») et en particulier :

- les critères retenus par l'opérateur pour sélectionner les « fact-checkers » (notamment : nationalité et langue de travail pour un partenariat sur une zone géographique donnée) ainsi que les contreparties qui leur sont offertes et dans quelles conditions,
- si les contenus sur la plateforme pouvant être soumis aux « fact-checkers » sont laissés à leur libre appréciation ou limités (le cas échéant, selon quels critères),
- 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).

N/A

3.4. Le cas échéant, il serait utile que le Conseil dispose de :

- la liste des structures de « fact-checking » partenaires de la plateforme ;
- la liste des contenus suggérés par la plateforme aux « fact-checkers » pour vérification et la liste des contenus proposés à la vérification par les « fact-checkers »,
- le nombre de contenus ayant fait l'objet d'une intervention des « fact-checkers » sur la plateforme (décliné par types d'interventions possibles),
- le budget consacré par l'opérateur au « fact-checking » (incluant les contreparties financières éventuellement versées aux « fact-checkers »).

N/A

3.5. L'opérateur est invité à décrire les moyens technologiques et humains permettant de traiter les contenus dont le « fact-checking » a montré qu'ils contenaient ou étaient susceptibles de contenir une fausse information, notamment les mesures prises le cas échéant (ex.:



déréférencement) et leur gradation, ainsi que l'existence de voies de recours pour les auteurs des contenus concernés.

N/A

3.6. Le Conseil souhaite également connaître la 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 (ex. : commentaire, like).

N/A

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

Les indications suivantes s'appliquent aux opérateurs qui auraient mis en œuvre des mesures relatives à la lutte contre les comptes propageant massivement de fausses informations au titre des mesures complémentaires prévues par l'article 11 de la loi du 22 décembre 2018.

4.1. Le Conseil souhaite savoir comment l'opérateur appréhende la notion de comptes propageant massivement de fausses informations et quelles sont les mesures et moyens (humains et technologiques) qu'il a pris afin de :

- détecter de tels comptes (ainsi que ceux ayant recours aux « deepfakes »),
- faire obstacle à ces comptes (notamment la procédure, les mesures prises et leur gradation : avertissement, restriction des droits, mise en quarantaine, déréférencement, suppression du contenu, du compte, etc.) et à l'accélération et à la viralité des contenus contenant de fausses informations qu'ils véhiculent.

La manipulation des plateformes peut prendre de nombreuses formes et nos règles visent à traiter un large éventail de comportements interdits, notamment le spam à motivation commerciale, qui vise généralement à détourner le trafic ou 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 paraître les comptes ou le contenu plus populaires ou plus actifs qu'ils ne le sont ; les activités coordonnées, qui tentent d'influencer artificiellement les conversations par l'utilisation de comptes multiples, de faux comptes, d'automatisation et/ou de script.

Twitter continue de développer les outils et systèmes de détection nécessaires pour lutter contre l'automatisation malveillante de notre service. Par exemple, Twitter donne la priorité à l'identification des activités suspectes sur les comptes, comme le fait de Tweeter à un volume exceptionnellement élevé avec le même hashtag ou de mentionner le même @nomdutilisateur sans réponse du compte concerné. Lorsque nous identifions une telle activité, nous pouvons demander à une personne utilisant le service de confirmer le contrôle humain du compte ou de son identité. Twitter a également accru son travail de vérification destiné à détecter les comptes automatisés, tels que les reCAPTCHA, qui exigent que les individus identifient des parties d'une image ou tapent des mots affichés à l'écran, et les demandes de réinitialisation de mot de passe qui protègent les comptes potentiellement compromis.

Chaque mois, nous vérifions des millions de comptes qui tentent de spammer ou de manipuler la plateforme Twitter de quelque manière que ce soit. Une vérification antispam implique des tâches comme la confirmation d'un numéro de téléphone ou la résolution d'un reCAPTCHA - de telles tâches sont faciles à réaliser pour de vraies personnes, mais difficiles ou coûteuses pour les comptes de spam.

Au cours du premier semestre 2019, nous avons lancé en moyenne 16 millions de vérification de ce type par mois. Bien que nous ayons observé une diminution globale des défis antispam lancés en 2019 par rapport à 2018, nous attribuons cette diminution, au moins en partie, à l'attention particulière portée à la dissuasion de la création de faux comptes lors de l'inscription plutôt qu'à la réponse au comportement d'un compte après l'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 d'environ 144 000 applications pendant une période de six mois allant du premier au troisième trimestre de 2019. Nous y sommes parvenus grâce à des mesures à la fois automatisées et proactives.

En outre, notre processus amélioré d'intégration des développeurs est conçu pour empêcher les mauvais acteurs d'accéder à notre plateforme API dès l'origine, et depuis le lancement du programme en juillet 2018, environ 800 000 cas d'utilisation ont été examinés et vérifiés au regard de leur conformité aux politiques avant que le développeur ne se voie accorder ou refuser l'accès.

Notre motivation avec ces changements est de réduire la charge qui pèse sur les personnes qui utilisent Twitter pour signaler les spams et les automatisations malveillantes. Ces améliorations technologiques ont entraîné une réduction correspondante du nombre de signalements de spam par les personnes utilisant Twitter. Les signalements de spam sont des signalements de personnes qui utilisent Twitter après avoir reçu une interaction, telle qu'un abonnement, une mention ou un message direct, d'un compte suspecté de spam. Au cours du premier semestre 2019, 4 544 096 signalements de spam ont été soumis par des utilisateurs après que des personnes utilisant Twitter aient reçu une interaction provenant d'un compte suspecté de spam.

Nous avons également supprimé les comptes verrouillés du décompte des abonnés afin que ces chiffres reflètent plus précisément la portée et la popularité réelles d'un compte. Les comptes sont verrouillés lorsque nos systèmes détectent une activité inhabituelle et forcent un changement de mot de passe ou une autre vérification. Si la vérification n'a pas été relevée 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 des Likes à d'autres personnes. La suppression des comptes verrouillés des comptes d'abonnés permet de s'assurer que les gens peuvent faire confiance à la fiabilité des informations des profils sur Twitter.

Les tactiques de manipulation, de spam et d'abus d'API employées par les mauvais acteurs sont en constante évolution, et nous nous engageons à améliorer sans cesse notre réponse et leur prévention.

4.2. S'agissant de l'information aux utilisateurs sur les mesures de détection et de traitement des comptes propageant massivement de fausses informations, l'opérateur décrira les dispositifs de suivi et de statistiques de ces mesures, 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 auxquelles ils donnent accès : 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,
- l'accès et le traitement possible de ces données par l'utilisateur, via une fonctionnalité sur le service ou permettant de les télécharger.

La transparence est au cœur du fonctionnement de Twitter. Notre rapport de transparence fournit des détails sur le nombre de comptes qui sont vérifiés par des systèmes automatisés en tant que spam potentiel.

Afin de permettre aux universitaires et au public de mieux comprendre les opérations d'information dans le monde entier et de permettre un examen indépendant de ces tactiques par des tiers sur notre plateforme, nous publions sur Twitter des archives complètes des opérations d'information soutenues par des États. Nous avons mis à disposition tous les comptes et contenus associés à des opérations d'information potentielles que nous avons trouvés sur notre service depuis 2016 et nous continuons à publier de nouvelles données à mesure que nous détectons de nouvelles activités.

4.3. Le Conseil examinera 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 poussée à l'utilisateur (ou nécessite-t-elle une recherche de sa part) et, le cas échéant, dans quelle(s) situation(s) ?**

Dans le centre d'assistance, Twitter fournit une description détaillée des comportements interdits en ce qui concerne la manipulation de notre plateforme et les politiques de spam. Une section "Que se passe-t-il si vous violez cette politique" indique également quelles sont les conséquences des violations mentionnées ci-dessus.

Comme pour toutes les règles de Twitter, ces explications ont été rafraîchies avec un langage simple et clair et des exemples de ces comportements.

<https://help.twitter.com/en/rules-and-policies/platform-manipulation>

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 :

- **le nombre de comptes propageant massivement de fausses informations détectés sur le service, et la proportion qu'ils représentent par rapport au 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'un traitement 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,**
- **le cas échéant, les recettes générées pour la plateforme par de tels comptes et celles qu'elle leur a reversées en 2019.**

Chaque mois, Twitter vérifie des millions de comptes qui tentent de spammer ou de manipuler la plateforme de quelque manière que ce soit.

Number of accounts challenged monthly Jan 1 - Jun 30, 2019

Month	Challenges
January	19,522,273
February	17,036,751
March	16,552,753
April	13,787,154
May	14,848,125
June	15,376,611
TOTAL	97,123,667

Depuis juin 2017, nous avons également suspendu plus de 117 000 applications malveillantes pour abus de notre API, qui sont collectivement responsables de plus de 1,5 milliard de Tweets de mauvaise qualité.

https://about.twitter.com/en_us/advocacy/elections-integrity.html#data

Tout au long de l'année 2019, nous nous sommes également appuyés sur les réalisations de l'année précédente, lorsqu'en octobre 2018, nous avons divulgué sur Twitter les premières archives complètes des opérations d'information soutenues par des entités étatiques.

Ces archives sont aujourd'hui les plus importantes de ce type dans le secteur. En 2019, de nouveaux corpus de données ont été ajoutés à l'archive, permettant ainsi au public, aux gouvernements, aux universitaires et aux chercheurs d'étudier ces comportements, d'apprendre et de renforcer les capacités d'éducation aux médias pour l'avenir afin de lutter contre la manipulation et la désinformation.



Les données comprennent plus de 160 millions de Tweets et plus de huit téraoctets de médias. En 2019, les opérations d'interférence étrangère potentielles que nous avons observées sur Twitter sont les suivantes :

Bangladesh

En collaboration avec nos pairs du secteur, nous avons identifié et suspendu un très petit nombre de comptes provenant du Bangladesh pour s'être livrés à des manipulations coordonnées de la plateforme. Les Tweets étaient entièrement en bengali et portaient sur des thèmes politiques régionaux.

Iran

En août 2018, nous avons été informés par un pair de l'industrie d'une possible activité malveillante sur leur service concernant une tentative de campagne d'influence que nous avons identifiée comme potentiellement située en Iran. Après avoir reçu des informations de leur part, nous avons entamé une enquête sur notre service afin de mieux comprendre ces réseaux. Dès que nous avons découvert une activité malveillante, nous avons immédiatement averti les forces de l'ordre et informé le public. Depuis lors, grâce à une analyse scientifique continue des données, nous avons identifié et suspendu 2 617 comptes malveillants supplémentaires qui, selon nous, pourraient avoir des origines en Iran. Twitter a été en contact étroit avec nos pairs du secteur à ce sujet et a partagé avec eux des informations détaillées sur les comptes malveillants. Ce processus multilatéral de partage d'informations se poursuit, nous permettant ainsi qu'à nos pairs du secteur de travailler ensemble pour mieux comprendre et identifier les activités malveillantes.

Russie

En septembre 2018, Jack Dorsey a témoigné au Capitole sur les activités de l'Agence russe de recherche sur Internet (IRA), révélant que Twitter avait suspendu un total de 3 841 comptes. Grâce aux analyses et enquêtes en cours, nous continuons à nous appuyer sur notre compréhension contextuelle de ces réseaux de comptes pour améliorer notre capacité à trouver et suspendre cette activité le plus rapidement possible à l'avenir. Ceci est particulièrement vital alors que des groupes tels que l'IRA font évoluer leurs pratiques en réponse aux efforts de suspension de l'ensemble du secteur. Nos efforts en cours ont permis de découvrir 418 comptes supplémentaires. Nous ne pouvons pas attribuer définitivement ces comptes à l'IRA, bien que la plupart semblent provenir de Russie, et qu'une grande partie du comportement imite l'activité de comptes antérieurs liés à l'IRA. Nous encourageons les chercheurs et autres experts à les étudier et à contribuer à la sensibilisation du grand public à ces activités en ligne.

Venezuela

Twitter a supprimé 764 comptes situés au Venezuela. Nous ne sommes pas en mesure de lier définitivement les comptes situés au Venezuela aux opérations d'information d'un gouvernement étranger contre un autre pays. Cependant, ces comptes sont un autre exemple de campagne étrangère de contenu spammé axée sur des thèmes politiques de division, et le comportement que nous avons découvert est similaire à celui utilisé par les comptes potentiels de l'IRA russe. Nous les avons divulgué par excès de prudence et nous attendons avec attention les analyses des chercheurs. En outre, nous avons supprimé 1 196 comptes situés au Venezuela qui semblent être engagés dans une campagne d'influence soutenue par l'État et visant des publics nationaux. Nous avons partagé des informations sur ces comptes avec nos pairs de l'industrie et nous continuons à enquêter sur les activités malveillantes provenant du Venezuela, ciblant à la fois des publics au Venezuela et à l'étranger.

Émirats arabes unis et Égypte

Nous avons supprimé un réseau de 271 comptes provenant des Émirats arabes unis (EAU) et de l'Égypte. Ces comptes étaient interconnectés dans leurs objectifs et leurs tactiques : une opération d'information à multiples facettes visant principalement le Qatar, et d'autres pays comme l'Iran. Elle a également amplifié les messages de soutien au gouvernement saoudien. Nous avons également trouvé des preuves que ces comptes ont été créés et gérés par DotDev, une entreprise technologique privée opérant aux EAU et en Égypte. Nous avons suspendu définitivement DotDev, et tous les comptes qui y sont associés, de notre service. En outre, nous avons suspendu un groupe distinct de 4 248 comptes opérant uniquement à partir des EAU, principalement dirigés vers le Qatar et le Yémen. Ces comptes employaient souvent de fausses



personnes et Tweetaient sur des questions régionales, telles que la guerre civile yéménite et le mouvement Houthi.

Arabie Saoudite

Nos enquêtes ont également permis de détecter un petit groupe de six comptes liés à l'appareil médiatique public saoudien qui étaient engagés dans des efforts coordonnés pour amplifier des messages bénéfiques au gouvernement saoudien. Bien qu'actifs, les comptes-rendus de cet ensemble se présentaient comme des organes de presse indépendants tout en Tweetant des récits favorables au gouvernement saoudien.

Par ailleurs, nous avons également suspendu de façon permanente le compte Twitter de Saoud al-Qahtani pour violation de nos politiques de manipulation des plateformes. Ce compte n'est pas inclus dans les archives divulguées.

Nous continuons à suspendre une série de comptes pour d'autres types de violations du spam politique en Arabie Saoudite.

Espagne

Nous avons supprimé 259 comptes que nous avons identifiés comme renforçant faussement l'opinion publique en ligne en Espagne. Exploités par le Partido Popular, ces comptes ont été actifs pendant une période relativement courte, et consistaient principalement en de faux comptes se livrant à des activités de spamming ou de Retweet pour accroître l'engagement.

Équateur

Au début de l'été 2019, nous avons supprimé un réseau de 1 019 comptes en Équateur liés au parti politique PAIS Alliance. Ce réseau, composé en grande partie de faux comptes, était principalement chargé de diffuser des informations sur l'administration du président Moreno, en se concentrant sur des questions concernant les lois équatoriennes sur la liberté d'expression, la censure gouvernementale et la technologie. Les tactiques les plus couramment utilisées étaient la manipulation de hashtag et le Retweet de spam.

Chine (RPC)/Hong Kong

En août 2019, nous avons révélé que nous avons identifié un réseau de plus de 200 000 faux comptes basés en RPC qui tentaient de semer la discorde au sujet du mouvement de protestation à Hong Kong. Nous avons également publié des ensembles de données supplémentaires concernant 4 301 comptes qui ont été les plus actifs dans cette opération d'information afin de sensibiliser le public.

5. Information des utilisateurs sur la nature, l'origine, les modalités de diffusion des contenus et l'identité des personnes versant des rémunérations en contrepartie de la promotion des contenus d'information

Les indications suivantes s'appliquent aux opérateurs qui auraient mis en œuvre des mesures relatives à l'information des utilisateurs sur la nature, l'origine et les modalités de diffusion des contenus et à l'information des utilisateurs sur l'identité de la personne physique ou la raison sociale, le siège social et l'objet social des personnes morales leur versant des rémunérations en contrepartie de la promotion de contenus d'information se rattachant à un débat d'intérêt général au titre des mesures complémentaires prévues par l'article 11 de la loi du 22 décembre 2018.

5.1. En premier lieu, l'opérateur renseignera le Conseil sur le traitement des contenus sponsorisés en général et sur les informations fournies à l'utilisateur à leur sujet :

- sont-ils clairement distingués des autres contenus, et comment ?
- la plateforme permet-elle à l'utilisateur d'identifier les critères qui l'ont conduite à lui proposer tel contenu sponsorisé, et quelles sont les informations qu'elle lui fournit ?
- l'origine du contenu sponsorisé est-elle identifiée et affichée (ainsi que celle du sponsor, si elle diffère) ?
- le service fournit-il des indications sur les modalités de diffusion des 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 précise-t-il ses politiques publicitaires dans des documents publics disponibles en français, notamment en matière de lutte contre la désinformation sur son service ?**
- **où et comment la plateforme appelle-t-elle les utilisateurs à la vigilance sur les contenus sponsorisés ayant fait l'objet de signalements ?**

La politique de Twitter en matière de publicité est organisée autour de plusieurs principes clés: assurer la sécurité des utilisateurs; promouvoir des contenus honnêtes et les cibler de manière responsable; interdire la diffusion de spam, de contenus nuisibles ou d'autres contenus dérangeants; fixer des normes éditoriales élevées pour les contenus des publicités Twitter créés et être informé des processus concernant les publicités Twitter qui viennent en soutien de nos politiques en la matière.

Aussi, dans le cadre des publicités commerciales, avant qu'un compte ne puisse faire de la publicité sur Twitter, il doit obligatoirement répondre à certains critères. Par exemple, les publicités doivent inclure un lien URL qui fonctionne dans leur biographie, elles doivent représenter la marque et le produit faisant l'objet de la publicité et ne peuvent induire l'utilisateur en erreur afin d'engager avec le contenu en utilisant des propos exagérés ou sensationnels ou des incitations à agir trompeuses. En outre, les comptes nouvellement créés sont soumis à un examen pendant un certain temps avant de pouvoir commencer à faire de la publicité sur Twitter.

Lorsque les annonceurs sur Twitter choisissent de promouvoir leur contenu avec des publicités Twitter, leur contenu peut être soumis à un processus de vérification. Cette procédure vise à garantir la qualité et la sécurité de la plateforme de publicité Twitter. Ce processus permet à Twitter de vérifier que les annonceurs se conforment à ses politiques en matière de publicité. Toutes les politiques publicitaires sont régulièrement révisées et mises à jour afin d'être conformes à la législation en vigueur et aux évolutions des produits. Nous investissons beaucoup dans ces domaines pour maintenir notre produit et nos processus à jour.

Les Tweets sponsorisés à travers 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 sponsorisés, de la même manière qu'avec les contenus organiques.

Les utilisateurs peuvent voir les détails des raisons pour lesquelles ils ont reçu une publicité spécifique en cliquant sur l'annonce. Ces informations renvoient également l'utilisateur vers le paramètre "Personnaliser les publicités" qui lui permet de contrôler s'il reçoit des publicités basées sur les centres d'intérêt sur Twitter.

Les politiques relatives aux publicités sur Twitter sont accessibles à tous sur <https://business.twitter.com/en/help/ads-policies.html> et sont adaptées à la fois à la langue du pays et au regard des lois locales en vigueur.

Twitter permet de signaler les publicités directement dans l'annonce et prend des mesures si les publicités sont en violation des politiques publicitaires de Twitter.

5.2. En second lieu, le Conseil souhaite connaître la définition des contenus 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 et leur nombre sur la plateforme en 2019. L'opérateur peut également faire part au Conseil de tout système de vérification particulier qu'il aurait mis en place pour les annonceurs faisant la promotion de contenus d'information se rattachant à un débat d'intérêt général sur le service et indiquer, le cas échéant, le nombre d'annonceurs ayant fait l'objet de cette procédure spécifique.

Jusqu'au 22 novembre 2019, Twitter définissait les publicités relatives à "défense d'une cause" ou "liée à un débat d'intérêt général" comme des publicités défendant une question législative d'importance nationale, des publicités faisant référence à une élection, un parti politique ou un candidat clairement identifié. En France, Twitter a interdit ce type de publicités, à l'exception des



publicités d'appel au vote. Ce dernier type de publicité est affiché dans le Centre de Transparence des Publicités pendant 7 jours.

En novembre 2019, Twitter a pris la décision d'interdire à l'échelle globale la publicité de contenus politiques. Cette décision a été prise sur la conviction que la portée d'un message politique doit se mériter et non s'acheter.

Nous définissons le contenu politique comme un contenu qui fait référence à un candidat, un parti politique, un membre du gouvernement ou du Parlement élu ou nommé, une élection, un référendum, une mesure de vote, une législation, un règlement, une directive ou une décision de justice. Les publicités qui font référence 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 quelque type que ce soit émanant de candidats, de partis politiques ou de représentants du gouvernement.

Par ailleurs, Twitter limite la publicité - et exige la certification des annonceurs - pour les publicités qui éduquent, sensibilisent et/ou appellent les gens à agir en rapport avec l'engagement civique, la croissance économique, la gestion de l'environnement ou les causes d'équité sociale. Nous avons pris cette décision en nous basant sur les deux convictions suivantes :

- La publicité ne doit pas être utilisée pour obtenir des résultats politiques, judiciaires, législatifs ou réglementaires ; cependant, la publicité axée sur une cause peut faciliter le débat public sur des sujets importants.
- La publicité qui utilise le micro-ciblage présente des défis entièrement nouveaux par rapport à la conversation citoyenne qui ne sont pas encore entièrement compris.

Les publicités dédiées à une cause sont également limitées en termes de ciblage et se limitent au ciblage géographique, par mot-clé et par intérêt. Aucun autre type de ciblage n'est autorisé, y compris les audiences personnalisées.

- Le ciblage géographique ne peut se faire qu'au niveau de l'État, de la province ou de la région et au-delà. Le ciblage au niveau des codes postaux n'est pas autorisé.
- Le ciblage par mot-clé et par intérêt ne peut inclure des termes associés à un contenu politique, à des annonceurs interdits ou à des tendances ou affiliations politiques (par exemple, "conservateur", "libéral", "élections politiques", etc.)

5.3. S'agissant de l'information du public à l'égard de ces contenus sponsorisés d'information se rattachant à un débat d'intérêt général, l'opérateur pourra préciser comment ils sont distingués sur le service et quelles informations fournit la plateforme à l'utilisateur à leur sujet (notamment : identité de la personne physique ; raison sociale, siège social et objet social des personnes morales).

Le centre de Transparence des publicités a été créé pour montrer toutes les annonces publicitaires en cours dans le cadre d'un effort plus large visant à accroître la transparence sur Twitter. Actuellement, les Tweets sponsorisés, les annonces vidéo et les annonces vidéo en continu apparaissent dans le Ads Transparency Center.

5.4. Dans la mesure où le dispositif différerait du dispositif de signalement général, le Conseil souhaite que l'opérateur lui indique si l'utilisateur peut signaler un contenu sponsorisé pour « fausse information » (ou équivalent), et en particulier un contenu sponsorisé d'information se rattachant à un débat d'intérêt général. Le cas échéant, il souhaite également connaître les modalités de signalement (faut-il avoir un compte pour signaler le contenu, quelle est la démarche pour signaler ces contenus...) et 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). L'opérateur pourra ici préciser le nombre de signalements reçus et traités relatifs à des contenus sponsorisés et, plus particulièrement, aux contenus sponsorisés 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 ». Enfin, il pourra détailler les moyens (technologiques, humains et financiers) mis en œuvre pour traiter ces signalements.



Lorsque les annonceurs sur Twitter choisissent de promouvoir leur contenu avec des publicités Twitter, leur compte et leur contenu sont soumis à un processus d'approbation. Le processus d'approbation est conçu pour soutenir la qualité et la sécurité de la plateforme de publicités Twitter. Twitter dispose d'un formulaire pour les signalements relatifs aux publicités Twitter. Bien que les annonceurs soient responsables du contenu de leur profil, des Tweets et de toutes les publicités qu'ils diffusent sur la plateforme Twitter Ads, nous examinons les violations signalées dans le cadre de la politique relative aux publicités Twitter, des règles et des conditions d'utilisation de Twitter et prenons les mesures qui s'imposent. Parmi ces politiques, Twitter interdit globalement les publicités qui contiennent des contenus inappropriés, notamment les contenus dénaturés, les contenus synthétiques, manipulés ou trompeurs 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 liés à des événements sensibles est également interdite. Twitter utilise la technologie et l'examen humain pour mettre en oeuvre ces politiques.

5.5. L'opérateur précisera les moyens, technologiques ou autres, permettant aux 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 sponsorisé ; etc.).

Il pourra également indiquer s'il a mis en place des mesures relatives à la sécurité des marques (brand safety) permettant de garantir aux annonceurs que leurs contenus sponsorisés ne seront pas associés à des comptes, pages ou contenus propageant de fausses informations.

Chez Twitter, nous prenons la sécurité des marques très au sérieux. Nous investissons continuellement dans ces domaines pour maintenir nos produits et nos processus à jour. Nous utilisons une combinaison de technologies (filtres algorithmiques visuels et textuels), de ressources humaines et d'outils de contrôle pour nous assurer que les messages des marques sont diffusés dans un environnement sécurisé. Nous investissons continuellement dans ces domaines afin de maintenir nos produits et nos processus à jour. Voici quelques-uns des domaines sur lesquels nous nous concentrons en matière de sécurité des marques :

- Des politiques strictes qui définissent les contenus à proximité desquels les publicités ne peuvent être distribuées :
Nos politiques en matière de publicité s'appliquent non seulement aux contenus que nous permettons aux annonceurs de promouvoir sur notre plateforme, mais aussi aux types de contenus qui les entourent.
- Une combinaison de différentes méthodes pour déterminer quel contenu peut, ou ne peut pas, faire l'objet de publicités en fonction de ces politiques:
Nous examinons activement le contenu et les signaux contextuels de nos produits publicitaires pour nous assurer que nous créons de bonnes expériences pour nos produits publicitaires. Cela est fait par :
 - L'apprentissage automatique : il est utilisé pour identifier et reconnaître les contenus potentiellement sensibles sur la base de filtres algorithmiques visuels et textuels
 - Revue humaine : pour identifier les contenus qui pourraient ne pas convenir à la publicité, identifier les contenus potentiellement sensibles (par exemple : catastrophe naturelle, décès ou soulèvements politiques), former l'apprentissage machine et examiner les contenus qui ont été signalés par les utilisateurs.
 - Auto-identification : les utilisateurs peuvent signaler leur propre compte pour marquer leurs médias comme contenant du matériel pouvant être sensible.
 - Outils de signalement des utilisateurs : tout utilisateur peut signaler une publicité comme n'étant pas conforme à nos politiques en matière de publicité sur Twitter. Les publicités signalées peuvent faire l'objet d'une revue humaine.
- Partenariats avec des éditeurs de contenu premium.
 - Twitter place la publicité sur des vidéos et des clips produits par des professionnels, dont plus de 200 éditeurs de qualité télévisuelle comme les grands réseaux de télévision et les ligues sportives. Les annonceurs peuvent consulter une liste blanche totalement transparente et exclure certains éditeurs. Nous demandons aux éditeurs d'accepter d'adhérer à nos politiques publicitaires. Les conséquences pour des



éditeurs qui enfreignent nos politiques sont importantes, y compris l'expulsion potentielle de dispositif.

- Nos publicités sont natives de la plate-forme et apparaissent dans des délais fixés en fonction des intérêts des utilisateurs.
 - Nous ne positionnons aucune publicité dans des formats superposés ou pop-up, et nos possibilités de vidéo pre-roll sont alignées sur des partenaires premium nommés qui partagent du contenu vidéo sûr pour les marques sur Twitter.

Les mesures que nous employons permettent de résoudre la majorité des problèmes potentiels. Lorsque des incidents rares et isolés se produisent, nous prenons des mesures immédiates pour apprendre ce qui s'est passé et nous agissons pour affiner nos systèmes et nos processus.

5.6. Enfin, l'opérateur indiquera au Conseil s'il dispose d'une base de données relative aux contenus sponsorisés diffusés sur le service, en général, et plus spécifiquement, aux contenus sponsorisés d'information se rattachant à un débat d'intérêt général ; il indiquera le cas échéant:

- **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, via une fonctionnalité sur le service ou permettant de les télécharger, peut y avoir accès.**

La transparence est une valeur fondamentale du travail que nous faisons chez Twitter. Notre engagement en faveur de la transparence est incarné par notre Centre de transparence des Publicités, un répertoire de toutes les publicités actuellement diffusées sur Twitter et des publicités de campagne politique lancées par des annonceurs certifiés pendant les élections européennes. Lancé à l'été 2018, puis étendu à l'UE, le Centre de Transparence des Publicités permet à tout un chacun - qu'il soit ou non utilisateur de Twitter - de consulter les Tweets sponsorisés qui ont été diffusés sur la plateforme, ainsi que les détails de la campagne en ligne tels que les informations de facturation, les dépenses publicitaires, les données d'impressions par Tweet et les données démographiques de ciblage des annonces diffusées.

Les informations peuvent être filtrées par compte, et non par pays pour le moment.

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

Les indications suivantes s'appliquent aux opérateurs qui auraient mis en œuvre des mesures relatives à l'éducation aux médias et à l'information au titre des mesures complémentaires prévues par l'article 11 de la loi du 22 décembre 2018.

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),**
- **les soutiens qu'il apporterait à des initiatives indépendantes émanant de journalistes et de chercheurs destinées à mieux comprendre et mesurer le phénomène de la désinformation.**

Nous pensons que les citoyens doivent être habilités à développer les compétences qui les aident à analyser de manière critique des documents pour en assurer la crédibilité et à poser les bonnes questions sur les informations qu'ils utilisent et partagent en ligne.



De notre côté, nous avons déjà pris des mesures pour accompagner les citoyens dans la compréhension des médias. Cela inclut notre ressource pédagogique dédiée, le Guide des Éducateurs sur Twitter, qui sensibilise les parents, les professeurs et les universitaires à la maîtrise des médias et de l'information. Nous avons également soutenu les efforts de nombreuses organisations à but non lucratif dans le monde entier qui travaillent à la promotion de l'éducation aux médias et à l'information.

Twitter et l'UNESCO travaillent ensemble depuis 2018 pour faire avancer la conversation sur la maîtrise des médias et de l'information. Nous avons lancé un emoji pour la Semaine mondiale de l'éducation aux médias et à l'information en 2018 et 2019, activé avec les hashtags #GlobalMILWeek, #ThinkBeforeSharing et #ThinkBeforeClicking et leurs traductions dans plus de 15 langues. Cela a permis d'accroître considérablement la visibilité et la participation à la conversation de la MIL Week au cours de ces deux années.

À l'occasion de la Semaine mondiale de l'éducation aux médias et à l'information 2019 (24-31 octobre), Twitter, en collaboration avec l'UNESCO, a produit la ressource "Enseigner & Apprendre avec Twitter". Il s'agit d'un manuel facile à lire pour les professeurs qui souhaitent exploiter les avantages de Twitter en tant qu'outil d'apprentissage en classe et à la maison, tout en recevant directement de l'UNESCO des conseils sur les bonnes pratiques en matière de maîtrise des médias et de l'information. Il illustre comment Twitter peut être utilisé pour l'apprentissage créatif des médias et de l'information et l'éducation des citoyens du monde, ainsi que la manière de faire face au cyber-harcèlement, de contrôler son empreinte numérique et de se protéger en ligne. Cette question se pose également dans le contexte des préoccupations relatives à la désinformation et de nos efforts pour mettre en avant des informations de qualité sur la plateforme.

Tout au long de l'année 2019 et au cours des années précédentes, Twitter a également organisé des formations et a soutenu de nombreuses campagnes et activités d'éducation aux médias dans l'UE et au-delà. À Paris, par exemple, nous avons organisé un événement d'éducation au numérique en partenariat avec le Centre de Liaison de l'Enseignement et des Médias de l'Information, au cours duquel 60 étudiants, un joueur de basket-ball olympique, un journaliste sportif et un enseignant ont été invités au bureau de Twitter France pour discuter et apprendre comment identifier la désinformation et le sexisme visant les femmes dans le sport en analysant des Tweets provenant de sources diverses. En France, Twitter est également partenaire de e-Enfance de longue date. Nous offrons une formation annuelle à leurs volontaires du service civique qui se rendent ensuite dans les écoles françaises afin d'éduquer les jeunes aux bonnes pratiques de sécurité et de sûreté en ligne.

Conformément à nos principes forts de transparence et dans le but d'améliorer la compréhension de l'influence étrangère et des campagnes d'information, nous avons publié les archives complètes et exhaustives des Tweets et des médias associés à d'éventuelles opérations d'information que nous avons trouvés sur notre service, y compris les 3 613 comptes qui, selon nous, étaient associés aux activités de l'*Internet Research Agency* sur Twitter depuis 2009. Nous avons mis ces données à disposition dans le but d'encourager la recherche ouverte et l'investigation de ces comportements de la part de chercheurs et d'universitaires du monde entier.

Avant la publication de ces données, Twitter a partagé des exemples d'interférences étrangères présumées dans les conversations politiques sur Twitter par l'*Internet Research Agency* et a fourni au public une notification immédiate s'il interagissait avec ces comptes. Nous avons lancé cette initiative unique pour améliorer la compréhension des universitaires et du public de ces campagnes coordonnées dans le monde entier, et pour permettre un contrôle indépendant et par des tiers de ces tactiques sur notre plateforme.

Nous reconnaissons également qu'en tant qu'entreprise privée, il existe des menaces que nous ne pouvons pas comprendre et auxquelles nous ne pouvons pas faire face seul. Nous devons continuer à travailler avec les élus, les entités gouvernementales, les pairs de l'industrie, les experts extérieurs et les autres parties prenantes afin que la communauté mondiale puisse comprendre le contexte complet dans lequel ces menaces se présentent.

Alors que nos enquêtes sur la manipulation de la plateforme à travers le monde se sont poursuivies, nous avons également ajouté plusieurs nouveaux ensembles de données tout en partageant également



des informations sur l'approche d'enquête interne de Twitter et sur la manière dont ces opérations complexes, parfois inter juridictionnelles, sont identifiées.

Les archives sont maintenant les plus importantes de ce type dans l'industrie. Nous sommes fiers du fait que des milliers de chercheurs utilisent ces ensembles de données qui contiennent plus de 30 millions de Tweets individuels et plus d'un téraoctet de médias. Grâce à nos archives, ces chercheurs ont mené leurs propres enquêtes et ont partagé leurs connaissances et leurs analyses indépendantes avec le monde entier.

6.2. L'opérateur décrira en quelques lignes chaque outil, action ou soutien, en s'attachant notamment à indiquer : sa nature, son objectif, en quoi il/elle consiste, sa date de lancement, son accessibilité et 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 (notamment en termes quantitatifs, ex. : nombre de personnes effectivement touchées), 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.

Voir réponse 6.1.

6.3. S'agissant spécifiquement des opérations qu'il promouvrait, il indiquera au Conseil s'il est à l'initiative de cette action de promotion ou s'il est sollicité par d'autres acteurs et comment elle s'opère.

N/A

6.4. Il renseignera le Conseil sur les relations qu'il entretient avec le milieu 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, le Conseil souhaite savoir si l'opérateur a conclu des partenariats avec le monde de la recherche pour mieux cerner certains comportements et phénomènes liés à la diffusion de fausses informations en ligne et 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 -, et si oui, sous quelles conditions cet accès est-il possible ?).

Twitter est une plateforme ouverte unique en son genre. Depuis 2006, les API de Twitter ont donné aux chercheurs et aux développeurs la possibilité d'accéder à ce qui se passe dans le monde. Twitter croit profondément à l'accès ouvert aux données pour étudier, analyser et contribuer à la conversation publique ; c'est pourquoi nous continuons de maintenir une API grand public. Les chercheurs utilisent les données de Twitter pour fournir un retour d'information précieux sur la manière dont les conversations et les interactions en ligne évoluent sur et hors de Twitter, ce qui est essentiel en période électorale. Nous continuons à fournir des moyens plus accessibles pour mettre les données et les informations à la disposition des chercheurs.

Les API de Twitter sont une source de données unique pour les universitaires. Elles sont utilisées chaque jour dans le monde entier dans un grand nombre de domaines, de la gestion des catastrophes aux sciences politiques. Chaque grande conférence de sciences sociales comporte probablement de nombreux articles basés entièrement ou en grande partie sur les données de Twitter. Notre service est la plus grande source de données publiques en temps réel de médias sociaux, que nous mettons à disposition via nos API.

Toutes nos données API sont publiques - aucune donnée privée d'utilisateur n'est incluse. Il s'agit donc de Tweets, de bios, de personnes suivies, de Tweets aimés, etc. Il n'y a pas d'adresses e-mail, de données IP, etc.



En tant que service ouvert unique, nos données sont utilisées chaque jour dans la recherche et notre centre de recherche est accessible au public: <https://developer.twitter.com/en/use-cases/academic-researchers>

Nous avons également alloué 1,9 million de dollars (une estimation de ce que nous avons gagné grâce aux annonces placées par Russia Today et Spoutnik, tous deux liés au gouvernement russe) pour financer des partenariats qui enrichissent la recherche universitaire sur la manière dont Twitter est utilisé pour l'engagement civique et les élections. Cette recherche explore l'utilisation de techniques de manipulation et de désinformation, en se concentrant dans un premier temps sur les élections, l'engagement civique et l'automatisation.

Nous travaillons avec un large éventail d'organisations pour nous aider à atteindre cet objectif, notamment:

- Annan Commission on Elections and Democracy in the Digital Age @kofiannanfdn
La Commission examine comment les pays, les entreprises et les citoyens peuvent exploiter les possibilités des nouvelles technologies pour servir nos démocraties tout en atténuant les risques.
- Atlantic Council's DFRL Lab @DFRLab
Le laboratoire de recherche en criminalistique numérique du Conseil atlantique est en train de créer un centre mondial d'analystes en criminalistique numérique qui suivra les événements en matière de gouvernance, de technologie, de sécurité.
- First Draft @firstdraftnews
First Draft est un projet du *Shorenstein Center on Media, Politics and Public Policy* de la *John F. Kennedy School of Government* de l'Université de Harvard qui vise à lutter contre la désinformation par le biais du travail de terrain, de la recherche et de l'éducation.
- EU DisinfoLab @DisinfoEU
L'EU DisinfoLab est une organisation non gouvernementale basée à Bruxelles qui a pour mission de lutter contre la désinformation grâce à une méthodologie innovante et un soutien scientifique à la communauté du contre-discours.
- City University @CityUniLondon
Le département de sociologie de la *City University* entreprend des recherches sur les aspects sociologiques des médias numériques en s'intéressant particulièrement aux effets croisés entre les réseaux sociaux en ligne et hors ligne.
- NDI @NDI
Le *National Democratic Institute* (NDI) dirige l'initiative d'organiser la coalition *Design 4 Democracy*, qui s'efforce de renforcer la démocratie à l'étranger à l'ère du numérique. Le NDI est soutenu dans cet effort par l'Institut républicain international (IRI), ainsi que par les organisations de nombreux membres du conseil consultatif de la coalition.
- Reporters Committee @rcfp
Le Reporters committee pour la liberté de la presse est une association à but non lucratif qui se consacre à l'aide aux journalistes depuis 1970.

7. Remarques générales

7.1. L'opérateur est invité à indiquer au Conseil le nombre global de contenus identifiés par lui comme comportant une fausse information sur son service en 2019, toutes sources de détection confondues (interne, utilisateurs, « fact-checkers »...), ainsi que le nombre d'entre eux ayant fait l'objet d'une mesure en conséquence, en déclinant en fonction du type de mesures adoptées.

Voir réponse au 1.5.

7.2. Le Conseil l'incite enfin à faire part de toute autre remarque qu'il souhaiterait porter à sa connaissance, en particulier sur les actions mises en place qui ne sont pas visées à l'article 11 de la loi du 22 décembre 2018 ni dans la recommandation du Conseil du 15 mai 2019 et dont la mise en place s'avérerait pertinente pour faire face aux enjeux évolutifs de la manipulation de l'information.



A cet égard, il pourra notamment porter à la connaissance du Conseil les actions menées de façon à identifier, par ses soins, les contenus susceptibles de constituer des fausses informations, et les mesures prises en conséquence (ou le cas échéant, indiquer si elles sont identiques à celles prises à l'issue du traitement des signalements d'utilisateurs).

N/A

7.3. L'opérateur est invité à faire part au Conseil des actions qu'il envisagerait de mener à l'avenir s'agissant de la mise en œuvre des différentes mesures et, plus globalement, de la lutte contre la désinformation.

N/A
