

# FACEBOOK

## 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

### Déclaration annuelle pour 2020

#### Facebook Ireland Limited

#### Questions liminaires : présentation du service

Conformément à l'Article 11 de la Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information (la « **Loi Fake News** »), Facebook Ireland Limited (« **Facebook Ireland** ») a rédigé les présentes réponses afin de fournir au Conseil supérieur de l'audiovisuel (« **CSA** ») une description des efforts entrepris pour lutter contre les fausses informations (la « **Déclaration 2020** »).

Cette Déclaration 2020 fournit des réponses détaillées au questionnaire du CSA. Les questions du CSA sont identifiées en caractères noirs ; les réponses aux questions du CSA sont identifiées en caractères bleus.

**Nom du service** : Les réponses portent sur les mesures mises en œuvre sur le service Facebook, sauf lorsqu'il est fait expressément référence au service Instagram.

**Nom et raison sociale de la société opératrice** : L'entité déclarante est Facebook Ireland (société constituée en vertu des lois de la république d'Irlande et ayant son centre d'activité à Dublin, et son siège social au 4 Grand Canal Square, Grand Canal Harbour Dublin 2). Facebook Ireland fournit le service Facebook et le service Instagram aux utilisateurs européens, en ce compris les utilisateurs français.

**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** : Anton'Maria Battesti, Directeur des affaires publiques de Facebook France SARL (« Facebook France »), a été mandaté pour être l'interlocuteur référent sur le territoire français pour l'application de la Loi Fake News, conformément à l'Article 13 de ladite loi.

# FACEBOOK

**Nombre de visiteurs uniques par mois de la plateforme en ligne, calculé sur la base de l'année précédant l'année d'exercice (2019) :** A la fin de l'année 2020 (T4), Facebook a rassemblé 2,8 milliards d'utilisateurs actifs mensuels, dont 40 millions d'utilisateurs actifs mensuels en France.

**Modalités de mise à disposition du service auprès du public français :** Les Conditions de service de Facebook sont accessibles [ici](#). Les Conditions d'utilisation d'Instagram sont accessibles [ici](#).

**Disponibilité :**

- **sur internet (URL du site internet) :**
- **via des applications (systèmes d'exploitation concernés) :**
- **via des assistants vocaux :**

Les services Facebook et Instagram sont accessibles aux adresses [www.facebook.com](http://www.facebook.com) et [www.instagram.com](http://www.instagram.com), ainsi que via leurs applications mobiles (disponible sur iOS (Apple Store) et Android (Play Store)).

**Langue du service sur le territoire français (par défaut) :** Les services sont disponibles en 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, les services disponibles sur le territoire français ne constituent pas une version française des services. La base des services est la même pour tous les utilisateurs, mais certaines fonctionnalités peuvent différer d'un pays à l'autre (en raison d'une réglementation spécifique par exemple).

**Description du service :**

- **Objet :** Facebook a pour mission de donner à tous la possibilité de créer une communauté et de rapprocher les personnes du monde entier. Celle d'Instagram est d'offrir à ses utilisateurs un service leur permettant de se rapprocher des gens et des choses qu'ils aiment.
- **Population(s) ciblée(s) (tranches d'âges, communautés d'intérêt...):** Les services Facebook et Instagram sont accessibles en France à toute personne âgée de 13 ans et plus, sous réserve de restrictions légales décrites dans les Conditions de service (par exemple si des utilisateurs ont une interdiction légale d'utiliser nos services).

# FACEBOOK

- **Types de contenus qui peuvent être publiés et consommés sur le service (vidéos, textes, sons, images, jeux vidéo...):** Les types de contenus qui peuvent être publiés sur les services Facebook et Instagram sont décrits dans les Conditions de service et les Standards de la communauté (Facebook) qui sont accessibles [ici](#) et les Règles de la communauté (Instagram) qui sont accessibles [ici](#).
- **Thématique(s) :** Voir description de l'objet des services.

**Modèle économique du service (types d'activités rémunératrices pour l'opérateur tirées de l'exploitation du service en France) :** Les utilisateurs ne paient pas pour l'utilisation des services Facebook et Instagram. Au contraire, les entreprises et les organisations nous paient pour diffuser aux utilisateurs des publicités de leurs produits et services. Le modèle économique de Facebook est décrit dans ses Conditions de service, accessibles [ici](#). Les Conditions d'utilisation d'Instagram sont accessibles [là](#).

**Effectifs employés aux activités du service dans le monde et en France :** L'effectif mondial était de 58 604 personnes au 31 décembre 2020. En France, l'effectif est supérieur à 270 personnes.

## **Comptes de résultat pour l'activité du service en France en 2020 :**

Les revenus de Facebook France en 2019 sont accessibles au public sur le site Infogreffe.fr. Les indicateurs pour 2019 seront disponibles dans les semaines à venir sur ce même site.

Les mesures et indicateurs cités dans cette Déclaration 2020 concernent l'année 2020 (puisque les réponses aux questions du CSA doivent couvrir les activités de l'année précédente), sauf lorsque l'année 2021 est expressément citée.

Nous avons inclus autant d'informations que possible concernant la France (selon les demandes du CSA). Toutefois, dans la mesure où nos politiques sont globales, les indicateurs cités (par exemple les statistiques) s'entendent au niveau mondial.

## **Questions générales**

### **Définition des fausses informations**

# FACEBOOK

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.

Comme mentionné à la Question 1.4, les experts, gouvernements et plateformes ne peuvent s'accorder sur la définition de "désinformation", et il n'existe aucun moyen de tracer une ligne nette entre information fausse, satire et simple opinion.

Chez Facebook, nous définissons la désinformation comme tout contenu faux ou trompeur. Cette définition peut couvrir différents types de contenus, comme cela est expliqué ci-dessous (voir section ci-dessous sur notre stratégie de lutte contre les fausses informations). Cette définition est utilisée à l'échelle mondiale, pour l'ensemble de nos services.

## ***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.

La lutte contre les fausses informations et la désinformation est au cœur de nos préoccupations.

Nous avons poursuivi nos efforts pour assurer la sécurité de nos utilisateurs et renforcer notre lutte contre les abus. Nos méthodes pour ce faire incluent le blocage et la suppression des faux comptes, l'identification et la suppression des comptes des acteurs mal intentionnés, la réduction de la distribution des fausses informations et de la désinformation, ainsi que davantage de contexte concernant ce que les utilisateurs voient sur nos services et plus spécifiquement la transparence accrue concernant les publicités d'intérêt général, en ce compris de nature politique.

Nous partageons le souhait de nos utilisateurs de voir du contenu de qualité sur nos services. C'est pourquoi notre stratégie de lutte contre la désinformation consiste en trois phases : supprimer, réduire, informer.

- Nous supprimons le contenu qui viole nos Standards de la communauté (Facebook) ou nos Règles de la communauté (Instagram), lesquels interdisent notamment les faux comptes, les *spams*, les vidéos manipulées et trompeuses, la pression sur les électeurs pour ne pas voter et la désinformation contribuant à causer un dommage physique.

# FACEBOOK

- Nous réduisons la distribution du contenu qui ne viole pas directement nos Standards de la communauté ou nos Règles de la communauté, mais qui est toutefois préjudiciable à l'authenticité de la plateforme, en réduisant la diffusion d'un contenu qui est signalé par nos partenaires de vérification indépendants dans le fil d'actualité Facebook ou Instagram Explorer.
- Nous informons nos utilisateurs en leur donnant davantage de contexte sur les contenus qu'ils souhaitent lire, croire et partager, en leur offrant davantage de transparence et de contrôle, ainsi qu'en promouvant l'éducation aux médias.

A ces fins, nous avons, notamment :

- augmenté le nombre de nos partenaires tiers de vérification indépendants et offert à nos utilisateurs davantage de contrôle sur leurs informations et les contenus qu'ils voient dans leur fil d'actualité.
- apporté plus de transparence sur nos politiques et nos opérations ; à cet égard, nous publions désormais notre [Rapport d'Application des Standards de la communauté \(Community Standards Enforcement Report\)](#) tous les trimestres afin de montrer comment nous appliquons nos propres Standards de la communauté ; ce rapport, ainsi que la méthodologie qui y est décrite, sont vérifiés de façon indépendante, via le groupe responsable de la transparence des données (*Data Transparency Advisory Group (DTAG)*) dont le rôle consiste à évaluer si les indicateurs que nous partageons avec le Rapport d'Application des Standards de la communauté sont exploitables et précis.
- travaillé en permanence à l'amélioration de nos algorithmes, nous permettant de détecter et de supprimer plus facilement tout comportement contrevenant à nos Standards de la communauté ou Règles de la communauté, ce qui nous permet de supprimer des millions de faux comptes chaque jour, les empêchant ainsi de lancer des campagnes d'informations coordonnées, comme celles qui sont souvent utilisées pour influencer des élections.
- renforcé notre collaboration avec les gouvernements, les experts externes et les autres sociétés.

Les utilisateurs peuvent trouver des informations relatives à la lutte contre les fausses informations et la désinformation sur nos services dans nos politiques. La politique de Facebook

# FACEBOOK

en matière de fausses informations et de désinformation se trouve dans nos Standards de la communauté accessibles [ici](#) et dans nos pages dédiées [ici](#) (Facebook) et [ici](#) (Instagram).

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.

- Depuis mars 2020, nous avons supprimé plus de 12 millions de contenus sur Facebook et Instagram contenant de la désinformation liée à la COVID-19 pouvant entraîner des dommages physiques imminents, tels que des contenus relatifs à de fausses mesures préventives ou à des remèdes artificiels.
- En janvier 2021, nous avons supprimé plus de 290 000 contenus sur Facebook et Instagram à l'échelle mondiale, dont plus de 13 000 contenus au sein de l'Union Européenne contenant de la désinformation liée à la COVID-19 pouvant entraîner des dommages physiques imminents, tels que des contenus relatifs à de fausses mesures préventives ou à des remèdes artificiels.
- Entre mars et octobre 2020, nous avons labellisé environ 167 millions de contenus sur Facebook, sur la base d'articles écrits par nos partenaires de vérification indépendants. Ainsi, ces contenus ont disposé d'une étiquette avertissant qu'il s'agit d'une fausse information sur la COVID-19.
- Au cours du seul mois d'avril 2020, nous avons labellisé environ 50 millions de contenus liés à la COVID-19 sur Facebook, sur la base d'environ 7 500 articles écrits par nos partenaires de vérification indépendants. Lorsque les gens voyaient ces avertissements, 95 % du temps ils ne visualisaient pas le contenu d'origine.

En règle générale, l'opérateur a-t-il accru et 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

# FACEBOOK

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.

1. Nous avons adapté nos politiques et apporté divers changements en réponse à la pandémie de la COVID-19.

Depuis que la COVID-19 a été déclarée urgence sanitaire mondiale en janvier 2020, nous nous sommes efforcés de partager avec nos utilisateurs des informations de références fournies par des experts de la santé et d'empêcher la diffusion, sur nos applications, de fausses informations néfastes concernant la COVID-19.

Depuis janvier 2020, nous travaillons avec (i) les ministères nationaux de la santé et des organisations mondiales telles que l'Organisation mondiale de la santé (OMS) et l'UNICEF pour les aider à diffuser en temps utile des informations de référence sur la COVID-19, et (ii) les principales institutions de l'Union Européenne et le Centre européen de prévention et de contrôle des maladies (ECDC). Nous avons travaillé pour partager avec les personnes des informations de référence en matière de santé, de différentes manières, par exemple en redirigeant les utilisateurs qui recherchent des informations sur la COVID-19 sur Facebook ou Instagram vers les autorités sanitaires, en fournissant un module d'information qui relaie les informations officielles dans le fil d'actualité Facebook pour que tous nos utilisateurs puissent les voir, et en lançant un « *Centre d'information sur le coronavirus (COVID-19)* » sur Facebook qui fait office de source de référence pour tous les utilisateurs souhaitant se tenir informés des dernières nouvelles sur le sujet, que ce soit des informations provenant des autorités sanitaires, des ressources ou des conseils pour rester en bonne santé et en sécurité (voir le Centre d'information sur le coronavirus [ici](#)). Ainsi, nous avons connecté plus de 2 milliards de personnes aux ressources des autorités sanitaires via notre Centre d'information sur le coronavirus et via des pop-ups sur Facebook et Instagram, avec plus de 600 millions de personnes cliquant dessus pour en savoir plus.

Pour plus d'informations et de chiffres, nous vous invitons à consulter les réponses de Facebook à la Commission Européenne sur la désinformation liée à la COVID-19 qui fournissent des détails sur les changements réalisés par Facebook sur ses politiques et les actions entreprises pour lutter contre les fausses informations: [Reports on January actions – Fighting COVID-19 Disinformation Monitoring Programme | Shaping Europe's digital future \(europa.eu\)](#). Voir aussi les Pages d'aide Facebook sur les mises à jour des politiques et des protections relatives à la COVID-19 [ici](#).

# FACEBOOK

## 2. Nous avons également étendu notre programme de vérification des faits

Notre programme de vérification des faits est un élément clé de notre stratégie à volets multiples visant à réduire la propagation de fausses informations sur nos plateformes. C'est pourquoi, depuis janvier 2020, nous avons pris un certain nombre de mesures supplémentaires pour aider nos partenaires de vérification indépendants à dénoncer les fausses informations sur la COVID-19.

En avril 2020, nous nous sommes associés à l'International Fact-Checking Network (IFCN) de Poynter pour lancer un programme de subvention de 1 milliards de dollars américains afin d'aider nos partenaires de vérification indépendants dans leurs travaux sur la COVID-19. Depuis le lancement de ce programme, nous avons accordé des subventions à 21 organisations de vérification des faits dans le monde, dont cinq au sein de l'Union Européenne: 20 minutes (France) ; Ellinika Hoaxes (Grèce), 15min.lt (Lituanie), Newtral (Espagne) et Pagella Politica (Italie).

Nous continuons à faire évoluer notre programme de vérification des faits pour refléter les contenus que nos partenaires voient sur nos plateformes. Notre approche consiste à prendre des mesures efficaces contre les fausses informations, tout en tenant compte du contenu qui ne serait pas tout à fait faux, mais qui devrait bénéficier d'un contexte supplémentaire. À cette fin, nous avons annoncé deux nouvelles options d'évaluation ("Modifié" et "Contexte manquant") pour donner à nos partenaires de vérification indépendants plus de latitude pour mieux refléter leurs recherches, et pour s'assurer que les personnes qui consultent ces publications disposent d'informations plus précises pour décider des contenus qu'elles doivent lire, croire et partager. Il est renvoyé aux Questions 3.3 à 3.7 à cet égard.

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.



# FACEBOOK

Nous nous efforçons constamment d'identifier et de mettre fin aux campagnes coordonnées qui cherchent à manipuler le débat public sur nos services. Ainsi, nous partageons nos conclusions sur les comportements inauthentiques coordonnés (CIB) que nous détectons et les supprimons de nos services.

Dans le cadre de nos enquêtes sur les [opérations d'influence](#), nous nous concentrons sur le comportement, et non sur le contenu, car c'est le meilleur moyen de mettre fin aux abus ; c'est pourquoi notre travail d'enquête et de mise en œuvre est indépendant du lieu et du contenu. Nous définissons le comportement inauthentique coordonné comme des efforts coordonnés pour manipuler le débat public dans un but stratégique, où les faux comptes sont au cœur de l'opération. Notre approche du CIB et plus largement des opérations d'influence ("IO") est fondée sur le comportement et les acteurs.

Cela signifie que nous recherchons des comportements violant nos règles spécifiques manifestés par des acteurs malveillants, plutôt que des contenus violant nos règles, c'est-à-dire certains de nos Standards de la communauté, telles que la désinformation (voir section ci-dessus) et les discours de haine. Par conséquent, lorsque des réseaux de CIB sont retirés, c'est en raison de leur comportement et non du contenu qu'ils ont publié. Pour un aperçu complet de la manière dont nous réagissons face aux comportements inauthentiques, voir la Question 2.9 et [ici](#).

Nous retirons les réseaux d'opérations d'influence au sein de l'Union Européenne et partout dans le monde, qui violent notre politique en matière de comportement inauthentique. Ces démarches ne sont pas spécifiques à une liste de 20 types de contenus et sont présentées dans notre rapport mensuel sur les comportements inauthentiques coordonnés, qui comprend le nombre de comptes, de pages, de groupes et le montant total des dépenses publicitaires ainsi que des exemples de contenus postés par ces réseaux.

Nous partageons également ces informations avec des chercheurs indépendants tiers, tels que le [DFRLab](#) de l'Atlantic Council, l'Internet Observatory de Stanford et [Graphika](#), qui publient leurs propres rapports, ainsi qu'avec les services répressifs, les organismes gouvernementales (y compris le Service européen d'action extérieure et la Commission européenne), les responsables politiques et les partenaires industriels concernés, afin qu'ils puissent également prendre les mesures appropriées.

Tous les rapports sur les réseaux que nous avons démantelés pour comportement inauthentique coordonné peuvent être consultés [ici](#).

# FACEBOOK

## ***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.

Facebook s'engage à trouver le bon équilibre et à protéger les droits fondamentaux de ses utilisateurs. En octobre 2019, nous avons mis à jour les valeurs qui fondent nos Standards de la communauté pour faire spécifiquement référence aux principes des droits de l'homme. Au début de l'année dernière, le Global Network Initiative (GNI) a rendu son évaluation biennale de Facebook, qui a déterminé que l'entreprise fait des efforts de bonne foi pour mettre en œuvre les principes du GNI avec une amélioration au fil du temps. L'évaluateur a indiqué que Facebook avait « [renforcé son examen systématique de la vie privée et de la liberté d'expression](#) ». Le 16 mars 2021, nous avons également lancé notre [politique en matière de droits de l'homme](#) qui définit les standards que nous nous efforcerons de respecter et d'appliquer à nos produits.

### 1. Fausses informations et liberté d'expression

Facebook est une plateforme pour les voix du monde entier.

Nous modérons le contenu partagé par des milliards de personnes de façon à faire toujours prévaloir la liberté d'expression, autant que possible. Mais il y a des [exceptions](#). Par exemple, nous reconnaissons le droit de dire quelque chose qui n'est pas vrai. En effet, le fait qu'une publication Facebook soit exacte ou non n'est pas en soi une raison de la bloquer. Mais réduire la propagation de fausses informations sur Facebook est également une responsabilité que nous prenons très au sérieux. Nous reconnaissons que cette mission est délicate et sensible. Notre but est d'aider les gens à rester informés sans réprimer les débats publics, qui peuvent être constructifs. La frontière est également ténue entre fausses informations, satire ou opinion.

Les droits de l'homme accordent le même droit d'expression à ceux qui affirment que la terre est plate qu'à ceux qui affirment qu'elle est ronde – et Facebook aussi. Parfois, un faux contenu peut enfreindre d'autres de nos règles, mais ce n'est pas toujours le cas. Plutôt que de bloquer le contenu car il est faux, nous préférons donc en [réduire la diffusion dans le fil d'actualité, lorsque ce contenu est jugé faux par nos partenaires](#) de vérification indépendants. Nous dirigeons également le public vers les articles écrits par nos partenaires de vérification indépendants.

# FACEBOOK

À cet égard, nous adaptons nos politiques lorsque cela est nécessaire et travaillons avec des organisations indépendantes qui signalent les fausses informations et les rumeurs qui peuvent contribuer à des conséquences violentes et réelles. Nous évaluons ensuite leur signalement et supprimons la publication si elle viole nos politiques. (Pour plus de détails, voir l'article suivant : <https://about.fb.com/news/2018/08/hard-questions-free-expression/>).

Comme le dit Mark Zuckerberg, « nous veillons à ce que les canulars ne deviennent pas viraux. Nous mettons particulièrement l'accent sur les fausses informations qui pourraient entraîner des dommages physiques imminents, comme des conseils de santé trompeurs disant que si vous avez un AVC, pas besoin d'aller à l'hôpital (...) ». Compte tenu des sensibilités culturelles et des points de vue divergents sur ce que les gens considèrent comme du contenu dangereux, nous travaillons avec des partenaires de vérification indépendants pour arrêter la propagation des canulars viraux. (voir discours « Standing For Voice and Free Expression », accessible à l'adresse <https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression/>).

Notre approche des fausses informations est également guidée par le principe selon lequel nous pensons qu'il ne relève pas du rôle d'une société de technologie privée de se faire arbitre de la vérité, c'est pourquoi notre approche repose sur des partenariats avec des experts externes et des organisations tierces qui nous aident à distinguer le vrai du faux.

Nous considérons que la réduction de la diffusion de fausses informations - plutôt que leur suppression pure et simple (à moins que la fausse information en question ne viole directement nos Standards de la communauté ou nos Règles de la communauté parce qu'elle entraîne ou peut entraîner un préjudice physique) - permet un juste équilibre entre, d'une part, un lieu d'expression mis à la disposition des utilisateurs et, d'autre part, une communauté sûre et authentique.

## 2. Contenu qui viole nos Standards de la communauté

En outre et comme mentionné précédemment, nous supprimons le contenu qui viole nos Standards de la communauté (Facebook) ou nos Règles de la communauté (Instagram).

À cet égard, nos Standards de la communauté font expressément référence à l'équilibre, que nous établissons, entre la liberté d'expression/le droit à l'information et la lutte contre les fausses informations :

# FACEBOOK

- Dans sa section introductive « *Réitérer notre engagement pour l'expression* », en affirmant que :

*« L'objectif de nos Standards de la communauté a toujours été de créer un espace d'expression et de donner une voix aux personnes. Cela n'a pas changé et ne changera pas. Pour créer des communautés et rapprocher les gens, il faut pouvoir partager des points de vue, des expériences, des idées et des informations variés. Nous souhaitons que les personnes puissent s'exprimer ouvertement sur les sujets qui comptent pour elles, même si d'autres peuvent marquer leur désaccord ou y trouveraient à redire. Dans certains cas, nous autorisons - à des fins de sensibilisation - la publication de contenu qui va à l'encontre de nos Standards de la communauté s'il est pertinent et d'intérêt public. Nous ne prenons cette décision qu'après avoir évalué l'impact positif pour l'intérêt public par rapport à la nuisance potentielle, et nous nous appuyons sur les principes universels des droits de l'Homme. Nous prenons ainsi en considération l'intérêt médiatique du contenu publié par tout individu, y compris les agences de presse et les utilisateurs particuliers. Par exemple, nous avons autorisé des contenus illustrant graphiquement la guerre ou les conséquences de la guerre lorsqu'ils s'avèrent importants pour la réflexion publique.*

*Notre engagement en faveur de la liberté d'expression est primordial, mais nous sommes conscients qu'Internet est une source de formes d'abus nouvelles et croissantes. Lorsque nous limitons la liberté d'expression, nous le faisons pour le respect de l'une ou de plusieurs des valeurs suivantes :*

- **Authenticité** : *Nous voulons nous assurer que le contenu que les personnes voient sur Facebook est authentique. Nous sommes convaincus que l'authenticité crée un meilleur environnement pour le partage, et nous ne voulons donc pas que les personnes qui utilisent Facebook mentent sur leur identité ou leurs activités.*
- **Sécurité** : *Nous nous engageons à faire de Facebook un endroit sûr. Les formes d'expression qui menacent les autres peuvent les intimider, les exclure ou les réduire au silence, et ne sont pas autorisées sur Facebook.*
- **Confidentialité** : *Nous nous engageons à protéger votre vie privée et vos informations. La confidentialité donne aux personnes la liberté d'être elles-mêmes, et de choisir quand et comment elles partagent des informations sur Facebook. Cela facilite la création de liens.*

# FACEBOOK

- **Dignité** : Nous pensons que chacun mérite les mêmes droits et la même dignité. Nous attendons de chaque personne qu'elle respecte la dignité d'autrui, et qu'elle ne harcèle pas et ne rabaisse pas les autres.

Nos Standards de la communauté s'appliquent à tout le monde, partout dans le monde, et à tous les types de contenu. Ils ont été conçus pour être exhaustifs. Par exemple, du contenu qui ne serait pas considéré comme haineux pourrait malgré tout être supprimé parce qu'il enfreint une autre politique. Nous savons que les mots ont différentes significations et n'affectent pas les personnes de la même manière selon leur communauté locale, leur langue et leurs origines. Nous mettons tout en œuvre pour prendre en compte ces nuances tout en appliquant nos politiques de manière juste et cohérente pour chaque utilisateur et sa liberté d'expression. » Voir également <https://about.fb.com/news/2019/09/updating-the-values-that-inform-our-community-standards/>

## 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 continuons d'intensifier nos efforts pour supprimer les fausses déclarations sur Facebook et Instagram, notamment sur la COVID-19, les vaccins contre la COVID-19 et les vaccins en général dans le contexte de la pandémie. Depuis décembre 2020, nous avons éliminé les fausses déclarations signalées par des experts en santé publique concernant les vaccins contre la COVID 19.

À la suite de consultations avec des organismes de santé de premier plan, y compris l'OMS, nous élargissons la liste des fausses déclarations que nous supprimerons pour inclure d'autres déclarations démenties concernant la COVID-19 et les vaccins. Nous rendons publique notre liste complète des déclarations (voir [ici](#)), et nous [interdisons déjà ces](#) déclarations dans les publicités. Ces nouvelles politiques nous aideront à continuer de lutter contre les fausses informations sur la COVID-19 et les vaccins. Nous appliquons cette politique en mettant particulièrement l'accent sur les pages, les groupes et les comptes qui enfreignent ces règles. Les groupes, pages et comptes sur Facebook et sur Instagram qui partagent à plusieurs reprises ces déclarations démenties peuvent être complètement supprimés. Les déclarations concernant la COVID-19 ou les vaccins qui n'enfreignent pas ces politiques seront toujours admissibles à un examen par nos

# FACEBOOK

partenaires de vérification indépendants, et si elles sont jugées fausses, elles seront labellisées et rétrogradées.

En outre, nous continuons d'améliorer les résultats de recherche sur nos services. Lorsque les gens recherchent du contenu lié au vaccin ou à la COVID-19 sur Facebook, nous promovons les résultats pertinents et de référence et fournissons des sources tierces pour donner accès aux gens à des informations d'experts sur les vaccins. Sur Instagram, en plus d'afficher des résultats faisant autorité dans la recherche, nous avons commencé à rendre plus difficile la recherche de comptes qui découragent les gens à se faire vacciner. Pour plus d'informations à ce sujet, voir : <https://about.fb.com/news/2020/04/covid-19-misinfo-update/#removing-more-false-claims> et les Pages d'aide Facebook <https://www.facebook.com/help/230764881494641/>

Nous lançons également un outil dans le Centre d'information sur le coronavirus qui informe les utilisateurs des dates et lieux où se faire vacciner et leur donne un lien pour prendre un rendez-vous. Et pour continuer à aider les gens à trouver des informations de référence sur la COVID-19, nous avons lancé le Centre d'information sur le coronavirus sur Instagram.

Nous adoptons une approche similaire en ce qui concerne la crise climatique et avons lancé un « [Centre d'information sur le climat](#) » sur Facebook pour donner accès aux gens à des actualités scientifiques, des informations accessibles et des ressources exploitables provenant des principales organisations mondiales de lutte contre le changement climatique. Le Centre d'information sur le climat comporte des analyses détaillées qui vont au-delà de simples faits, ainsi que des façons de s'impliquer. Il contient également des informations pertinentes sur l'endroit où vit le lecteur. Nous avons créé le Centre d'information sur le climat et les informations qu'il contient sur le modèle du Centre d'information sur le coronavirus. Le Centre d'information sur le climat est déjà disponible en France. En plus de développer ce Centre, nous l'améliorons constamment. Nous avons ajouté une section dont l'objet est de dénoncer les mythes climatiques communs, selon lesquels, par exemple, une quantité excessive de dioxyde de carbone dans l'atmosphère nuirait à la vie végétale de la terre et les populations d'ours polaires seraient en déclin en raison du réchauffement climatique. Pour signaler les mythes à l'aide de faits actuels et spécifiques, nous avons fait appel à des experts en communication climatique de l'Université George Mason, du Yale Program on Climate Change Communication et de l'Université de Cambridge. Pour plus d'informations, voir : <https://about.fb.com/news/2021/02/connecting-people-with-credible-climate-change-information/>

# FACEBOOK

Enfin, notre réseau de partenaires de vérification indépendant peut évaluer et évalue les contenus liés aux sciences du climat.

## 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.

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.

Nous avons mis en place un dispositif de signalement directement accessible et visible qui permet aux utilisateurs de signaler des fausses informations.

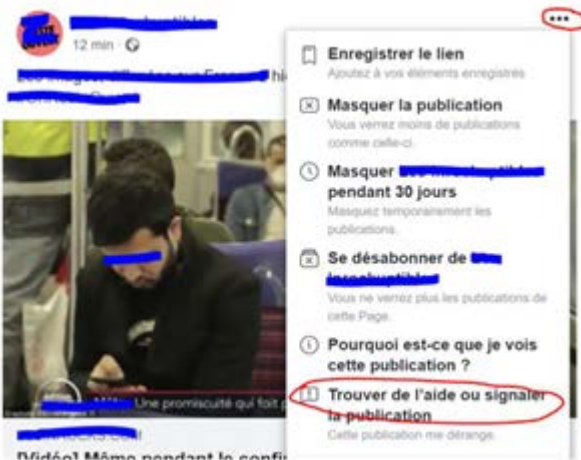
Le signalement de fausses informations dans une publication, dans un commentaire ou sur une page oblige maintenant l'utilisateur à préciser si les fausses informations sont liées à des questions de santé, de politique, d'enjeu social ou autre chose avant de finaliser le rapport. En effet, il nous a souvent été reproché de ne pas offrir aux utilisateurs la possibilité de préciser si leur signalement est dû à une fausse information pour des raisons de santé. Nous avons donc jugé nécessaire d'ajouter une sous-catégorie pour les fausses informations telles que « fausses informations - santé » et de les hiérarchiser.

Pour signaler une publication, un commentaire, un groupe ou une page comme une fausse information sur Facebook, l'utilisateur doit :

1. Cliquer sur  à côté de la publication/commentaire/groupe/page qu'il/elle souhaite signaler comme fausse.
2. Sélectionner « *Trouver de l'aide ou signaler la publication/commentaire/groupe/page* ».

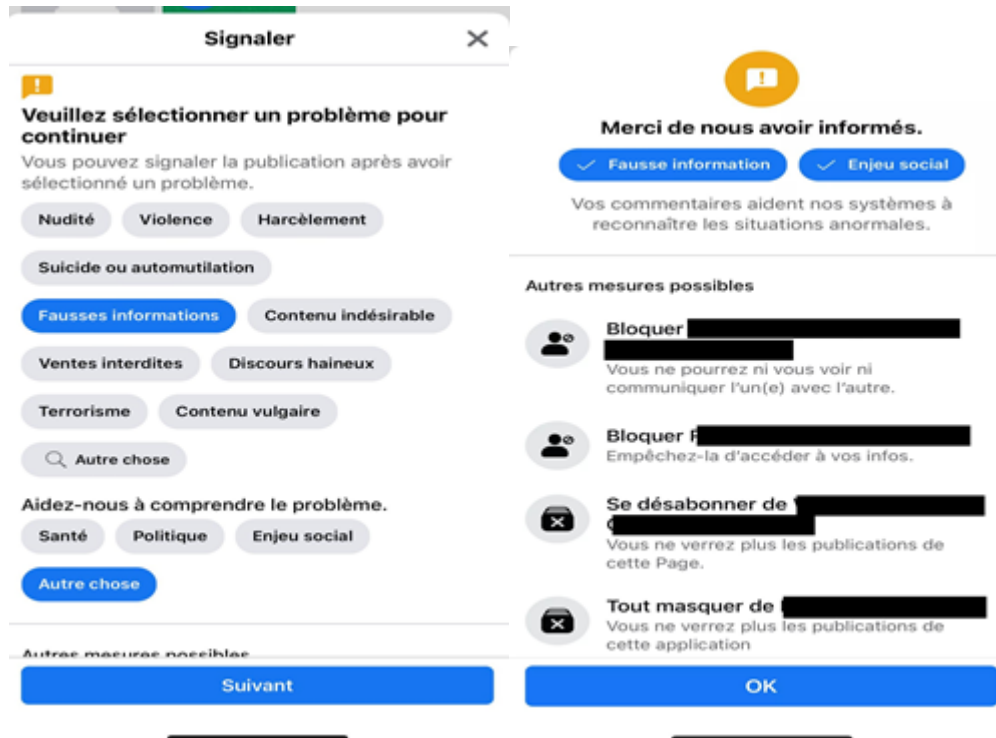
# FACEBOOK

3. Cliquer sur « *Fausse* informations ».
4. Cliquez sur « *Santé* », « *Politique* », « *Enjeu social* » ou « *Autre chose* ».
5. Cliquez sur « *Suivant* ».
6. Ci-dessous des captures d'écran montrent l'expérience utilisateur :





# FACEBOOK



Pour signaler un courrier *spam*, la procédure est identique à celle décrite ci-avant. Au lieu de sélectionner « *Fausse information* » à l'étape 3, l'utilisateur sélectionne alors « *Contenu indésirable* ».

Pour signaler un faux compte, l'utilisateur doit :

1. Consulter le profil du faux compte.

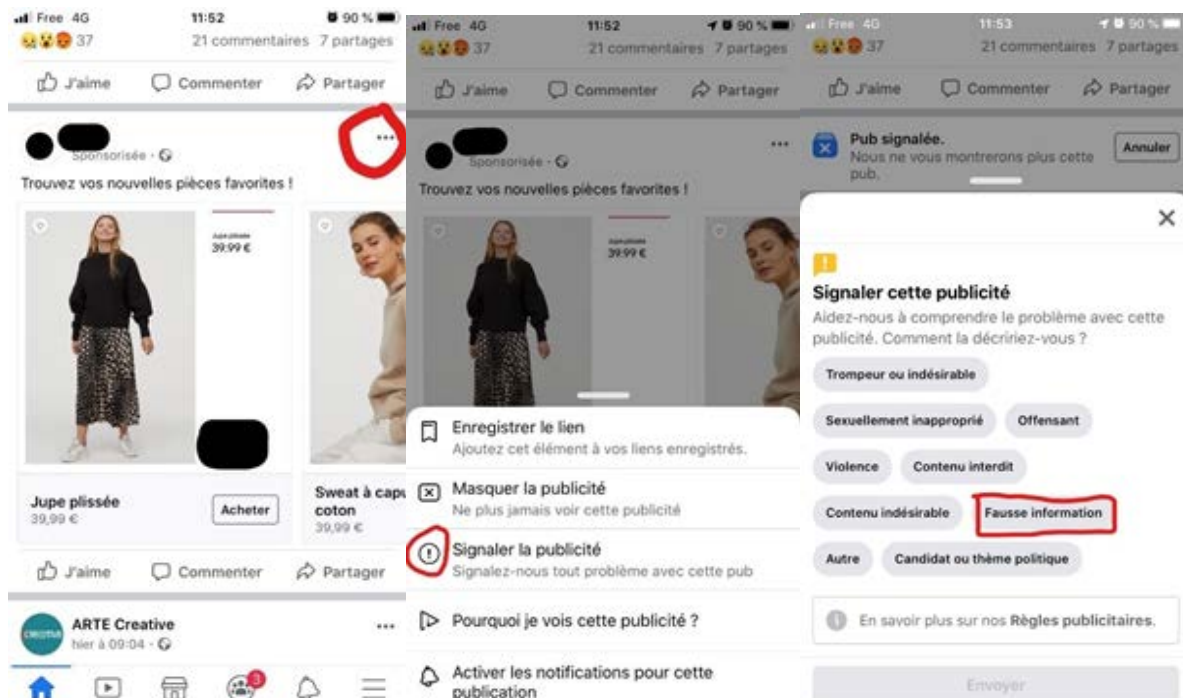
Si l'utilisateur ne peut le trouver, il/elle peut tenter de rechercher le nom utilisé dans le profil ou demander à ses amis s'ils peuvent lui envoyer un lien vers celui-ci.

2. Cliquer sur **...** sur la photo de couverture et sélectionner « *Trouver de l'aide ou signaler un profil* ».
3. Suivre les instructions qui s'affichent pour signaler le faux compte.

# FACEBOOK

Pour signaler une publicité qui est, selon l'utilisateur, fausse, trompeuse ou indésirable (*spam*), l'utilisateur doit :

1. Cliquer sur **⋮** en haut à droite de la publicité qu'il souhaite signaler.
2. Cliquer sur « *Signaler la publicité* » et sélectionner « *fausse information* », « *trompeur ou indésirable* » ou « *contenu indésirable* »).



# FACEBOOK



Notre Page d'aide intitulée « *Signalement de quelque chose* » explique en détail comment signaler différents types de contenus, y compris, par exemple, les profils, pages, groupes, publications, publicités, etc.


Exemple des explications données par Facebook sur ses Pages d'aide :

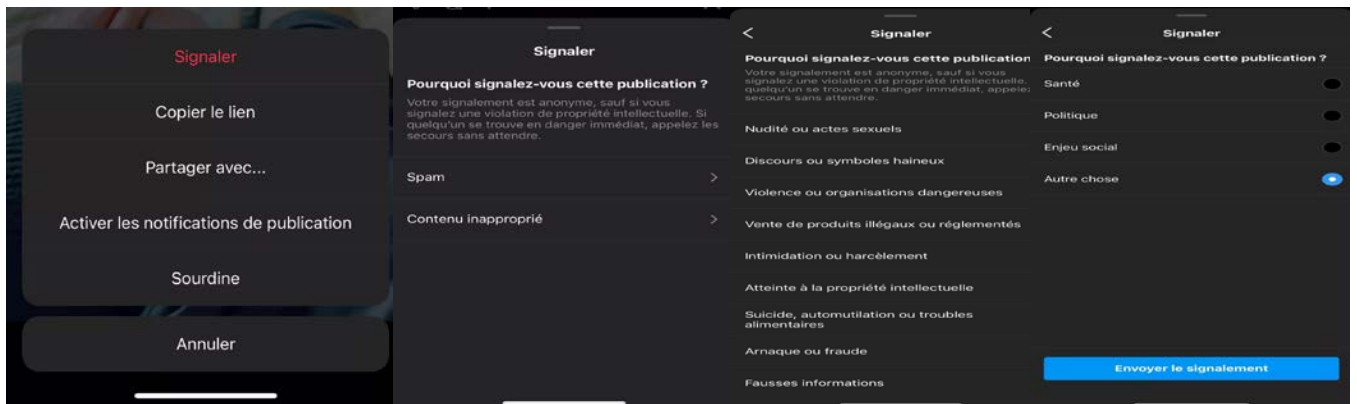
Pour signaler une fausse information sur une publication :

1. Cliquez sur \*\*\* à côté de la publication que vous souhaitez signaler comme contenant de fausses informations.
2. Cliquez sur **Trouver de l'aide ou signaler la publication**.
3. Cliquez sur **Fausse information**, puis sur **Suivant**.
4. Cliquez sur **Terminé**.

De même, pour signaler un article comme une fausse information sur Instagram, l'utilisateur peut :

# FACEBOOK

1. Cliquer sur  à côté de la publication qu'il souhaite marquer comme fausse.
2. Cliquer sur « Signaler ».
3. Cliquer sur « Contenu inapproprié ».
4. Cliquer sur « Fausse information ».
5. Cliquer sur « Santé », « Politique », « Enjeu social » ou « Autre chose ».
6. Cliquer sur « Envoyer le signalement ».



Une fois le contenu signalé, Facebook permet aux utilisateurs de suivre le traitement de leur signalement :

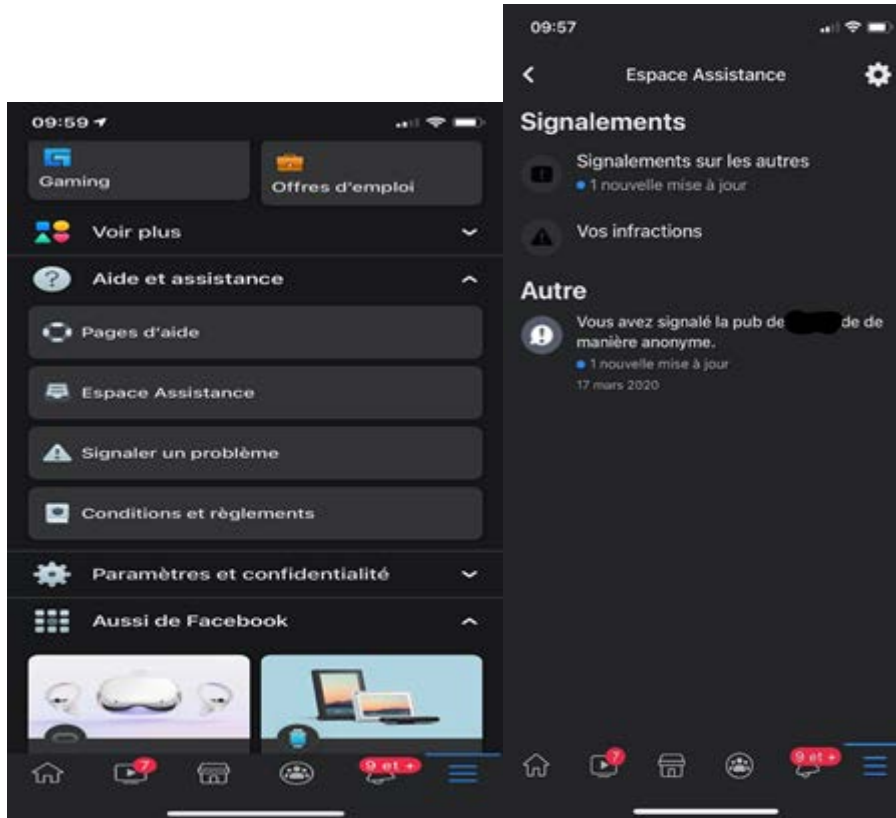
# FACEBOOK

- Capture d'écran des Pages d'aide de Facebook :

The screenshot shows the Facebook help page for reporting content. On the left is a navigation menu titled 'Facebook - Mode d'emploi' with options like 'Création d'un compte', 'Ajout d'amis', 'Votre page d'accueil', 'Messagerie', 'Stories', 'Vos photos et vidéos', 'Vidéos sur Watch', 'Pages', 'Groupes', 'Événements', 'Collectes de fonds et dons', and 'Facebook Pay'. The main content area has the heading 'Puis-je vérifier le statut d'un élément que j'ai signalé à Facebook ou annuler un signalement ?' and a sub-heading 'Aide ordinateur' with a 'Partager un article' button. The text explains that users can check the status of their report in the 'Espace Assistance' and provides a list of actions: clicking on a report, annulling a report, and requesting a review. A note states that reports cannot be annulled if they have not yet been reviewed.

- Capture d'écran de l'Espace assistance de Facebook :

# FACEBOOK



- Page d'aide Facebook « [Puis-je vérifier le statut d'un élément que j'ai signalé à Facebook ou annuler un signalement ?](#) » indiquant que les utilisateurs peuvent suivre l'évolution de leur signalement en cliquant dans l'Espace assistance.

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 ;

# FACEBOOK

- 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.

Comme mentionné ci-dessus, nous avons mis en place un dispositif de signalement directement accessible et visible qui permet aux utilisateurs de signaler des fausses informations. Voir Question 1.1 pour plus de détails sur ce dispositif.

Nous disposons d'une équipe d'accessibilité dédiée et centralisée pour définir, développer et mettre en œuvre les exigences d'accessibilité des utilisateurs, la formation, les bonnes pratiques, les outils, les technologies, les métriques et tout autre support pour faire progresser les travaux d'accessibilité dans l'ensemble de nos établissements. En outre, nous nous référons aux commentaires et aux tests des utilisateurs et aux directives d'accessibilité au contenu Web du W3C, WCAG 2.1 AA, pour informer sur la manière dont nous mettons en œuvre l'accessibilité sur le service, y compris pour le signalement de contenu. Notre équipe d'accessibilité a étudié et amélioré l'accessibilité de nos flux de signalement afin de promouvoir la pleine utilisation et la pleine jouissance de nos services par les personnes handicapées.

Nous faisons progresser l'accessibilité numérique en appliquant des technologies de pointe comme l'intelligence artificielle pour créer de nouvelles fonctionnalités et de nouveaux services au service des personnes handicapées, par exemple notre technologie de description automatique de photos pilotée par l'intelligence artificielle (« texte alternatif automatique » ou « AAT ») décrit des objets dans des photos à des personnes malvoyantes grâce à une technologie d'assistance appelée « lecteur d'écran ». Plus de 80 % des images affichées sur Facebook et Instagram contiennent désormais « AAT » ; notre outil d'accessibilité de reconnaissance faciale aidant les personnes malvoyantes à en savoir plus sur qui apparaît sur leurs photos (sous réserve des paramètres des autres utilisateurs) ; le sous-titrage codé aide les personnes malentendantes et utilisant l'intelligence artificielle, nous fournissons un sous-titrage généré automatiquement pour Facebook Live, Workplace Live et Instagram TV, ainsi que pour les publicités et les pages.

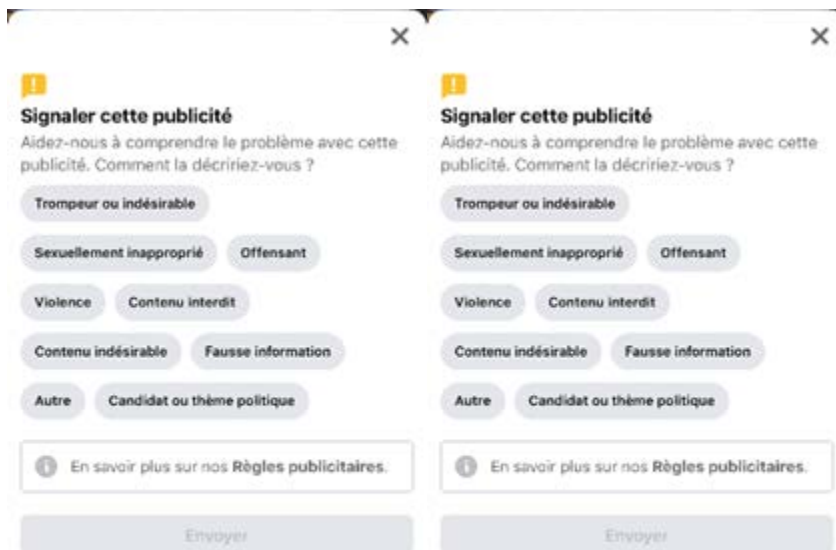
# FACEBOOK

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.

Facebook fournit les mêmes moyens de signaler des contenus quel que soit le mode d'accès au service ou le type de contenu en cause.

- Captures d'écran des outils de signalement pour les publications à partir d'une variété de dispositifs montrant que l'option de signalement « *fausse information* » est présentée aux utilisateurs de la même manière. Par exemple:

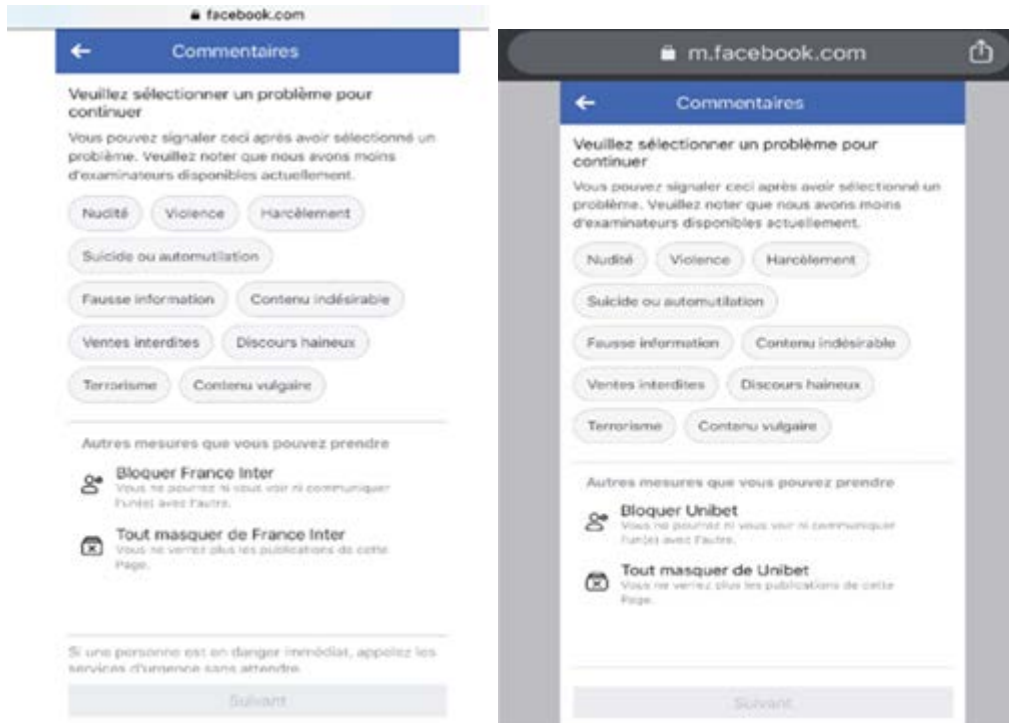
*Capture d'écran 1 - outil signalement sur l'application du téléphone (IOS/Android)*





# FACEBOOK

Capture d'écran 2 – outil de signalement sur le navigateur du téléphone (IOS – Safari /Android – Google Chrome)



# FACEBOOK

Capture d'écran 3 - outil de signalement sur ordinateur (Google Chrome/Internet Explorer)



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'assortir le signalement d'explications, d'une pièce jointe ou d'un lien, etc.).

Il est renvoyé à la Question 1.1.

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.

Il est renvoyé à la Question 1.2.

## **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.

# FACEBOOK

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.

Nous avons une approche systématique et nous nous efforçons de répondre aux signalements des utilisateurs concernant d'éventuelles violations de nos Standards de la communauté (pour Facebook) ou Règles de la communauté (pour Instagram) autant que possible dans les meilleurs délais, en hiérarchisant les signalements pouvant créer une situation de danger. Ces dernières années, l'équipe responsable de la sécurité sur Facebook et Instagram était constituée d'environ 35 000 personnes, dont environ 15 000 modérateurs de contenus au niveau mondial. Le travail de ces équipes (formation, statut, lieu, organisation du travail) est décrit [ici](#).

Lorsque les utilisateurs nous signalent du contenu, nous examinons ce contenu afin de déterminer si celui-ci viole nos Standards de la communauté (Facebook) ou nos Règles de la communauté (Instagram). Si nous déterminons qu'un contenu viole nos politiques, nous le supprimons. Ce contenu devient alors totalement inaccessible. Notre équipe « *Community Operations* » utilise et applique des directives claires lors de l'examen des signalements. Nous fournissons également à nos équipes des outils d'examen du contenu signalé et du contexte disponible requis pour identifier le problème et déterminer si un contenu viole nos Standards ou nos Règles. Dans les Standards de la Communauté et le Guide d'accompagnement du CSER, Facebook publie les standards et directives d'application détaillés utilisés par notre équipe « *Community Operations* ». De plus amples informations sur la façon dont nous revoyons le contenu et le rôle de la technologie sont disponibles [ici](#).

Si le contenu d'un utilisateur est supprimé en raison d'une violation de nos Standards ou Règles, l'utilisateur recevra dans la plupart des cas une notification de la suppression et pourra, demander un examen supplémentaire. Le contenu restera inaccessible pendant la durée de ce nouvel examen. Pour conserver toute objectivité dans leur jugement, les personnes en charge de ce deuxième examen ignorent que la publication a déjà été examinée.

Si le deuxième examen conduit à une décision conforme à la décision initiale, le contenu demeurera supprimé de Facebook ou d'Instagram. En cas de désaccord avec la décision initiale, il y aura une troisième revue, laquelle sera définitive.

# FACEBOOK

De même, si nous recevons un signalement d'un utilisateur et ne supprimons pas le contenu, l'utilisateur qui a fait le signalement peut demander un examen supplémentaire, qui suivra la même procédure que celle décrite ci-dessus.

Nous avons également mis en place un conseil de surveillance auprès duquel les utilisateurs peuvent contester les décisions prises par Facebook concernant un contenu. Toute personne en désaccord avec la décision de Facebook de suppression d'un contenu et qui a épuisé toutes les voies de recours dispose de 15 jours pour formuler un recours auprès du conseil de surveillance. Facebook peut également soumettre directement les cas importants et difficiles au conseil. Enfin, Facebook souhaite aussi permettre aux utilisateurs de soumettre des recours concernant le contenu qui demeure sur le service. Le conseil de surveillance peut déterminer, parmi les cas soumis par Facebook ou un utilisateur, si Facebook a pris la bonne décision et peut faire des recommandations.

**Ceci étant dit, il convient de noter qu'en matière de désinformation, les *spams* et faux comptes sont, en très grande majorité, détectés grâce à l'automatisation avant ou sans un signalement d'utilisateur, alors que les signalements des fausses informations servent de signal pour nos modèles d'apprentissage automatique (*machine learning*) :**

- Dans le cas des *spams* et faux comptes, la grande majorité des violations de nos Standards ou Règles de la communauté sont détectées grâce à l'automatisation plutôt que par des signalements des utilisateurs ; plus de 99,5 % des *spams* et faux comptes sont identifiés par Facebook avant ou sans un quelconque signalement d'un utilisateur (voir le graphique ci-dessous à la Question 1.4).
- Dans le cas des fausses informations qui ne violent pas directement nos Standards ou Règles de la communauté, les signalements des utilisateurs ne déclenchent pas le processus d'examen décrit ci-avant. Nous utilisons ces signalements comme un signal, qui va alimenter et aider nos modèles d'apprentissage automatique à identifier les contenus potentiellement faux et à déterminer si un contenu doit faire l'objet d'un examen par nos partenaires de vérification indépendants (*fact checkers*). Plus il y a de signalements de fausses informations et plus le modèle d'apprentissage automatique sera efficace. La procédure est la même pour Instagram. Parmi les critères utilisés pour détecter une fausse information, nous tenons notamment compte de l'existence, sous une publication, de nombreux commentaires exprimant une défiance (par exemple : « *c'est faux* »), ou du fait qu'un profil ait déjà publié de fausses informations par le passé. Enfin, les contenus

# FACEBOOK

identifiés comme faux sur Facebook vont également l'être sur Instagram. Ainsi, si un contenu est identifié comme faux ou partiellement faux sur Facebook, un contenu identique sera automatiquement identifié comme tel sur Instagram (et vice versa).

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 ;

Il est renvoyé aux Questions 1.3.1, 1.3.3 et 1.4 de cette Déclaration infox 2020.

Comme expliqué dans cette Déclaration 2020, nous avons mis en place des **moyens humains et/ou automatisés** de détection proactive des contenus diffusant de fausses informations. Nous comptons notamment sur les ressources humaines/ partenaires de vérification indépendants pour l'instruction. L'intelligence artificielle est également un outil déterminant pour détecter les contenus préjudiciables (voir un aperçu [ici](#)). L'articulation entre les moyens humains et automatiques dépend du contenu en cause. En outre, nous avons adapté nos politiques et apporté divers changements en réponse à la pandémie de la COVID-19 (voir Questions générales de cette Déclaration 2020 et [ici](#)).

- 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*...).

Facebook fait partie du « Groupe de Contact Permanent » dirigé par le ministère de l'Intérieur. Nous nous réunissons régulièrement pour améliorer notre processus et contribuer au travail de

# FACEBOOK

coordination avec les autorités françaises. En termes de résultats, vous pouvez trouver tous les chiffres concernant les requêtes des autorités françaises [ici](#). Entre janvier et décembre 2020, nous avons répondu à 82% des demandes d'accès à des données pour la France. En cas de crise, nous disposons également d'un processus et d'un canal spécifiques pour les demandes urgentes provenant des autorités françaises.

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.

Nous combinons l'intelligence humaine et la technologie pour soumettre de fausses informations potentielles à nos partenaires de vérification indépendants. Ces derniers peuvent également identifier de manière proactive le contenu à évaluer et à revoir.

Fondamentalement, les technologies automatisées ne peuvent pas déterminer ce qui est vrai ou faux, mais elles jouent un rôle important dans notre action contre la désinformation.

Comme évoqué ci-dessus, notre système de lutte contre la désinformation est hybride en ce qu'il utilise à la fois l'intelligence humaine et la technologie (voir Question 1.3.1). Par exemple, nous utilisons les avis des utilisateurs qui marquent le contenu qu'ils voient dans leur fil d'actualité comme potentiellement faux. Nous observons également d'autres comportements des utilisateurs, comme les personnes qui commentent et affirment ne pas croire telle ou telle publication.

Nous utilisons également des modèles d'apprentissage automatique pour améliorer constamment notre capacité à anticiper les fausses informations. Nous réinjectons dans ce modèle les évaluations de nos partenaires de vérification indépendants, afin d'améliorer au fil du temps notre capacité à prévoir les contenus qui pourraient être faux. C'est un processus de développement continu. Par exemple, nos modèles d'apprentissage automatique identifient les liens vers des articles pouvant être faux, et les prédictions de ces modèles peuvent être utilisées par nos partenaires de vérification indépendants. Nos partenaires examinent ensuite les contenus, vérifient les faits, et évaluent leur véracité. Les contenus faux sont rétrogradés dans le fil d'actualité afin d'être moins visibles pour les utilisateurs. Les évaluations des partenaires de vérification indépendants permettent ensuite à notre modèle d'apprentissage automatique de s'améliorer, ce qui crée un processus cyclique. Pour plus d'informations à ce sujet, voir

# FACEBOOK

<https://www.facebook.com/journalismproject/programs/third-party-fact-checking/how-it-works>.

Dans le cas d'Instagram, nous utilisons également notre technologie pour tenter d'identifier les écrits ayant fait l'objet d'une vérification factuelle qui sont apposés sur des images, par exemple dans des mèmes (à savoir un élément ou un phénomène repris et décliné en masse sur Internet). Cette fonctionnalité est particulièrement importante dans le cas d'Instagram, où les fausses informations apparaissent souvent sous la forme d'un mème ou d'un texte superposé sur une image. Si le texte apposé sur l'image a été identifié comme faux, nous le retirerons (ainsi que toute copie) de la section Explorer (icône loupe) et des pages de hashtag (recherche d'informations via un hashtag) et nous réduirons également sa visibilité dans le fil d'actualité et les *stories*. En outre, comme pour Facebook, nous travaillons sur Instagram avec des partenaires de vérification indépendants pour nous aider à identifier, examiner et marquer les fausses informations.

Nos modèles d'apprentissage automatique nous aident également à identifier les copies des contenus d'information ayant été rétrogradées. Par exemple, après qu'un partenaire en France ait qualifié de fausse l'affirmation selon laquelle vous pouvez sauver une personne ayant un AVC en piquant son doigt avec une aiguille et en la faisant saigner, nous avons pu identifier plus de 1.400 liens et plus de 20 domaines diffusant la même affirmation, et ainsi limiter leur diffusion.

Notre modèle d'apprentissage automatique pour les appâts à clics (*clickbait*) fonctionne de la même manière (le *clickbait* est un problème relié à celui de la désinformation, les deux allant souvent de pair). Nous avons examiné des centaines de milliers de titres de publications pour décider s'il s'agissait de *clickbait* ou pas, en considérant par exemple et de manière distincte, si le titre traite de manière exagérée un simple détail d'une histoire ou dissimule des informations. Une équipe de Facebook a examiné des milliers de titres en utilisant ces critères, chaque membre de l'équipe validant le travail des autres, afin d'identifier de grands ensembles de titres racleurs. A partir de ce travail, nous avons identifié les expressions communément utilisées dans les titres racleurs qui ne sont pas utilisées dans les autres titres.

Il s'agit du même mode de fonctionnement que la majorité des filtres à *spams*. Lorsque nous identifions des contenus d'information présentant les signaux des *clickbaits*, le modèle prédit la probabilité que la publication soit un *clickbait*. Pour l'avenir prévisible, tous ces systèmes conserveront un fonctionnement hybride. La façon dont la technologie et les utilisateurs signalent des faux contenus devront s'améliorer afin de s'adapter aux changements à venir.

# FACEBOOK

Concernant plus particulièrement les signaux, nous pouvons utiliser différentes approches pour identifier un faux contenu, sachant que le contenu en soi est seulement l'un des signaux potentiellement importants. Les autres types de signaux à prendre en compte sont notamment la source du contenu, la personne qui le partage, l'intérêt des personnes pour le contenu, etc. Les prédictions du modèle peuvent être très efficaces même sans considérer le contenu en lui-même et les méthodes d'analyse du langage naturel jouent un rôle dans le système global de plusieurs façons. Tirer parti de la façon dont les faux contenus se propagent, par rapport à d'autres contenus d'information non faux, en modélisant les aspects de ces séquences dans le temps est un domaine dans lequel nous investissons activement.

Par exemple :

- Notre modèle d'apprentissage automatique s'appuie sur des méthodes d'analyse du langage naturel – par exemple, un des signaux que nous utilisons, en plus des retours des utilisateurs indiquant qu'un contenu d'information peut être faux, sont les commentaires mettant en cause la véracité de ce contenu. Nous le faisons dans plusieurs langues et devons faire preuve d'une grande finesse pour distinguer ce qui relève de la suspicion et ce qui relève de l'humour.
- Le modèle inclut des signaux liés à la source. Savoir qu'une page ou un site Internet a partagé du faux contenu par le passé est une bonne indication que cela se reproduira à l'avenir.
- Nous utilisons également les informations communiquées par nos partenaires de vérification indépendants pour entraîner notre modèle d'apprentissage automatique, afin que nous puissions identifier davantage de nouveaux contenus d'information potentiellement faux et de manière plus rapide.
- Nous ne divulguons pas tous les signaux utilisés dans ce modèle afin qu'ils ne puissent pas être contournés par des utilisateurs malintentionnés.
- Nous vérifions également si les revendications exprimées dans un contenu sont similaires à celles de contenus qui auraient déjà été qualifiés de faux par nos partenaires de vérification indépendants.
- Nous vous invitons à lire le post accessible à l'adresse ci-après pour vous informer sur notre utilisation de l'IA pour traiter davantage de langues :



# FACEBOOK

<https://code.fb.com/mlapplications/expanding-automatic-machine-translation-to-more-languages/>

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).

Les Questions 1.3.1 et 1.4 détaillent les **différentes étapes de l'instruction**.

Les partenaires de vérification indépendants jouent un rôle important dans l'instruction concernant un contenu. Ils examinent et évaluent l'exactitude des déclarations au moyen d'outils fiables, notamment en interrogeant les sources primaires, en consultant les données publiques et en effectuant des analyses des médias, y compris des photos et des vidéos. Les **options** d'évaluation que les partenaires de vérification indépendants peuvent appliquer au contenu sont décrites dans la section consacrée aux systèmes de vérification des informations (voir Question 3.3).

En ce qui concerne la **procédure d'instruction spécifique pendant la période électorale**, nous protégeons l'intégrité des élections.

Protéger l'intégrité des élections tout en préservant la liberté d'expression est une priorité absolue pour Facebook. En nous appuyant sur les leçons du passé et les commentaires d'experts et de décideurs de tous les horizons politiques, nous avons fait des investissements substantiels dans nos équipes et technologies afin de mieux sécuriser les élections, et nous les déployons là où elles auront le plus grand impact. Notre stratégie de protection des élections s'applique non seulement en période critique, mais aussi tout au long de l'année. Cette stratégie s'articule autour de trois axes : prévenir les interférences ; supprimer les contenus préjudiciables et réduire les fausses informations ; augmenter la transparence.

Un élément clé de notre stratégie de prévention des interférences consiste à travailler avec les autorités gouvernementales, les forces de l'ordre, les experts en sécurité, la société civile et d'autres entreprises technologiques afin de mettre un terme aux menaces émergentes en établissant une ligne de communication directe, en partageant nos connaissances et en identifiant les possibilités de collaboration. Nous avons fait évoluer nos opérations de sécurité afin de démanteler les campagnes d'opérations d'influence et d'identifier les menaces

# FACEBOOK

émergentes. Nos équipes d'enquêteurs spécialisés recherchent activement et démantèlent les réseaux coordonnés de comptes, de pages et de groupes inauthentiques qui cherchent à manipuler le débat public. Nous avons créé des équipes spécialisées, qui travaillent avec des experts de premier plan, pour mettre fin à ce type d'activité, ce qui a entraîné la suppression de plus de 100 réseaux de comportement inauthentique coordonné au cours des quatre dernières années. Nous avons mis à jour notre politique de comportement inauthentique pour améliorer encore notre capacité à contrer de nouvelles tactiques et à agir plus rapidement contre l'éventail de pratiques trompeuses que nous voyons sur nos services, qu'elles soient étrangères ou nationales, étatiques ou non étatiques.

En ce qui concerne la **procédure d'instruction spécifique en situation de crise (par exemple en cas de crise sanitaire)**, nous augmentons notamment nos efforts pour supprimer les fausses déclarations sur Facebook et Instagram, en particulier concernant la COVID-19, les vaccins contre la COVID-19 et les vaccins en général dans le contexte de la pandémie ainsi que dans celui de la lutte contre le changement climatique. Pour plus de détails, voir Questions générales.

## 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.

L'**issue de l'instruction** est détaillée dans notre stratégie de lutte contre les fausses informations dans les Questions générales ainsi que dans les Questions 1.3.1 et 1.4.

Là encore, nous comptons notamment sur nos partenaires de vérification indépendants pour nous permettre d'agir en cas de fausse information potentielle. Une fois qu'un partenaire de vérification indépendant a évalué un contenu (ce dernier pouvant être faux, modifié, partiellement faux ou contexte manquant), nous prenons des mesures. Par exemple, nous pouvons réduire la distribution du contenu en cause, ajouter un avertissement de partage, des notifications de partage, des étiquettes de désinformation, ou supprimer des avantages pour les récidivistes. Pour plus d'informations à ce sujet, voir Question 3.6.

# FACEBOOK

Dans le cadre de la crise sanitaire liée à la COVID-19, nous n'autorisons pas un certain nombre de déclarations concernant la COVID-19 telles qu'énumérées ci-dessus. Nous nous efforçons d'éliminer le contenu lié à la COVID-19 qui contribue au risque de préjudice réel, y compris sur la base de nos politiques relatives à la violence imminente ou de préjudice physique.

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 Question 1.1 détaille les **informations communiquées à l'auteur du signalement**, la Question 1.3.1 les informations communiquées à **l'auteur du contenu signalé** et la Question 1.3 les informations communiquées aux **différents utilisateurs**.

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.

La Question 1.3.1 fait référence **aux voies de recours internes** mises en place.

*Il précisera notamment :*

- *quelles sont les décisions de la plateforme susceptibles d'être contestées* ; La plupart des décisions.
- *quels sont les **utilisateurs en droit d'exercer un recours, dans quel délai la possibilité de recours doit leur être présentée à l'issue de la décision et par quelle voie*** ; Les utilisateurs disposent d'un délai de 30 jours pour obtenir un réexamen d'une décision de Facebook sur un contenu. Une fois qu'ils ont épuisé les possibilités de réexamen, les utilisateurs disposent d'un délai de 15 jours pour présenter un recours au conseil de surveillance.

# FACEBOOK

- *quelles sont les équipes en charge de l'instruction de ces recours ; L'équipe en charge d'enquêter sur ces recours est constituée de nos modérateurs de contenus dont le travail est décrit [ici](#).*
- *quels sont les critères de recevabilité des recours ; Il n'y a pas de critères spécifiques.*
- *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) ; Aucune information supplémentaire ne peut être fournie par les utilisateurs. Le recours consiste en un réexamen du contenu, sans contexte supplémentaire fourni par l'utilisateur.*
- *le délai moyen de traitement d'un recours ; Fournir un délai moyen de traitement des recours peut avoir une incidence sur la façon dont nos modérateurs de contenus font leur travail. En effet, certains contenus sont faciles à examiner, d'autres sont beaucoup plus compliqués et leur examen peut prendre plus de temps.*
- *si les informations sont fournies systématiquement aux utilisateurs s'agissant des recours internes ou s'il existe des exceptions, prévues ou de fait. Voir notre réponse au premier tiret.*

## 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.

Nous avons pris diverses initiatives pour évaluer de façon indépendante nos procédures et nos actions. A titre d'exemple :

- **Le Groupe consultatif sur la transparence des données**

Nous avons créé le Data Transparency Advisory Group (DTAG) en 2018. Il s'agit d'un organe indépendant composé d'experts internationaux en matière de mesure, de statistiques, de

# FACEBOOK

criminologie et de gouvernance. Leur tâche consiste à fournir une évaluation publique indépendante concernant la question de savoir si les indicateurs que nous partageons dans le [Rapport sur l'application des Standards de la communauté](#) (CSER) fournissent des mesures exactes et significatives relatives aux défis de modération du contenu de Facebook et de notre travail pour y remédier. [https://law.yale.edu/sites/default/files/area/center/justice/document/dtag\\_report\\_5.22.2019.pdf](https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf) Pour plus de détails, voir : <https://about.fb.com/news/2019/05/dtag-report/>

- **Vérification indépendante des paramètres du Rapport sur l'application des Standards de la communauté**

En août 2020, nous avons lancé une Demande de Proposition à des auditeurs externes afin qu'ils effectuent une vérification indépendante de nos indicateurs utilisés dans le cadre de notre [Rapport sur l'application des Standards de la communauté](#) (CSER). Cette vérification sans précédent vise à donner aux parties prenantes externes, y compris les régulateurs, une confiance supplémentaire dans l'exactitude des chiffres que nous communiquons au sujet des contenus préjudiciables dans le CSER. Aucune entreprise ne devrait noter ses propres obligations et la crédibilité de nos systèmes devrait être gagnée, et non supposée. Nous croyons que des vérifications et des évaluations indépendantes sont déterminantes pour nous rendre responsables et nous aider à faire mieux. Pour plus de détails, voir <https://about.fb.com/news/2020/08/independent-audit-of-enforcement-report-metrics/>

- **Partenariat pour la confiance et la sécurité numériques**

En février 2021, des entreprises technologiques de premier plan ont officiellement lancé le [Partenariat pour la confiance et la sécurité numériques](#), une initiative unique en son genre visant à promouvoir un Internet plus sûr et plus fiable.

Le partenariat s'engage à élaborer des bonnes pratiques pour l'industrie, examinées au moyen d'évaluations internes et indépendantes par des tiers, afin d'assurer la sécurité et la confiance des consommateurs lorsqu'ils utilisent les services numériques. Regroupant une gamme diversifiée d'entreprises technologiques les plus innovantes de différentes tailles et modèles commerciaux, le Partenariat pour la confiance et la sécurité numériques regroupe Discord, Facebook, Inc, Google, Microsoft, Pinterest, Reddit, Shopify, Twitter et Vimeo.

# FACEBOOK

Dans le cadre de son lancement, le partenariat publie une série de cinq engagements clés convenus par les entreprises participantes pour assurer la sécurité et la confiance des consommateurs lors de l'utilisation des services numériques. Également connu sous le nom de « Cadre de bonnes pratiques du Partenariat pour la confiance et la sécurité numériques », le partenariat considère ce qui suit comme étant les bonnes pratiques de l'industrie :

Engagement 1 : Identifier, évaluer et ajuster les risques liés au contenu et à la conduite dans le développement de produits.

Engagement 2 : Adopter des processus explicatifs pour la gouvernance des produits, y compris quelle équipe est responsable de la création des règles et de la manière dont les règles évoluent.

Engagement 3 : Mener des opérations d'application de la loi pour mettre en œuvre la gouvernance des produits.

Engagement 4 : Évaluer et améliorer les processus associés aux risques liés au contenu et à la conduite.

Engagement 5 : S'assurer que les politiques de confiance et de sécurité pertinentes sont publiées auprès du public et rendre compte périodiquement au public et aux autres parties prenantes des mesures prises.

Chaque entreprise effectuera un examen de ses propres pratiques de confiance et de sécurité avant que le Partenariat pour la confiance et la sécurité numériques ne présente un rapport sur l'état de l'industrie, qui évaluera la mise en œuvre des pratiques par les entreprises. Cet effort n'est pas axé sur les politiques ou les décisions individuelles de l'entreprise, mais se concentre sur la façon de réaliser et d'évaluer la confiance et la sécurité en tant que domaine en évolution. Le partenariat collaborera avec les défenseurs des consommateurs/utilisateurs, les responsables politiques, les forces de l'ordre, les ONG compétentes et divers experts de l'industrie afin d'aider à développer ces bonnes pratiques.

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

- le **nombre de contenus signalés** comme fausses informations ;
- le nombre de procédures de traitement mises en œuvre ;

# FACEBOOK

- 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**.

Comme cela a été mentionné à la Question 1.3.1 ci-dessus, Facebook associe les signalements des utilisateurs et les technologies d'apprentissage automatique avancées pour identifier et prendre des mesures concernant les contenus qui violent les Standards de la communauté (Facebook) ou les Règles de la communauté (Instagram). Par conséquent, nous ne considérons pas que le nombre de signalements reçus soit un indicateur significatif, en particulier lorsqu'il s'agit de contenu lié à la désinformation. L'une des raisons en est que les experts, gouvernements et plateformes ne peuvent s'accorder sur la définition de « désinformation », et il n'existe aucun moyen de tracer une ligne nette entre information fausse, satire et simple opinion.

**Nous nous concentrons plutôt sur le nombre de contenus retirés car violant nos Standards ou Règles de la communauté, et ce quelle que soit la manière dont ils ont été identifiés.** Les indicateurs et statistiques sont fournis aux utilisateurs dans notre [Rapport sur l'application des Standards de la communauté](#) (CSER). Fournir un délai moyen de traitement du signalement peut avoir une incidence sur la façon dont nos modérateurs de contenus font leur travail. En effet, certains contenus sont faciles à examiner, d'autres sont beaucoup plus compliqués et leur examen peut prendre plus de temps.

En ce qui concerne le budget, comme mentionné ci-avant, l'équipe responsable de la sécurité sur Facebook et Instagram est constituée d'environ 35 000 personnes, et aujourd'hui, notre budget lié à la sécurité est supérieur à l'ensemble des revenus de notre entreprise au moment de notre introduction en bourse au début de la décennie.

# FACEBOOK

## (a) Les faux comptes

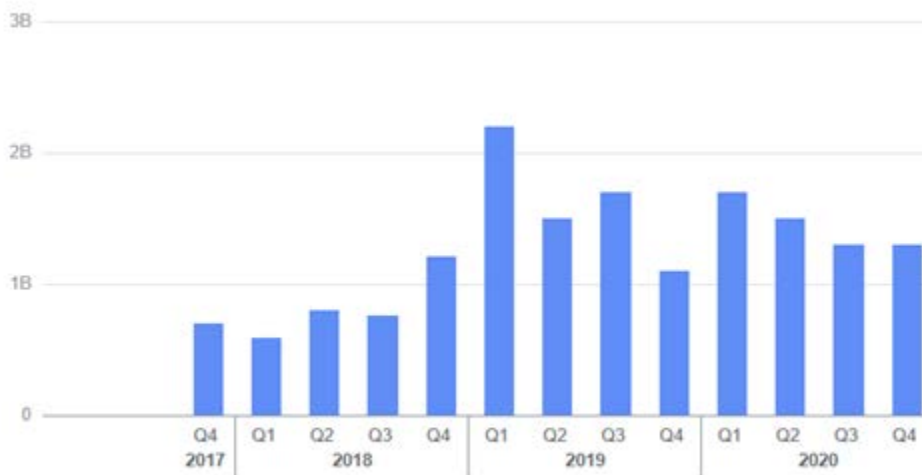
Nous souhaitons supprimer le plus de faux comptes possibles. Nous traitons en priorité les faux comptes qui ont pour objectif de causer des dommages.

Grâce à notre technologie de détection, nous bloquons des millions de tentatives de création de faux comptes chaque jour et en détectons des millions supplémentaires souvent dans les minutes suivant leur création. Les chiffres qui suivent n'incluent pas les tentatives bloquées de création de comptes. Comme indiqué dans notre dernier [Rapport d'application des Standards de la communauté](#), nous estimons que les faux comptes représentent environ 5 % de nos utilisateurs actifs mensuels, au niveau mondial, sur Facebook durant le 4e trimestre 2020.

Le graphique ci-dessous détaille le nombre de faux comptes que nous avons désactivés :

### ACCOUNTS ACTIONED

How many fake accounts did we take action on?



Le nombre de faux comptes désactivés a augmenté, passant de 1,1 milliard de comptes au 4<sup>e</sup> trimestre 2019 à 1,7 milliard au 1<sup>er</sup> trimestre 2020, en grande partie en raison de l'augmentation des attaques automatisées par les acteurs mal intentionnés qui tentent de créer une grande



# FACEBOOK

quantité de comptes en même temps. Puis, au 2<sup>e</sup> trimestre 2020, 1,5 milliard de faux comptes ont été désactivés, et 1,3 milliard aux 3<sup>e</sup> et 4<sup>e</sup> trimestres 2020. Dans leur majorité, ces comptes ont été identifiés dans les minutes suivant leur création, avant qu'ils ne fassent partie de notre population d'utilisateurs actifs mensuels. Pour les comptes qui ne sont pas initialement détectés et désactivés, nous constatons que beaucoup d'entre eux sont utilisés dans des campagnes de *spam* et sont créés pour des raisons financières.

Nous nous efforçons de supprimer les faux comptes qui abusent de nos services des trois différentes manières décrites ci-dessous :

- **Empêcher la création de faux comptes** : Le meilleur moyen de lutter contre les faux comptes est, en premier lieu, d'empêcher leur création sur Facebook. Nous avons créé une technologie de détection capable de détecter et bloquer les comptes avant même leur création. Nos systèmes utilisent différents critères pour ce faire, par exemple l'indication que des comptes sont créés en masse depuis un même endroit. Dans ce cas, nous pouvons par exemple bloquer totalement certaines adresses IP, afin qu'elles ne puissent plus accéder à nos systèmes et créer des comptes.

Comme précisé supra, dans la mesure où il est impossible d'identifier le nombre de tentatives de création d'un compte que nous avons bloqué (car, par exemple, nous empêchons un ensemble d'adresses IP d'accéder à notre site), le [Rapport d'application des Standards de la communauté](#) précité n'inclut pas les données relatives aux tentatives infructueuses de création de faux comptes que nous avons bloqués.

Nous pouvons toutefois estimer que nous empêchons chaque jour la création de millions de faux comptes à l'aide de ces systèmes de détection.

- **Suppression des faux comptes lors de leur création** : Nos systèmes de détection avancés recherchent les faux comptes potentiels dès leur création, en détectant les signes de comportement malveillant. Ces systèmes associent des signaux tels que l'utilisation d'adresses électroniques suspectes, les actions suspectes et d'autres signaux préalablement associés à d'autres faux comptes que nous avons supprimés. Les comptes que nous supprimons actuellement sont, pour la plupart, bloqués dans les minutes suivant leur création, avant qu'ils ne puissent occasionner un quelconque dommage.

# FACEBOOK

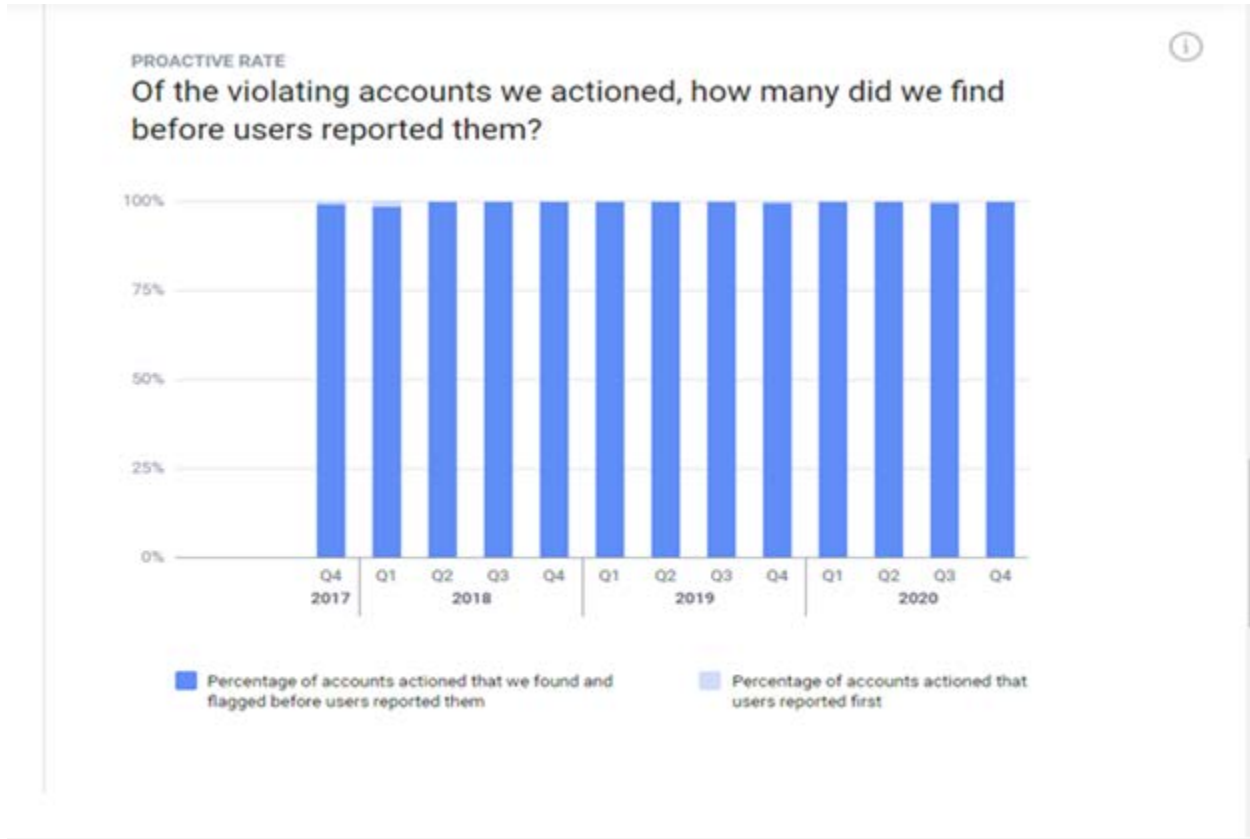
Les comptes que nous désactivons lors de leur création sont inclus dans nos statistiques sur le nombre de faux comptes supprimés. Ces statistiques peuvent varier en raison de la survenance d'attaques peu sophistiquées, ce qui est souvent le cas. Ces attaques, aisément identifiables, ont une influence significative sur nos statistiques, alors même qu'elles représentent un risque minime pour les utilisateurs. Par exemple, un *spammer* peut essayer de créer 1 000 000 comptes en un court laps de temps, à partir de la même adresse IP. Dans ce cas, nos systèmes vont détecter le problème et supprimer rapidement ces faux comptes. Ce nombre de faux comptes supprimés va s'ajouter au nombre global de comptes supprimés. Toutefois, ces comptes ayant été supprimés extrêmement rapidement, ils n'ont jamais été considérés comme actifs et ne seront donc pas pris en compte notamment dans notre évaluation de la prévalence de faux comptes parmi les utilisateurs actifs mensuels.

- **Suppression des faux comptes déjà actifs sur Facebook** : Certains comptes peuvent parvenir à passer au travers des deux filtres ci-dessus et devenir actifs sur le service. C'est le cas lorsque ces comptes ne montrent pas immédiatement de signes de fausseté ou de malveillance, aussi nous leur accordons le bénéfice du doute jusqu'à ce qu'ils montrent des signes d'activité malveillante. Ces comptes sont repérés par nos systèmes de détection lorsque ces derniers identifient un tel comportement ou lorsque les utilisateurs de Facebook nous les signalent. Nous utilisons un certain nombre de signaux concernant le mode de création du compte et son utilisation afin de déterminer un taux de probabilité qu'il soit faux et le désactiver lorsque ce taux est élevé.

Les comptes que nous supprimons à ce stade sont pris en compte dans nos statistiques relatives à nos actions sur les comptes. Si ces comptes sont actifs sur la plateforme, nous les comptabilisons également dans notre indicateur de prévalence. La prévalence des faux comptes mesure combien de faux comptes actifs existent parmi nos utilisateurs actifs mensuels au cours d'une période donnée.

Le graphique ci-dessous révèle que 99,6 % des comptes que nous supprimons ont été détectés avant tout signalement :

# FACEBOOK



# FACEBOOK

## (b) Concernant les fausses informations

Il est renvoyé aux Questions générales pour notre approche de la lutte contre les fausses informations.

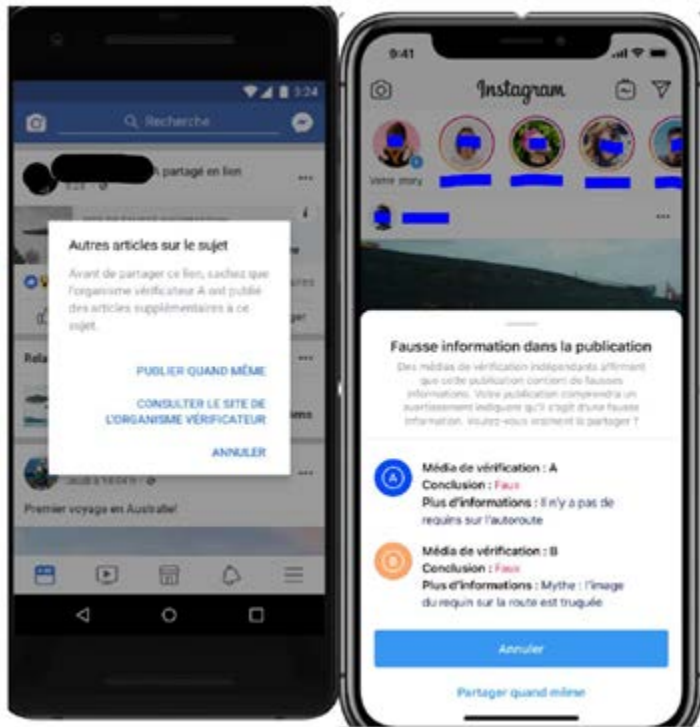
Le fil d'actualité est classé sur la base de centaines de milliers de signaux, et certains signaux peuvent identifier des appâts à clics (*clickbait*), fausses informations ou appâts à interagir (*engagement bait*). Nous utilisons d'autres signaux pour identifier les contenus faux, notamment la source du contenu, la personne qui le partage et l'intérêt des personnes pour le contenu. Les partenaires de vérification indépendants peuvent également porter à notre attention les tendances en matière de fausses informations et leurs sources. En effet, les partenaires de vérification indépendants du monde entier peuvent, à leur discrétion, émettre des doutes sur la véracité d'un contenu, que ce soit sur Facebook ou sur Instagram, et nous le signaler via un outil où ils peuvent renseigner leur évaluation et y joindre le lien de référence qui détaille leur analyse.

Nous nous efforçons également de rendre plus difficile le partage d'un contenu jugé comme faux par les partenaires de vérification indépendants. Lorsque les utilisateurs tentent de partager un tel contenu, une fenêtre pop-up apparaît pour les informer que le contenu a été jugé comme faux par un partenaire de vérification indépendants et les encourage à consulter l'analyse faite par ce partenaire de vérification et à bien réfléchir avant de partager un tel contenu (voir ci-dessous pour des exemples des captures d'écran sur Facebook et sur Instagram) :

# FACEBOOK

Facebook (à gauche) et Instagram (à droite) :

Les captures d'écran seront mises à jour si nécessaire.



- (i) Suppression des fausses informations contribuant à causer un dommage physique

Concernant les fausses informations pouvant contribuer à causer un dommage physique imminent, nous avons remarqué que, dans certains cas, les utilisateurs publient du contenu qui, en lui-même, ne viole pas nos Standards de la communauté (Facebook) ou nos Règles de la communauté (Instagram) et semble inoffensif, mais qui, pris dans le contexte spécifique d'un pays ou d'une région (par exemple, en cas de tensions ethniques et religieuses ou d'antécédents de violences), peut constituer une incitation à, ou même causer, un dommage physique réel. Ainsi, nous supprimons les fausses informations qui peuvent contribuer à une violence imminente ou à un dommage physique. Pour appliquer cette politique, nous avons fait appel à des organisations locales et institutions internationales, qui sont souvent les premières à prendre

# FACEBOOK

connaissance des informations inexactes ou trompeuses pouvant contribuer à des dommages réels. Nous avons établi des canaux pour que ces organisations nous contactent et nous transmettent de tels contenus rapidement.

Les fausses déclarations sur Facebook et Instagram concernant la COVID-19, les vaccins contre la COVID-19 et les vaccins en général sont un exemple de fausses informations contribuant à causer un dommage physique.

- (ii) Suppression des fausses informations contribuant à la pression sur les électeurs

Dans le cadre de nos efforts constants pour éviter que les utilisateurs utilisent Facebook ou Instagram de façon malintentionnée durant les élections, nous avons, en automne 2019, renforcé notre politique de lutte contre la pression sur les électeurs, soit contre les comportements visant à dissuader ou empêcher les utilisateurs de voter. Nos politiques couvrent de nombreux types de contenus d'ingérence dans le vote qui apparaissent sur la plateforme, en ce compris :

- La présentation erronée des dates, lieux, heures et moyens de vote ou d'inscription des électeurs (exemple : « *Vote par texto !* ») ;
- La présentation erronée des personnes habilitées à voter, des conditions requises pour voter, du mode de comptabilisation des votes et des informations et/ou documents devant être fournis pour voter ; et
- Les menaces de violence concernant le vote, l'inscription des électeurs ou le résultat d'une élection.

Nous supprimons ce type de contenu quelle que soit sa provenance.

En outre, nous sommes conscients que d'autres types de contenus, comme les discours de haine, peuvent créer une forte pression sur les électeurs. C'est la raison pour laquelle nos politiques de lutte contre les discours haineux interdisent d'empêcher des utilisateurs de participer à la vie politique en raison de leur nationalité, origine ethnique ou leur religion (par exemple, il sera interdit de dire à des utilisateurs de ne pas voter pour un candidat en raison de la nationalité de ce dernier, ou de déclarer que les personnes d'une certaine religion ne devraient pas être autorisées à se voir confier une charge publique).

# FACEBOOK

Certaines publications qui nous sont signalées peuvent nécessiter un examen additionnel. En effet, nous ne pouvons pas vérifier la véracité de l'intégralité des publications relatives aux bureaux de vote du monde entier (par exemple : « *Inondation d'une école élémentaire, Bureau de vote fermé* »). Dans de tels cas, nous envoyons le contenu en cause à nos partenaires de vérification indépendants pour examen. Le contenu qualifié de faux sera rétrogradé dans le fil d'actualité et accompagné d'une information additionnelle sur le même sujet, rédigée par nos modérateurs (ce que nous appelons les « *articles liés* »). Nous continuons également à enrichir et à développer nos partenariats afin de développer une expertise sur les tendances en matière de pressions sur les électeurs, dans le but de détecter rapidement tout contenu contrevenant.

### (iii) Vidéos manipulées et trompeuses

Parfois, les utilisateurs partageant du contenu sur nos services s'inscrivent dans un but de manipulation des médias, afin de tromper les lecteurs.

Ce type de manipulation peut être réalisé à l'aide d'une technologie simple ou plus sophistiquée, qui utilise l'intelligence artificielle ou des techniques d'apprentissage approfondi (*deep learning*) afin de créer des vidéos qui déforment la réalité – communément appelées *deepfakes* (ou hypertrucages). Bien que des vidéos de ce type soient encore rares sur Internet, elles représentent un défi pour notre secteur et pour la société, car leur usage devient plus fréquent.

Notre approche à cet égard est diversifiée, et va de l'examen des contenus générés par l'intelligence artificielle et des comportements trompeurs comme les faux comptes, au partenariat avec le monde universitaire, le gouvernement et le secteur privé afin de démasquer les personnes à l'origine de ces contenus.

Ici, la collaboration est clé. Nous avons engagé des discussions avec plus de 50 experts à travers le monde dans les secteurs technique, politique, médiatique, juridique, civique et académique, afin de mieux orienter notre stratégie de développement et d'améliorer notre technologie de détection de manipulation des médias.

Grâce à ces partenariats et discussions, nous renforçons continuellement nos politiques à l'encontre des vidéos de manipulation trompeuses identifiées comme *deepfakes*. Depuis 2020, nous avons supprimé les contenus vidéos manipulés et trompeurs qui remplissent les critères suivants :

# FACEBOOK

- Le contenu a été édité ou raccourci – pour des raisons autres que de clarté ou de transparence – d’une façon qui n’est pas évidente pour des profanes et qui pourrait laisser croire que le sujet de la vidéo tient des propos qu’il n’a en réalité jamais prononcés ; et
- Le contenu est le produit d’une intelligence artificielle ou d’un apprentissage automatique qui fusionne, remplace ou superpose des contenus dans une vidéo qui semble vraie.

Cette politique ne s’applique pas à des contenus parodiques ou satiriques, ou à des vidéos qui ont été modifiées seulement dans le but de supprimer ou de modifier l’ordre des mots.

Nous sommes également impliqués dans l’identification des contenus manipulés, parmi lesquels les *deepfakes*, qui sont les plus difficiles à détecter. Dans cette optique, nous avons lancé en septembre 2019 le *Deepfake Detection Challenge*, qui a incité des personnes du monde entier à entreprendre des recherches et à créer des outils en libre accès (*open source*) pour détecter les *deepfakes*. Ce projet, que nous subventionnons à hauteur de 10 millions de dollars américains, réunit des organisations de tous les secteurs, parmi lesquels le Partenariat sur l’AI (*Partnership on AI*), Cornell Tech, l’Université Berkeley en Californie, le MIT, WITNESS, Microsoft, et d’autres membres de la société civile et des communautés technologiques, médiatiques et académiques.

Par ailleurs, nous nous sommes associés avec Reuters, la société fournissant le plus de nouvelles multimédia dans le monde, pour aider les sociétés d’information du monde entier à identifier les *deepfakes* et les médias manipulés à l’aide d’une formation en ligne gratuite. Les sociétés d’information comptent de plus en plus sur l’aide de tiers pour obtenir de grandes quantités d’images et de vidéos, et identifier les visuels manipulés est un défi de taille. Cette formation vise à aider les sociétés de presse à mener à bien ce travail.

En plus du fait que nos partenariats prennent de l’importance et que nos propres connaissances évoluent, nos politiques de lutte contre les médias manipulés ont aussi vocation à évoluer. Entre temps, nous travaillons en interne et avec d’autres partenaires de notre secteur pour trouver des solutions réellement efficaces.

## (iv) Vidéos diffusées en direct sur Facebook Live

Nous avons pour objectif de minimiser les risques d’abus réalisés lors de l’utilisation de la fonctionnalité « Live » (diffusion de contenu en direct), tout en permettant aux utilisateurs d’utiliser chaque jour le Live de manière positive. Nous avons réfléchi à plusieurs moyens pour nous assurer que le Live ne soit pas utilisé dans le but de causer un dommage ou de propager la



# FACEBOOK

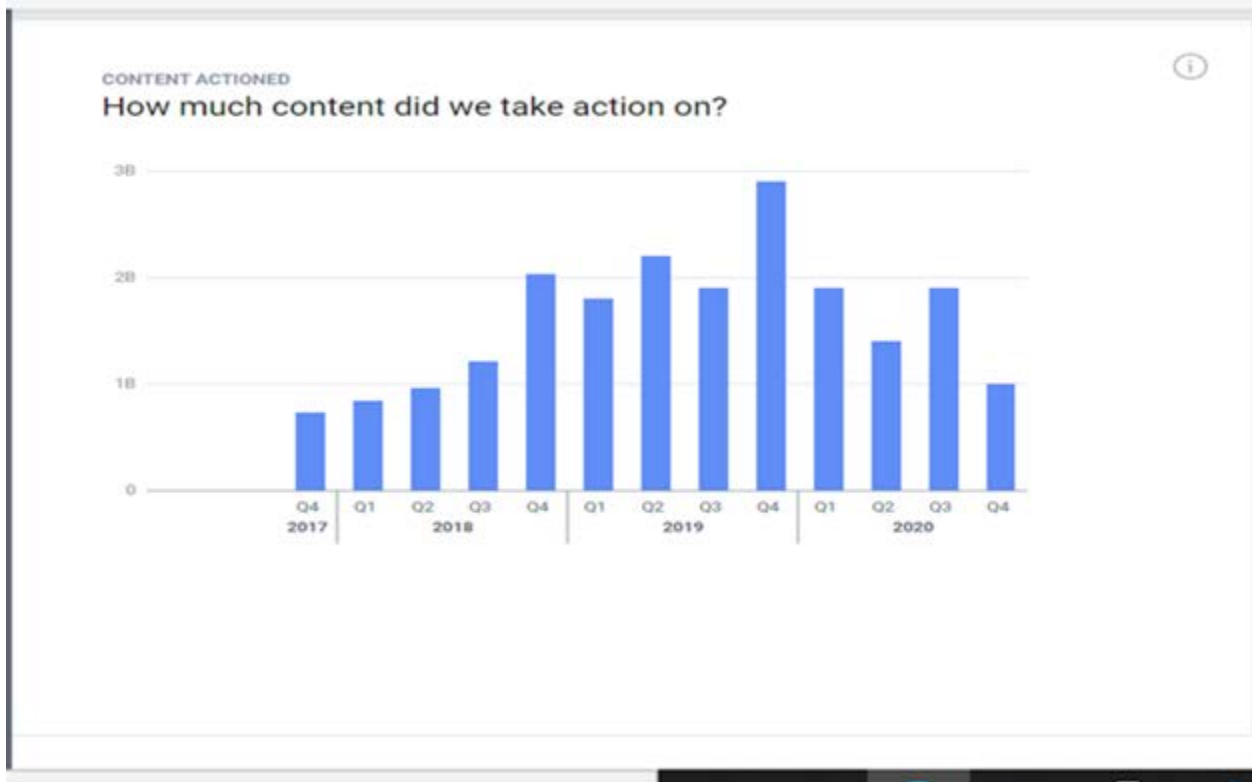
haine et la désinformation. Dans le cadre de cette réflexion, nous empêchons désormais les personnes ayant violé certaines règles de Facebook – notamment notre politique sur les « *Individus et Organismes Dangereux* » – d'utiliser Facebook Live. Nous appliquons également une même politique de suspension du Live pour de nombreuses autres violations. Ainsi, toute personne qui enfreint nos politiques les plus graves se verra interdire l'utilisation de Live durant une période déterminée – par exemple 30 jours – à compter de sa première infraction. Ces restrictions peuvent également être étendues à d'autres domaines, tels que l'interdiction pour l'utilisateur de créer des publicités sur Facebook.

## (v) Concernant les *spam*

Les *spammers* posent des défis de taille car ils développent constamment de nouvelles tactiques pour échapper à nos systèmes de détection. Nous avons alors récemment investi dans des améliorations de notre modèle d'apprentissage automatique, en vue de détecter plus précisément les *spams* et de prendre toutes les mesures qui s'imposent.

Au 1<sup>er</sup> trimestre 2020, les chiffres montrent une forte baisse des contenus qualifiés de *spam* que nous avons retirés par rapport au 4<sup>e</sup> trimestre 2019. Au 3<sup>e</sup> trimestre 2020, nous avons apporté des mises à jour à notre technologie qui nous ont aidés à lutter contre les commentaires de masse, ce qui a augmenté les contenus actionnés au cours de ce trimestre. Toutefois, en raison des fluctuations des volumes et des tendances associées au *spam*, la répression a diminué depuis juillet. Pour continuer à améliorer la précision, nous avons ajusté notre technologie de détection proactive au 4<sup>e</sup> trimestre. En conséquence, les contenus actionnés ont diminué, passant de 1,9 milliard de contenus au 3<sup>e</sup> trimestre 2020 à 1 milliard au 4<sup>e</sup> trimestre 2020. En 2020, il y a eu un total de 2,3 millions de recours et nous avons restauré 560.600 contenus après recours.

# FACEBOOK



En outre, grâce à notre investissement dans une technologie plus performante de détection des *spams*, visant à permettre non seulement une meilleure identification de ceux-ci mais aussi à faciliter l'application des sanctions appropriées, 99,8 % des contenus ayant fait l'objet d'actions pour *spam* ont en réalité été identifiés par Facebook avant même d'être signalés par les utilisateurs.

# FACEBOOK

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 publiés 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.

Il est renvoyé à la Question 3.

## **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...).

Ces discussions visant à harmoniser le dispositif de signalement n'ont pas été mises en place. Cependant, nous avons engagé des discussions avec certains acteurs, notamment pour promouvoir un Internet plus sûr et plus fiable dans le cadre du Partenariat pour la confiance et la sécurité numériques (voir Question 1.3.8) et concernant les sujets abordés dans la Question 1.4 (Vidéos manipulées et trompeuses) qui mentionne les partenaires « industrie » de Facebook, ainsi que dans les Questions 3 et 6 de cette Déclaration 2020.

# FACEBOOK

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.

Dans son bilan de l'application et de l'effectivité des mesures mises en oeuvre par les opérateurs en 2019, le CSA a formulé les recommandations suivantes concernant le dispositif de signalement de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un scrutin (reproduites en *italique*) :

- Les opérateurs, notamment de plateformes d'agrégation de contenus, sont invités à continuer d'améliorer leur outil de signalement en termes **d'accessibilité, de clarté et d'ergonomie**. En particulier, le CSA les invite à simplifier cet outil afin que l'envoi d'un signalement puisse être finalisé en trois clics maximum.

Comme expliqué ci-dessus (voir Question 1.2), nous avons mis en place un dispositif de signalement facilement et directement accessible et visible qui permet aux utilisateurs de signaler des fausses informations. Nous avons également amélioré notre dispositif de signalement afin de permettre à l'utilisateur de préciser si les fausses informations sont liées à des questions de santé, de politique, d'enjeu social ou autre avant de finaliser le signalement. Comme le démontre cette Déclaration 2020, les utilisateurs sont pleinement informés de notre dispositif de signalement et de la façon dont nous traitons les signalements.

- Les plateformes sont incitées à prévoir, dans leur formulaire de signalement, une **catégorie « Fausse information »** ou une catégorie équivalente clairement énoncée et à faire en sorte que leur dispositif de signalement permette aux utilisateurs de fournir des **éléments de contexte** (explications, captures d'écran).

Comme il apparaît dans les captures d'écran ci-dessus, notre dispositif de signalement comprend une catégorie « Fausses informations » (voir Question 1.1).

- En se dotant **d'outils de comptage efficaces**, elles pourraient très utilement établir des données statistiques sûres et exhaustives sur les signalements pour fausses informations et leur traitement.

Nous fournissons des données statistiques efficaces, sûres et exhaustives ainsi que des évaluations de la quantité des contenus sur lesquels nous agissons pour violation de nos

# FACEBOOK

Standards de la communauté grâce à la catégorie des faux comptes et des *spams* ainsi que cela est décrit aux Questions 1.3.8 et 1.4.

- Les plateformes effectuant un contrôle interne des contenus signalés comme des fausses informations sont invitées à se doter de normes et sources de référence, notamment en envisageant des **partenariats avec tiers vérificateurs de faits**. À cet égard, la constitution d'une **banque documentaire publique** de ces documents de référence non confidentiels est à encourager. Elle permettrait, d'une part, aux opérateurs de mutualiser leurs ressources et, d'autre part, aux utilisateurs et à la société civile de pouvoir les étudier.

Comme expliqué ci-dessus, nous travaillons en partenariat avec des partenaires de vérification indépendants depuis des années (voir Question 3). Les partenaires de vérification indépendants jouent un rôle clé dans l'instruction et l'action sur le contenu.

- Les plateformes sont encouragées à procéder à des **évaluations**, en interne, du traitement des signalements et, auprès de leurs utilisateurs, de leur procédure de signalement.

Comme détaillé ci-dessus, nous avons développé des systèmes pour notre auto-évaluation et des systèmes pour évaluer les mesures et les canaux que nous avons mis en œuvre (voir Question 1.3.8).

- L'information des utilisateurs mériterait d'être améliorée sur plusieurs points : la politique suivie par la plateforme à l'égard des publications propageant de fausses informations, les procédures de traitement des signalements, les éléments permettant le contrôle (normes et sources), les partenariats avec les fact-checkers, les voies de recours contre les décisions prises par la plateforme (conditions d'exercice, délai, modalités en interne de traitement du recours), de manière générale et, à l'issue du traitement d'un signalement, à l'attention des concernés.

Nous avons partagé des informations publiques sur tout ce qui précède, ceci de manière récurrente et régulière.

- Les opérateurs sont encouragés à développer et à formaliser une **procédure d'urgence** pour le signalement de fausses informations en cas de circonstances exceptionnelles.

# FACEBOOK

Il est renvoyé à nos explications concernant la COVID-19 ci-dessus (Questions générales et Question 1.3.5). Nous disposons également de canaux spécifiques pour interagir de manière proactive avec les autorités dans des cas spécifiques, notamment en cas d'urgence.

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.

Comme mentionné dans les Questions générales et à la Question 1.1, nous avons adapté nos politiques et apporté divers changements en réponse à la pandémie de la COVID-19.

## 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 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.

Les élections municipales en France n'ont pas modifié nos procédures habituelles en 2020.

Concernant la crise sanitaire, nous avons pris des précautions pour protéger nos équipes afin d'envoyer tous les agents contractuels qui effectuent de la modération de contenus travailler à domicile. Comme il y a moins de personnes disponibles pour modérer les contenus (avec intervention humaine), nous avons continué à accorder la priorité aux préjudices imminents et à nous fier davantage à la détection proactive dans d'autres domaines pour retirer tout contenu contrevenant. Cela n'a pas d'impact notable sur les utilisateurs. Ceci étant dit, cette approche peut présenter certaines limites et nous pourrions avoir des temps de réponse plus longs et commettre des erreurs en conséquence. Il s'agit d'une époque sans précédent, mais la sécurité

# FACEBOOK

et la sûreté de notre service continueront d'être assurées. Pour plus d'informations à ce sujet, voir les actualités de Facebook : <https://about.fb.com/news/2020/12/coronavirus/>

## **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.

Facebook encourage l'utilisation de certains concepts comme la loyauté, l'explicabilité et l'équité dans le cadre des mesures qu'elle met en place.

Facebook s'efforce de définir de bonnes pratiques pour développer une intelligence artificielle fiable, grâce au travail interne de son équipe dédiée à une Intelligence Artificielle (IA) Responsable, qui développe de nouveaux outils et processus pour améliorer davantage la loyauté et l'explicabilité de certains systèmes d'IA (nous pouvons notamment citer comme exemples les outils « *Pourquoi je vois cette publication ?* » (voir Question 2.7) et Fairness Flow (voir Question 2.6) ; Facebook a participé, aux côtés de la Commission et de nombreux États membres de l'Union Européenne, au groupe d'experts qui a aidé à l'élaboration des Principes relatifs à l'IA de l'OCDE, et nous travaillons actuellement, via l'Observatoire OCDE des politiques relatives à l'IA, à la mise en œuvre de ces principes. Facebook pilote également de nouvelles approches innovantes en matière de « bac à sable » réglementaire et de prototypage des politiques, y compris une collaboration actuelle avec [l'Autorité singapourienne de développement des médias Infocomm](#) pour tester de nouvelles idées sur l'explicabilité de l'IA, et nous serions heureux de participer à des collaborations publiques-privées similaires avec les décideurs politiques de l'Union Européenne.

Chez Facebook, nous considérons la loyauté comme un processus, et non simplement comme une propriété d'un système. Dans le cadre de l'objectif à long terme de Facebook en matière d'intelligence artificielle responsable, nous avons [publié des recherches](#) sur une approche holistique permettant de traiter les questions de loyauté dans un contexte réel de systèmes de

# FACEBOOK

production à grande échelle. Cette approche, qui est toujours en cours d'évaluation et de développement, examine la loyauté sous de multiples aspects connexes : au niveau du produit, de la politique et de la mise en œuvre.

La loyauté, au niveau du produit, implique de poser des questions telles que : « Les objectifs de ce produit sont-ils compatibles avec le fait de fournir aux personnes une valeur juste et de les traiter de manière loyale ? » Au niveau de la politique, il s'agit de poser des questions telles que : « Une politique interdisant certains types de comportements nuisibles au sein du produit répond-elle de manière adéquate aux préjudices uniques subis par certaines sous-populations ? ». Au niveau de la mise en œuvre, les questions de loyauté portent sur des détails plus techniques concernant les performances d'un système. Par exemple, « les personnes chargées d'étiqueter les ensembles de données exécutent-elles correctement les instructions relatives à l'étiquetage ? » et « les modèles prédictifs fonctionnent-ils aussi bien pour toutes les sous-populations ou génèrent-ils des erreurs qui affectent certains groupes plus que d'autres ? ».

En d'autres termes, une approche holistique de la loyauté permet de déterminer non seulement si les règles algorithmiques sont appliquées de manière appropriée à tous, mais également si les règles elles-mêmes - ou la structure dans laquelle elles se situent - sont loyales, justes et raisonnables.

Nous avons également créé un groupe de produits dédié, l'équipe Instagram Équité, qui se concentre sur une meilleure compréhension et la résolution des biais dans le développement de nos produits et des expériences des personnes sur Instagram. L'équipe Équité se concentrera sur la création de produits justes et équitables, ce qui comprend le travail avec notre équipe d'IA responsable afin de comprendre et de résoudre les problèmes potentiels d'équité algorithmique. L'équipe a également lancé la Carte de modèle d'apprentissage automatique, un modèle qui aide à s'assurer que l'apprentissage automatique est conçu en tenant compte de l'équité, et les a étendus à bon nombre de nos modèles d'intégrité.

Pour plus d'informations à ce sujet, voir <https://ai.facebook.com/blog/collaborating-on-the-future-of-ai-governance-in-the-eu-and-around-the-world/> et <https://about.fb.com/news/2020/09/an-update-on-our-equity-work/>



# FACEBOOK

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.

# FACEBOOK

## 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.

**Un des principaux systèmes de classement de contenus**, le fil d'actualité, détermine quelles publications apparaissent dans votre fil d'actualité, et dans quel ordre, en prédisant ce qui est le plus susceptible d'intéresser ou de faire participer les utilisateurs. Ces prédictions sont basées sur divers facteurs, y compris ce que les utilisateurs ont suivi, aimé ou ce à quoi ils ont participé récemment. Pour comprendre comment cela fonctionne en pratique, veuillez regarder <https://www.facebook.com/facebook/videos/206832444517607/>.

Le classement est le processus que nous utilisons pour organiser toutes ces publications. Le classement comporte généralement quatre éléments détaillés ci-dessous : l'inventaire des publications ; les signaux et les points de données qui peuvent éclairer les décisions de classement ; les prédictions que nous faisons ; et un indice de pertinence pour chaque publication. Pour en savoir plus sur notre algorithme de classement, veuillez consulter <https://about.fb.com/news/2018/05/inside-feed-news-feed-ranking/> et <https://www.facebook.com/business/help/718033381901819?id=208060977200861>.

Pour classer plus d'un millier de publications par utilisateur, par jour, pour plus de 2 milliards de d'utilisateurs, en temps réel, le processus doit être efficace. Nous procédons pour cela en plusieurs étapes stratégiquement organisées pour rendre le processus rapide et limiter les ressources de calcul nécessaires.

Dans un premier temps, le système recense toutes les publications possibles qui pourraient être classées pour un utilisateur. Cet inventaire éligible inclut toute publication partagée avec l'utilisateur par un ami, un groupe ou une page qu'il suit depuis sa dernière connexion et qui n'a pas été supprimée.

Pour s'assurer que les publications qui n'ont pas encore été vues soient prises en compte, nous appliquons une logique de remontée du contenu non lu : les publications récentes qui ont été classées pour l'utilisateur, mais qu'il n'a pas encore vues lors de ses connexions antérieures, sont aussi ajoutées à l'inventaire éligible pour cette connexion. Nous appliquons également une logique de remontée en fonction des actions de manière à ce que toute publication que

# FACEBOOK

l'utilisateur a déjà vue et qui a depuis généré une conversation intéressante entre ses amis soit aussi ajoutée à l'inventaire éligible.

Le système doit ensuite calculer le score de chaque publication à partir de divers facteurs, tels que le type de publication, la similitude avec d'autres éléments, ou encore la mesure dans laquelle la publication correspond à ce avec quoi l'utilisateur a tendance à interagir. Pour calculer ce score pour plus de 1 000 publications, pour chacun des milliards d'utilisateurs et en temps réel, nous exécutons ces modèles pour toutes les publications possibles en parallèle sur plusieurs machines dites de prévision.

Avant de combiner toutes ces prévisions pour obtenir un score unique, nous devons appliquer des règles supplémentaires. Nous attendons d'avoir ces premières prévisions pour pouvoir réduire le nombre de publications à classer, et nous les appliquons sur plusieurs passages pour économiser la puissance de calcul.

Tout d'abord, des processus d'intégrité sont appliqués à chaque publication. Ils visent à déterminer quelles mesures de détection de l'intégrité, le cas échéant, doivent être appliquées aux articles sélectionnés pour le classement. Lors du passage suivant, un modèle léger réduit le nombre de candidats à environ 500 publications parmi les plus pertinentes pour l'utilisateur. Le fait de classer moins de publications nous permet d'utiliser des modèles de réseaux de neurones plus puissants pour les prochains passages.

Ensuite, il y a le passage principal de la notation, au cours duquel se fait l'essentiel de la personnalisation. Ici, un score pour chaque histoire est calculé indépendamment, puis les 500 publications sont classées par ordre de score. Pour certaines, le score peut être plus élevé pour les Likes que pour les commentaires, car certaines personnes aiment s'exprimer davantage ainsi. Toute action qu'une personne entreprend rarement, comme par exemple, une prévision de Like très proche de zéro, joue automatiquement un rôle minime dans le classement, car la valeur prédite est très faible.

Enfin, nous effectuons un passage contextuel, au cours duquel des caractéristiques contextuelles, comme les règles de diversité des types de contenu, sont ajoutées pour s'assurer que le fil d'actualité de l'utilisateur offre un mélange équilibré de types de contenu, et éviter qu'il ne voit plusieurs vidéos à la suite les unes des autres. Toutes ces étapes de classement ne prennent que le temps nécessaire à l'utilisateur pour ouvrir l'application Facebook. Et en l'espace de quelques

# FACEBOOK

secondes seulement, il peut profiter d'un fil d'actualité dont le contenu a été spécialement conçu pour lui et l'explorer comme bon lui semble.

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) ?

Il est renvoyé à nos remarques générales concernant les systèmes de modération de contenus à la Question 1.3.3 de cette Déclaration 2020.

Nos outils d'intelligence artificielle signalent les problèmes potentiels à examiner et trouvent automatiquement de nouveaux cas de fausses informations précédemment identifiés.

Pour mieux appliquer les étiquettes de mise en garde à grande échelle, nous devons mettre au point de nouvelles technologies d'IA pour s'adapter aux quasi-duplications de fausses informations.

- Facebook a déployé « SimSearchNet++ », un logiciel d'intelligence artificielle identifiant et regroupant les différentes variantes d'une même image avec un très haut degré de précision. Cet outil est déployé dans le cadre de notre système d'indexation et de comparaison d'images de bout en bout, et fonctionne aussi bien sur Facebook que sur Instagram. SimSearchNet++ fonctionne sur de nombreuses variétés de manipulations d'images, telles que les recadrages, les flous et les captures d'écran – ce qui est particulièrement important pour une plateforme axée sur le visuel comme Instagram. SimSearchNet++ est particulièrement efficace pour faire correspondre des images similaires, ce qui nous permet d'anticiper efficacement ces correspondances. SimSearchNet++ fonctionne aussi sur les images avec du texte, grâce à un système de vérification par reconnaissance optique de caractères (OCR) qui permet de faire correspondre de telles images avec un haut degré de précision. SimSearchNet++ permet de supprimer les images ainsi identifiées de manière extrêmement précise, rendant ainsi cet outil parfaitement apte à détecter des cas de fausses informations, tout en

# FACEBOOK

déclenchant quelques faux positifs. Il est également très efficace pour détecter et faire correspondre des montages de fausses informations.

- La détection des fausses informations à grande échelle pose un autre problème : les fausses déclarations peuvent apparaître dans d'innombrables variantes au fil du temps. Nous avons développé un ensemble de systèmes pour prédire quand deux éléments de contenus ont la même signification même s'ils ont l'air très différents. (Par exemple, ils peuvent avoir des légendes avec un texte complètement différent qui fait la même déclaration, comme « *les masques sont dangereux* » et « *se couvrir le visage n'est pas sûr* ». Ou ils peuvent avoir des images différentes qui montrent le même sujet.).
- Nous avons introduit de nouveaux systèmes d'IA pour détecter automatiquement les nouvelles variations de contenus que des partenaires de vérification indépendants ont déjà démenties. Une fois que l'IA a repéré ces nouvelles variations, nous pouvons les signaler à nos partenaires de vérification indépendants pour qu'ils les examinent. Ce système s'appuie sur plusieurs technologies, dont ObjectDNA, qui est basé sur la recherche IA de Facebook. Contrairement aux outils de vision informatique classiques, qui examinent l'ensemble de l'image afin d'en comprendre le contenu, ObjectDNA se concentre sur des objets clés spécifiques de l'image tout en ignorant l'encombrement de l'arrière-plan. Cela nous permet de trouver des reproductions de la déclaration qui utilisent des morceaux d'une image que nous avons déjà signalée, même si les images globales sont très différentes les unes des autres. Notre IA utilise également l'intégration de phrases multilingues LASER, développée par les chercheurs en IA de Facebook. Cette technique intègre plusieurs langues conjointement dans un seul espace d'intégration partagé, ce qui nous permet d'évaluer plus précisément la similitude sémantique des phrases. Elle exploite à la fois le texte et les images, mais fonctionne également pour les contenus qui ne se composent que de l'un ou l'autre.
- Nous avons également pris des mesures pour lutter contre les *deepfakes*. Ces vidéos, qui utilisent l'IA pour montrer des personnes faisant et disant des choses qu'elles n'ont pas réellement faites ou dites, peuvent être difficiles à repérer, même pour un modérateur entraîné. En plus de demander à notre [Équipe rouge IA](#) d'anticiper les problèmes potentiels, nous avons déployé un modèle de détection des *deepfakes* à la pointe de la technologie avec huit réseaux neuronaux profonds, chacun utilisant l'épine dorsale EfficientNet. Il a été entraîné sur des vidéos à partir d'un ensemble de données unique

# FACEBOOK

commandé par Facebook pour le *Deepfake Detection Challenge*. Afin d'identifier de nouvelles vidéos de *deepfakes* que nos systèmes n'ont jamais vues auparavant, nous utilisons une nouvelle technique de synthèse de données pour mettre à jour nos modèles en temps quasi réel. Lorsqu'une nouvelle vidéo de *deepfakes* est détectée, nous générons de nouveaux exemples similaires de *deepfakes* pour servir de données d'entraînement à grande échelle pour notre modèle de détection de *deepfakes*. Nous appelons cette méthode GAN des GAN (GoG) car elle génère des exemples en utilisant des réseaux adversaires génératifs (GAN), une architecture d'apprentissage automatique où deux réseaux neuronaux se font concurrence pour devenir plus précis. La méthode GoG nous permet d'affiner continuellement notre système afin qu'il soit plus robuste et plus généralisé pour faire face à d'éventuels futurs *deepfakes*.

Pour plus de détails à cet égard, voir l'article suivant, qui détaille notre utilisation de l'IA pour lutter contre les fausses informations : <https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/> et Facebook IA <https://ai.facebook.com/>

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.

Comme expliqué à la Question 2.2 ci-dessus, l'équipe d'IA développe de nouveaux outils et processus pour améliorer davantage la loyauté et l'explicabilité de certains systèmes d'IA, notamment au moyen d'outils tels que Fairness Flow, qui peut aider à générer des indicateurs pour évaluer s'il y a des biais involontaires dans certains modèles.

En outre, nous développons des [partenariats](#), tels que la création d'un centre de recherche indépendant sur l'éthique de l'IA avec l'Université technique de Munich (TUM). L'institut d'Éthique de l'Intelligence Artificielle, qui bénéficie d'une subvention initiale de Facebook de 7,5 millions de dollars sur cinq ans, contribuera à faire progresser le domaine croissant de la recherche éthique sur les nouvelles technologies et explorera les questions fondamentales qui affectent l'utilisation et l'impact de l'IA. Le premier rapport de l'institut TUM pour l'Éthique de l'IA a été publié en 2020 et est disponible ici : <https://ieai.mcts.tum.de/wp-content/uploads/2021/03/IEAI-Annual-Report-2020.pdf>

# FACEBOOK

## *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.).

Comme expliqué ci-dessus, Facebook utilise des algorithmes pour une plus grande transparence et un meilleur contrôle de ses services.

### 2.7.1 Plus grande transparence et meilleur contrôle



Que ce soit pour les nouveaux utilisateurs ou pour ceux qui utilisent Facebook depuis plusieurs années, nous faisons en sorte que tous puissent facilement comprendre et paramétrer la façon dont leurs informations ont une influence sur les publicités et publications qu'ils voient. C'est la raison pour laquelle nous avons introduit des outils et continuons à ajouter davantage

# FACEBOOK

d'informations et de fonctionnalités, afin que nos utilisateurs puissent avoir davantage de contexte et de contrôle sur ce qu'ils voient sur Facebook.

(a) a. Activité sur Facebook

(i) i. Transparence des contenus publicitaires

Nous avons créé un outil intitulé « *Pourquoi je vois cette publicité ?* » accessible dans le menu déroulant en cliquant sur  en haut à droite de chaque publicité (voir capture d'écran ci-dessous ce qui apparaît après avoir cliqué sur  ) :

---

- Masquer la publicité**  
Ne plus jamais voir cette publicité
- Signaler la publicité**  
Signalez-nous tout problème avec cette pub
- Enregistrer le lien**  
Ajoutez à vos éléments enregistrés
- Pourquoi je vois cette publicité ?**

---

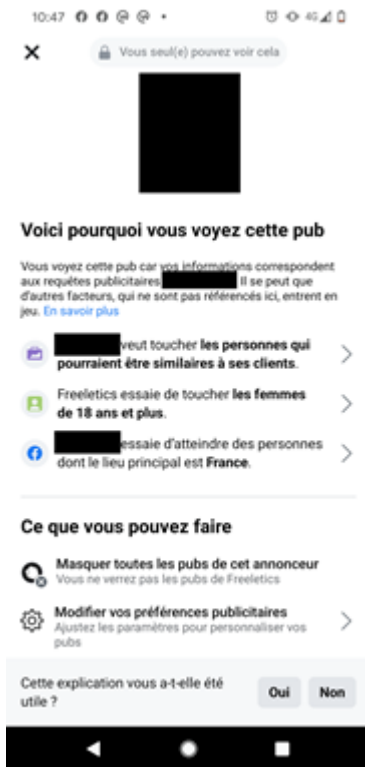
∨ Plus d'options

---



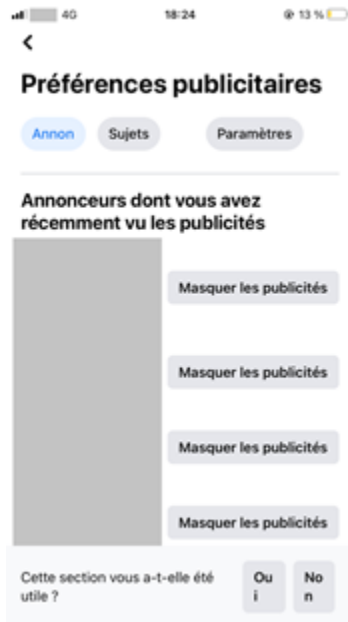
# FACEBOOK

En cliquant sur « *Pourquoi je vois cette publicité?* », l'utilisateur peut avoir des informations sur les raisons du ciblage comme par exemple, les informations démographiques, visite d'un site internet, les intérêts ou catégories qui les ont associés à une publicité spécifique (voir capture d'écran ci-dessous).



# FACEBOOK


Dans la section « *Ce que vous pouvez faire* » (voir capture d'écran ci-dessus), les utilisateurs peuvent également choisir de masquer toutes les publicités de cet annonceur ou de modifier leurs préférences publicitaires (Voir capture d'écran ci-dessous montrant ce l'utilisateur voit après avoir cliqué sur « *modifier vos préférences publicitaires* ») :



Enfin, la fonctionnalité « [Accéder à la bibliothèque publicitaire](#) » disponible sur Facebook et Instagram permet aux utilisateurs de voir toutes les publicités diffusées sur les produits Facebook, y compris Instagram.

# FACEBOOK

## (ii) Transparence des contenus organiques

Les utilisateurs peuvent mieux comprendre et mieux contrôler ce qu'ils voient de leurs amis, pages et groupes dans leur fil d'actualité via l'outil « *Pourquoi est-ce que je vois cette publication ?* », lequel se trouve dans le menu déroulant qui apparaît après avoir cliqué sur  à l'angle en haut à droite d'une publication (voir capture d'écran ci-dessous).



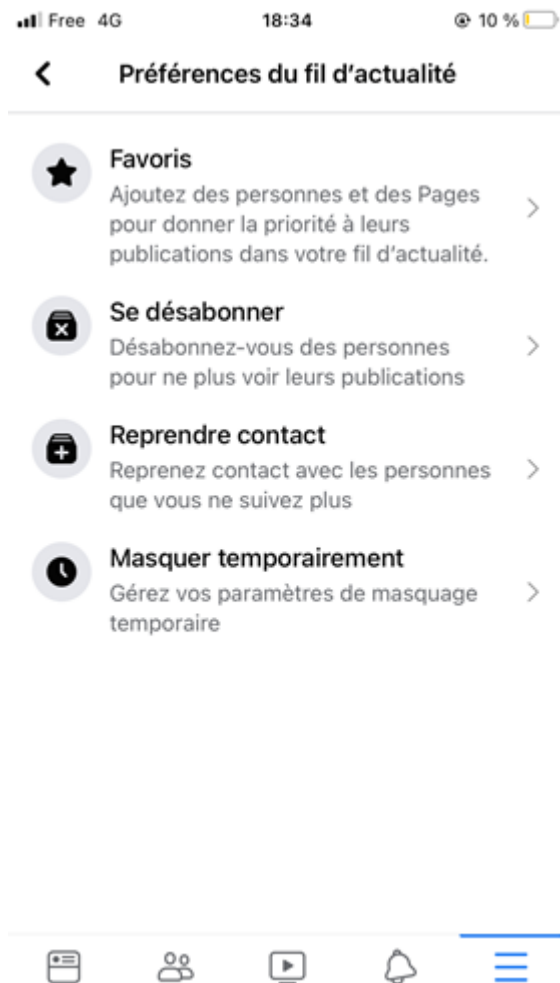
# FACEBOOK

Voir la capture d'écran ci-dessous montrant ce qui apparaît après avoir cliqué sur « Pourquoi est-ce que je vois cette publication ? » :



# FACEBOOK

En cliquant sur « *Modifier les préférences de votre fil d'actualité* » (voir ci-dessus), l'utilisateur peut personnaliser et contrôler ce qu'ils voient dans son fil d'actualité (voir capture d'écran ci-dessous montrant ce qui apparaît après avoir cliqué sur « *Modifier les préférences de votre fil d'actualité* ») :



# FACEBOOK

L'utilisateur peut également contrôler ce qu'il voit dans le fil d'actualité en cliquant sur les options, par exemple « *Masquer la publication* » figurant dans le menu déroulant après avoir cliqué sur les trois boutons en haut à droite de chaque publication (voir capture d'écran ci-dessous) :



## (b) Activité en dehors de Facebook

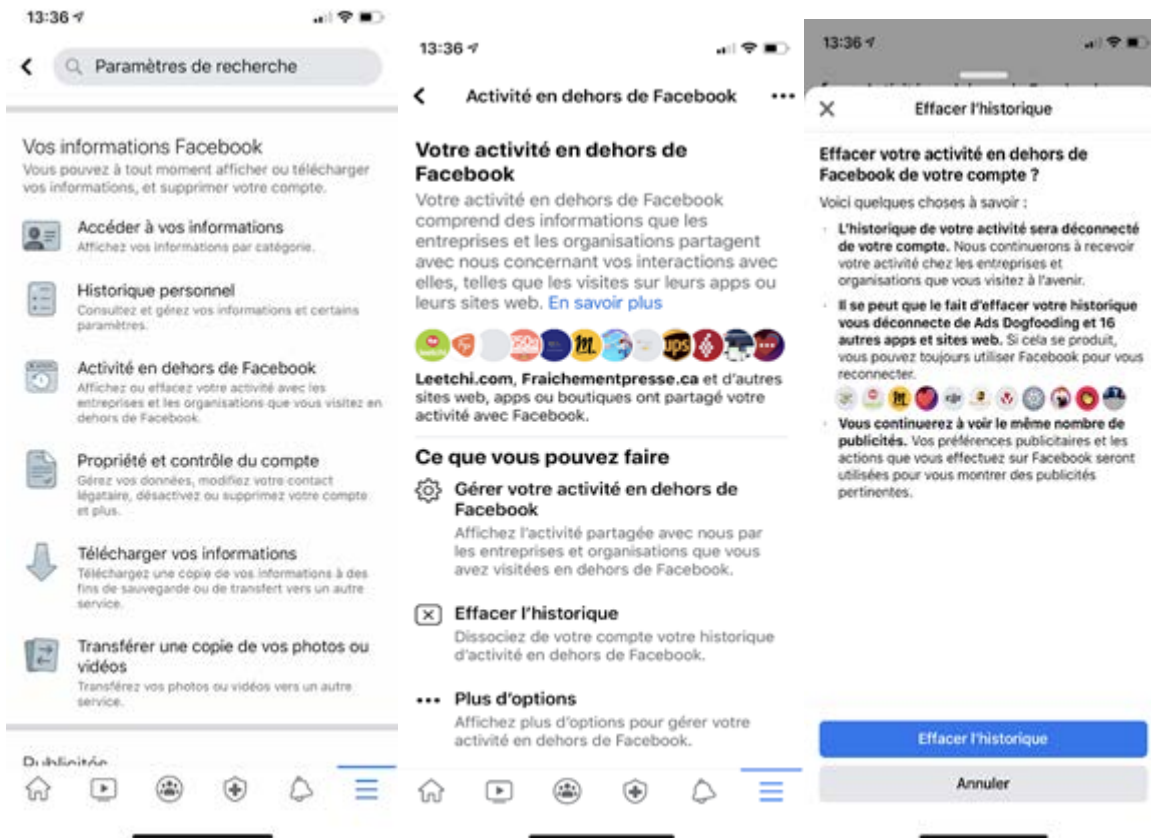
De même, à l'aide de l'outil Activité en dehors de Facebook, les utilisateurs peuvent déconnecter l'activité en dehors de Facebook qui a été associée à leur compte Facebook, sous réserve d'une certaine conservation pour des besoins de sécurité et de sûreté.

Comme illustré ci-dessous, les utilisateurs peuvent voir un résumé des données d'activité qui ont été partagées avec Facebook par des applications et des sites internet ; déconnecter les données d'activité tierces historiques de leur compte à l'aide de l'option « *Effacer l'historique* » ; et gérer si les données d'activité futures doivent être déconnectées de leur compte à l'avenir à l'aide de l'option « *Gérer l'activité future* ».

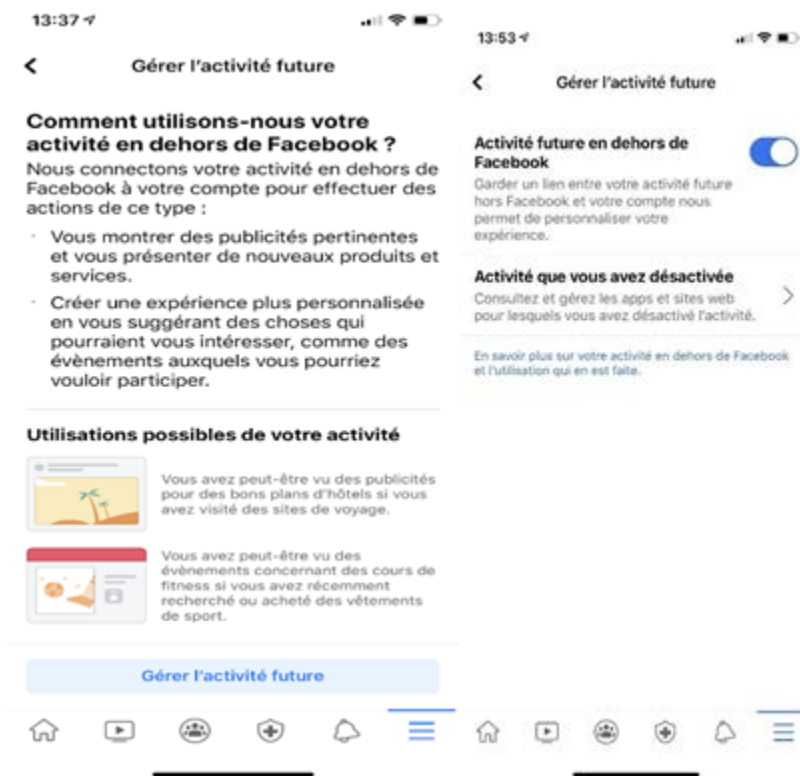
# FACEBOOK

L'outil Activité en dehors de Facebook est disponible dans le monde entier. Sur mobile (voir captures d'écran ci-dessous), cet outil est accessible lorsque l'utilisateur clique sur « Paramètres de confidentialité », puis « paramètres », puis « Activité en dehors de Facebook ».

Voir les captures d'écran de l'outil ci-dessous (sur téléphone mobile) :



# FACEBOOK



Les utilisateurs sont informés de ces systèmes de recommandation qui personnalisent les recommandations des utilisateurs des services Facebook et Instagram et les mises à jour de ces systèmes, notamment par des mises à jour sur les actualités Facebook et dans les Pages d'aide.

Pour des exemples à ce sujet, veuillez consulter : <https://about.fb.com/fr/news/2021/02/comment-le-fil-dactualite-de-facebook-prevoit-il-ce-que-vous-souhaitez-voir/> et <https://www.facebook.com/help/1257205004624246> <https://help.instagram.com/313829416281232>.

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.



# FACEBOOK

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 ?

Il est renvoyé à la Question 2.5 ci-dessus, qui comprend des informations mises à la disposition de tous les utilisateurs sur les systèmes de modération via les actualités Facebook (<https://about.fb.com/fr/news/>).

Les utilisateurs sont informés des mises à jour de ces systèmes de modération qui affectent de manière significative la manière dont le contenu est modéré, notamment par des mises à jour via les actualités Facebook et les Pages d'aide. Voir par exemple: <https://about.fb.com/news/2020/04/covid-19-misinfo-update> ; <https://www.intern.facebook.com/help/230764881494641>. Les utilisateurs sont également informés des résultats de la modération de contenus dans notre CSER.

# FACEBOOK

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

2.9 L'opérateur décrira **tout système qu'il a développé ou utilisé 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).

Dans le cadre de nos signalements réguliers de CIB, nous partageons des informations sur tous les réseaux que nous supprimons au cours d'un mois afin de permettre aux gens de voir plus facilement les progrès que nous faisons en un seul endroit.

Nous nous efforçons de mettre fin à deux types d'activités : 1) les comportements inauthentiques coordonnés dans le cadre de campagnes nationales non gouvernementales et 2) les comportements inauthentiques coordonnés au nom d'un acteur étranger ou gouvernemental.

Comportement inauthentique coordonné : Lorsque nous constatons que des campagnes nationales non gouvernementales incluent des groupes de comptes et des pages cherchant à tromper les gens sur leur identité et leurs activités tout en s'appuyant sur de faux comptes, nous supprimons les comptes, les pages et les groupes inauthentiques et authentiques directement impliqués dans cette activité.

Ingérence étrangère ou gouvernementale : Si nous découvrons des cas de CIB menés au nom d'une entité gouvernementale ou par un acteur étranger, nous appliquons les mesures d'application les plus larges, y compris le retrait de tous les biens sur la plateforme liés à l'opération elle-même et aux personnes et organisations qui la soutiennent.

Mise en œuvre continue : Nous surveillons les efforts déployés par les réseaux que nous avons précédemment supprimés visant à rétablir leur présence sur Facebook. Grâce à la détection automatisée et manuelle, nous retirons continuellement les comptes et les pages liés aux réseaux que nous avons supprimés par le passé.

Nos équipes continuent de se concentrer sur la recherche et la suppression des campagnes trompeuses dans le monde entier, tant nationales qu'à l'étranger. Résumé des constatations de [février 2021](#) :

# FACEBOOK

- Nombre total de comptes Facebook supprimés : 915
- Nombre total de comptes Instagram supprimés : 606
- Nombre total de pages supprimées : 86
- Nombre total de groupes supprimés : 21

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.

Dans son bilan de l'application et de l'effectivité des mesures mises en oeuvre par les opérateurs en 2019, le CSA a formulé les recommandations suivantes concernant la transparence des algorithmes (reproduit en *italique*) :

- Dans le cadre des futures déclarations, il souhaite que les opérateurs lui communiquent davantage d'éléments sur **l'intelligibilité et la redevabilité**, de leurs algorithmes d'organisation, de sélection et d'ordonnement des contenus. Il les invite également à lui transmettre les principes de ces approches tout comme les listes des critères et données utilisés par ces algorithmes. (...).

Comme nous l'avons expliqué en détail à la Question 2, nous fournissons au CSA de nombreuses informations sur les principes appliqués à nos algorithmes, notamment une vidéo explicative de l'ordonnement du fil d'actualité

- **Le CSA recommande une transparence accrue en matière de fonctionnement des algorithmes vis-à-vis des utilisateurs.** (...) les opérateurs devraient être proactifs pour proposer ces informations nécessaires à la compréhension et à la bonne utilisation de leurs services.

Facebook utilise des algorithmes pour une plus grande transparence et un meilleur contrôle de ses services et nous nous assurons que tous les utilisateurs sont informés et peuvent facilement comprendre ce qu'ils voient sur nos services. En effet, les utilisateurs sont immédiatement informés dans le menu déroulant situé dans le coin droit d'une publication/publicité en plus de nos politiques et de nos actualités (voir les Questions 2.4, 2.5 et 2.7).

# FACEBOOK

- **Le CSA recommande également transparence accrue en matière de capacité de réglage des algorithmes.** (...) Une attention doit également être portée à l'ergonomie afin de s'assurer que ce type de réglages est facilement accessible et ce en un minimum d'actions de la part de l'utilisateur.

Nous avons mis en place des outils et continuons d'ajouter plus d'informations et de fonctionnalités afin que les personnes puissent avoir plus d'informations et de contrôle sur ce qu'elles voient sur Facebook. Tous les utilisateurs peuvent déjà facilement comprendre et ajuster comment leurs informations influencent les publicités et les publications qu'ils voient (voir Question 2.7).

- Enfin, (...), une analyse dynamique des pratiques des opérateurs semble montrer un **recours grandissant aux algorithmes, notamment pour appuyer le travail de modération** des contenus. Cette tendance aurait été amplifiée par les contraintes sanitaires liées à la COVID-19, (...) le CSA invite les opérateurs à lui **fournir toutes les informations utiles sur les rôles et degrés respectifs de la part de curation humaine et de celle des algorithmes dans la détection d'une part, et le traitement d'autre part, des fausses informations** dans leur prochaine déclaration annuelle.

Il est renvoyé aux Questions 1.3.3 et 2.8.

### **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.

#### ***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.

# FACEBOOK

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) ;

Nous utilisons le modèle d'apprentissage automatique pour développer notre classificateur d'actualités, un modèle qui examine les publications individuelles sur Facebook et qui détermine si le contenu d'une publication correspond ou non à une actualité. Pour former un modèle d'apprentissage, les modérateurs créent des exemples de contenu que le classificateur essaie d'identifier (exemples positifs) et n'essaie pas d'identifier (exemples négatifs), afin que le modèle apprenne à faire la différence entre les deux.

Cette méthode d'identification des actualités au niveau du contenu est différente de l'Index des pages d'actualités défini ci-dessous.

Pour identifier les pages Facebook qui publient principalement du contenu d'actualités sur la plateforme, nous avons créé un index appelé Index des pages d'actualités. Toute page Facebook venant principalement de journalistes devrait être enregistrée en tant que page d'actualités sur Facebook dans l'Index des pages d'actualités.

Les éditeurs peuvent demander que leurs pages soient incluses dans l'Index des pages d'actualités. Pour être inclus, ils doivent se conformer aux règles que nous avons élaborées avec des organismes de presse, des universitaires et des organisations industrielles. Par exemple, les pages doivent se concentrer principalement sur les événements d'actualités et d'information, citer des sources et inclure des dates. Les pages qui publient principalement des publicités ou du contenu promotionnel ne sont pas admissibles. Les pages ne peuvent pas avoir été enregistrées pour avoir publié des articles qui ont été jugés faux par nos partenaires de vérification indépendants. Les pages incluses ne seront pas différentes sur Facebook, et l'inclusion n'aura pas d'impact sur le statut « vérifié » de la page (marque bleue).

Nous continuerons de travailler avec des organismes de presse, des universitaires et des organisations professionnelles pour explorer d'autres façons d'utiliser l'Index, notamment en identifiant les pages éligibles aux produits et aux programmes dédiés aux médias d'actualité.

# FACEBOOK

- **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 ;

Comme mentionné ci-dessus, au niveau du contenu, les modérateurs répondent à des questions précises afin d'évaluer si un contenu donné est une actualité. Ces questions et ces règles ont été développées en collaboration avec des spécialistes du journalisme et incluent des critères tels que : Ce contenu relate-t-il des événements actuels, des informations actuelles ou des enquêtes en cours, ou s'agit-il d'un rapport de la rédaction ou d'un rapport d'opinion ? Ce contenu est-il directement attribué à un auteur, à un journaliste ou à un créateur cité ? Ce contenu cite-t-il des sources pour appuyer ses affirmations ? Ce contenu fait-il preuve de transparence éditoriale (en incluant le nom complet de l'auteur ou des auteurs, des crédits ou un annuaire du personnel, par exemple) ? Ce contenu comprend-t-il des dates et / ou des horodatages ?

Si les modérateurs répondent à toutes les questions par l'affirmative, le contenu examiné est identifié comme une actualité. Les identifications effectuées par nos modérateurs alimentent le modèle d'apprentissage et entraînent le modèle à prédire automatiquement si un contenu est une actualité parmi les énormes quantités d'autres contenus sur Facebook.

Pour être inscrites à l'Index des pages d'actualités, les pages doivent suivre un ensemble de règles que nous avons élaborées avec des organismes de presse, des universitaires et des organisations industrielles :

**Authenticité et responsabilité** : Pour être éligibles à l'inclusion dans l'Index des pages d'actualités, les pages doivent être actives au cours des 90 derniers jours, avoir existé pendant au moins 90 jours et doivent être liées à un site internet fonctionnel avec un nom de [domaine vérifié](#).

Le nom de la page de l'éditeur doit avoir un lien avec son site internet ou l'entreprise qui la détient. De plus, les informations fournies par l'éditeur dans la section « À propos » de sa page peuvent être évaluées pour leur exactitude et vérifiées par rapport à leur statut de fournisseur d'actualités avec une équipe éditoriale composée de plusieurs personnes. En outre, la majorité des liens affichés sur la page de l'éditeur doivent provenir du site internet vérifié pour l'enregistrement des pages d'actualités.

# FACEBOOK

Pour protéger l'intégrité de notre service, nous attendons des éditeurs qui soumettent leurs pages d'actualités qu'ils confirment quelle société les détient et qu'ils se conforment aux [Règles générales applicables aux Pages](#), aux [Standards de la communauté](#) et aux [Conditions générales de la plate-forme](#) Facebook. Les pages qui choisissent de faire de la publicité et de monétiser doivent également se conformer à nos [Règles publicitaires](#) et [Règles de monétisation](#).

**Conditions :** L'Index des pages d'actualités est actuellement conçu comme un outil pour les sociétés éditrices d'actualités, y compris les journaux, les e-magazines et les émissions de télévision d'actualités, et non pour les particuliers. L'Index des pages d'actualités n'accepte pas les candidatures de pages dédiées aux journalistes individuels. Le contenu des candidats doit refléter le jugement d'une équipe de rédaction composée de plusieurs personnes.

**Contenu d'actualités :** Nous savons que les actualités couvrent un large éventail de sujets et peuvent être communiquées dans différents formats et styles. Selon les normes de l'industrie, nous nous attendons à ce que les éditeurs soumettent des pages d'actualités qui :

- Créent principalement du journalisme qui rapporte des événements d'actualités ou d'information en temps utile, et n'est pas de la parodie ou de la satire.
- Citent et/ou renvoient à des sources factuelles d'informations publiées. Les éditeurs qui citent la parodie et la satire comme des faits, ou qui citent à plusieurs reprises des éditeurs connus de fausses informations et de canulars, ne satisferont pas à cette exigence.
- Incluent les dates/l'horodatage sur le contenu publié.
- Publient principalement du contenu qui n'est pas généré par l'utilisateur ou agrégé à partir d'autres sites Internet.
- Fournissent des informations transparentes sur les rédacteurs ou les éditeurs (par ex., les noms complets des rédacteurs dans les biographies, ou des rédacteurs et des éditeurs dans un répertoire du personnel).

Il sera également tenu compte de l'environnement de liberté de la presse dans lequel opère un éditeur.

**Fausses informations :** Les pages et les éditeurs qui ont récemment et à plusieurs reprises partagé de fausses informations ne pourront pas s'inscrire comme pages d'actualités. Si nos

# FACEBOOK

partenaires de vérification indépendants trouvent qu'une page d'actualités enregistrée ou un éditeur partage à plusieurs reprises des fausses informations, l'enregistrement pour cette page peut être révoqué. Les éditeurs disposant de pages associées qui partagent de manière répétée de fausses informations peuvent également faire face à des privilèges restreints associés à l'enregistrement des pages d'actualités.

**Publicité et contenus sponsorisés** : Nous reconnaissons que les éditeurs d'actualités créent et publient une large gamme de contenus, y compris de la publicité et des promotions. Toutefois, la quantité de contenus publicitaires et promotionnels ne doit pas excéder la quantité de contenus d'actualités.

Dans un souci de transparence, nous exigeons également des éditeurs d'actualités enregistrés qu'ils distinguent clairement les contenus d'actualités du contenu d'affiliation, de promotion, de publicité et de marketing, afin que les utilisateurs puissent facilement faire la différence.

- 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 ;

Nous souhaitons permettre aux utilisateurs de mieux comprendre la provenance des contenus informationnels qu'ils visualisent sur Facebook afin qu'ils puissent faire des choix éclairés sur les contenus qu'ils lisent. Le [bouton contexte](#) de Facebook est conçu pour fournir plus de contexte et d'informations sur les éditeurs et les liens qui apparaissent dans le fil d'actualité afin que les gens puissent décider eux-mêmes de ce qu'ils veulent lire, croire et partager.

Le bouton contexte apparaît sous la forme d'un « *i* » dans le coin inférieur droit d'un article, afin de faciliter l'affichage de plus d'informations sur les sites internet et les éditeurs qui publient sur Facebook. Ce bouton permet ainsi d'obtenir des informations sur le site internet ou l'éditeur à l'origine de la publication d'un article, notamment sur la ligne éditoriale suivie par ce dernier, le nom de domaine de son site internet et sa date d'enregistrement, et inclut des liens vers d'autres articles publiés par ce site internet ou éditeur. Nous avons étendu le bouton Contexte aux images contenant des informations ayant été vérifiées sur Facebook.



# FACEBOOK

Voir capture d'écran ci-dessous :



Voir la capture d'écran ci-dessous montrant ce que l'utilisateur peut voir après avoir cliqué sur le « i » :

**A propos de ce contenu** ✕

 **Le Figaro** ✓  
Entreprise de radiodiffusion et de production médiatique à Paris S'abonner

Le Figaro est un journal quotidien français fondé en 1826, sous le règne de Charles X. Il est à ce titre le plus ancien quotidien de la presse française encore publié. Il est nommé d'après Figaro, le personnage de Beaumarchais, dont il met en exergue la réplique : Sans la liberté de blâmer, il n'est... [Afficher la suite](#)

Extrait de Wikipédia (Creative Commons)

 Premier partage du lien : **Avril 2020**

**AUTRE CONTENU DE CE SITE WEB**

 **LEFIGARO.FR**  
«Pas de déconfinement si le confinement n'est pas réussi», prévient Jérôme Salomon

 **LEFIGARO.FR**  
Coronavirus: Emmanuel Macron met en garde contre un déconfinement trop rapide

 **LEFIGARO.FR**  
Donald Trump a signé le décret limitant temporairement l'immigration  
Cette suspension sera d'une durée initiale de 60 jours.

**PARTAGES**

# FACEBOOK

- 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 ;

Nous utilisons à la fois l'intelligence artificielle et des modérateurs humains pour vérifier chaque demande par rapport à nos règles sur les actualités qui ont été élaborées en collaboration avec des éditeurs, des universitaires et des organisations industrielles. Nous surveillons de près le processus d'examen pour nous assurer que les modérateurs respectent systématiquement nos règles. Notre équipe de modérateurs est diversifiée et compte des membres ayant des parcours variés, notamment en matière de politiques, de journalisme et d'opérations.

- 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.

Nous réexaminons périodiquement les pages incluses dans l'Index des pages d'actualités pour nous assurer que rien n'a changé, et nous procédons également à un réexamen si des informations clefs de la page changent.

## ***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

# FACEBOOK

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*.

Nous souhaitons permettre aux utilisateurs de mieux comprendre la provenance des contenus d'actualités qu'ils visualisent sur Facebook afin qu'ils puissent faire des choix éclairés sur les contenus qu'ils lisent.

Le [bouton contexte](#) de Facebook est conçu pour fournir plus de contexte et d'informations sur les éditeurs et les liens qui apparaissent dans le fil d'actualité afin que les gens puissent décider eux-mêmes de ce qu'ils veulent lire, croire et partager (Voir Question 3.1 pour plus de détails).

En juin 2020, nous avons commencé à labelliser les médias qui, selon nous et au regard d'un ensemble de critères élaborés, pourraient être entièrement ou partiellement sous le contrôle éditorial d'un Etat en tant que [médias contrôlés par l'État](#).

Nous maintiendrons ces pages à un niveau de transparence plus élevé parce qu'elles conjuguent le pouvoir d'influence de l'opinion des médias avec le soutien stratégique d'un État.

Nous développons notre définition et nos règles pour les médias contrôlés par un État avec la contribution de plus de 55 experts du monde entier spécialisés dans les secteurs des médias, de la gouvernance, des droits de l'homme et du développement. Les experts consultés représentent des institutions universitaires, des organisations à but non lucratif et des organisations internationales de premier plan dans ce domaine, notamment Reporters sans frontières, le Centre d'assistance aux médias internationaux (*Center for International Media Assistance*), le Centre Européen de Journalisme (*European Journalism Center*), le projet d'Oxford Internet Institute's Project sur la propagande informatique, le Centre pour les médias, les données et la société (CMDs) de l'Université d'Europe centrale, le Conseil de l'Europe, l'UNESCO et d'autres.

Il est important de noter que notre [politique](#) fait volontairement la distinction entre les médias contrôlés par un État et les médias publics, que nous définissons comme toute entité qui est financée par l'État, qui conserve une mission de service public et qui peut démontrer que son contrôle éditorial est indépendant. Nous concentrons actuellement nos efforts d'information uniquement sur les médias contrôlés par un Etat.

# FACEBOOK

## **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 *deepfakes*. 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) ;

Notre approche de la lutte contre les fausses informations repose sur la **mise en œuvre de systèmes de vérification des informations**.

Pour lutter contre la propagation de la désinformation et fournir des informations plus fiables aux utilisateurs, nous nous sommes associés à des médias de vérification tiers et indépendants (également appelés « partenaires de vérification indépendants ») à l'échelle mondiale, qui sont certifiés par l'IFCN, un réseau non politisé. L'IFCN, filiale du *Poynter Institute*, une organisation de recherche en journalisme, se consacre à réunir des médias de vérification indépendants dans le monde entier.

Notre programme de vérification des informations a été lancé en 2016, et nous avons continué de l'étendre depuis pour couvrir plus de langues et de pays.

Ce programme vise à lutter contre la désinformation virale, c'est-à-dire les fausses déclarations avérées et les déclarations manifestement fausses, en particulier celles susceptibles d'être mensongères ou de nuire.

Les partenaires vérification des informations peuvent examiner le contenu publié sur Facebook et sur Instagram, y compris les contenus organiques et les contenus boostés. Ils peuvent également examiner des vidéos, des images, des liens ou des publications composées uniquement de texte.

En ce qui concerne les **critères** utilisés par Facebook pour sélectionner ses partenaires de vérification des informations, les partenaires de vérification indépendants sont tous certifiés par l'IFCN. Le Code de principes de l'IFCN comprend une série d'engagements que les organisations doivent respecter afin de promouvoir l'excellence dans la vérification des faits: impartialité et

# FACEBOOK

équité; transparence des sources; transparence du financement et de l'organisation; transparence de la méthodologie; politique de correction accessibles et honnêtes.

Pour plus de détails, veuillez consulter :

<https://www.facebook.com/journalismproject/programs/third-party-fact-checking/selecting-partners>

- 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 ;

Comme expliqué à la Question 1.3.3, nous effectuons une vérification des informations en externe, mais, pour détecter une fausse information potentielle, nous utilisons un certain nombre de signaux pour prédire le contenu qui pourrait être une fausse information, et le soumettre aux partenaires de vérification indépendants.

Il existe plusieurs options d'évaluation que les partenaires de vérification indépendants peuvent appliquer au contenu :

- Faux : Contenu qui n'a aucun fondement dans les faits.
- Modifié : Image ou contenu audio ou vidéo qui a été modifié ou synthétisé au-delà des ajustements réalisés à des fins de clarté ou de qualité, d'une façon qui pourrait induire en erreur.
- Partiellement faux : Contenu qui comporte certaines inexactitudes factuelles.
- Contexte manquant : Contenu susceptible d'induire en erreur sans contexte supplémentaire.

Bien que notre objectif soit d'identifier la désinformation, les partenaires de vérification des informations peuvent également nous informer si le contenu qu'ils ont examiné est vrai ou satirique.

Pour plus d'informations à ce sujet, voir <https://www.facebook.com/journalismproject/programs/third-party-fact-checking/how-it-works> et <https://www.facebook.com/business/help/341102040382165?id=673052479947730>

# FACEBOOK

- 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 ;

Tous les types de contenus sont susceptibles d'être soumis à vérification. Les partenaires de vérification indépendants peuvent examiner et évaluer les publications publiques publiées sur Facebook et Instagram, y compris les publicités, les articles, les photos, les vidéos et les publications composées uniquement de texte.

- 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) ;

Nous identifions les fausses informations potentielles ou tout contenu préjudiciable à l'aide d'une variété de signaux, tels que les commentaires des utilisateurs sur Facebook, et nous soumettons le contenu aux partenaires de vérification indépendants. Ces derniers peuvent également identifier de manière proactive du contenu.

Les partenaires de vérification indépendants examinent le contenu, vérifient les faits en cause et évaluent leur exactitude. Cela se produit indépendamment de Facebook, et peut inclure de faire appel à des sources, à des données publiques, à l'authentification des vidéos et des images, et plus encore.

Bien que Facebook soit responsable d'établir certaines lignes directrices concernant l'examen du contenu, ce sont en fin de compte les partenaires de vérification indépendants qui examinent et évaluent le contenu de manière indépendante.

- 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) ;

Comme mentionné ci-dessus, l'examen du contenu par les partenaires de vérification indépendants se fait indépendamment de Facebook, et peut nécessiter la consultation d'autres sources et informations.

Nous avons mis en place un dispositif basé sur des options d'évaluation pour les partenaires de vérification indépendants. Vous trouverez sur le lien suivant

# FACEBOOK

<https://www.facebook.com/business/help/341102040382165?id=673052479947730> les options d'évaluation pour les partenaires de vérification indépendants, ainsi que des exemples de contenus susceptibles de correspondre à chaque évaluation.

- 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 ;

Nous n'avons pas de dispositif d'urgence, notre dispositif global est décrit à la Question 3.3. Mais, les partenaires de vérification indépendants sont libres de mettre en place d'autres processus de leur côté. Au cours des derniers mois, nous avons vu plusieurs de nos partenaires de vérification indépendants réorienter leurs travaux vers les fausses informations sur la COVID-19, ce que nous avons soutenu en lançant un programme de subvention de 1 million de dollars américains pour aider les organisations à maintenir ou à augmenter leur capacité pendant cette période (voir la question suivante).

En outre, nous avons un outil dédié aux événements d'actualités intégré à notre outil de vérification des informations qui permet aux partenaires de vérification indépendants de filtrer rapidement le contenu qui comprend un ensemble de mots-clés liés à un événement important. Cela a été utilisé pour des événements comme la COVID-19, les élections dans le monde, et plus encore.

- 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).

Notre équipe a lancé plusieurs programmes pour soutenir l'industrie de vérification des informations grâce à son partenariat avec l'IFCN. Nous ne divulguons pas les détails financiers de nos accords confidentiels.

Comme mentionné dans les Questions générales, en avril 2020, nous avons annoncé notre [partenariat](#) avec l'IFCN pour lancer un programme de subvention de 1 million de dollars américains afin d'aider les partenaires de vérification indépendants dans leurs travaux sur la COVID-19. En plus de fournir un financement essentiel permettant aux partenaires de maintenir ou d'augmenter leurs capacités pendant cette période, les subventions soutiennent également des propositions telles que des traductions de vérification des informations de leurs langues

# FACEBOOK

maternelles vers différentes langues, la production de contenus multimédia (vidéos, infographies, podcasts) sur la COVID-19, une collaboration avec des experts de la santé pour une couverture scientifique et factuelle, un lancement d'initiatives de développement du public utilisant des formats innovants (tels que la communication hors ligne ou interactive) pour mieux atteindre les personnes avec des informations fiables, un relais des partenaires de vérification indépendants du contenu publié par les autorités publiques avec des informations fiables pour une meilleure communication sur la COVID-19.

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**

À ce jour, notre programme de vérification des informations compte plus de 80 partenaires dans plus de 60 langues et nous travaillons assidûment pour continuer à étendre ce programme en Europe et dans le monde entier. Depuis que nous avons signé le Code de bonnes pratiques de l'Union Européenne contre la désinformation, nous avons développé notre programme en accueillant de nouveaux partenaires. Notre programme inclut désormais 35 partenaires vérifiant des informations dans 26 langues européennes : albanais, bosniaque, bulgare, croate, tchèque, danois, néerlandais, anglais, français, allemand, géorgien, grec, hongrois, italien, letton, lituanien, macédonien, norvégien, polonais, portugais, russe, roumain, slovaque, serbe, espagnol, suédois et ukrainien.

En France, Facebook et Instagram ont conclu des partenariats avec plusieurs médias français. Les Décodeurs du Monde, Les Observateurs de France 24, AFP Factuel et 20 minutes pour Facebook. Instagram collabore avec 20 minutes, l'Agence France Presse, Les Décodeurs du Monde et Les Observateurs de France 24.

- **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,**

Comme mentionné ci-dessous, nous identifions les fausses informations potentielles à l'aide de plusieurs signaux, tels que les commentaires des utilisateurs sur Facebook, et nous présentons le contenu aux partenaires de vérification indépendants. Les partenaires de vérification indépendants peuvent également identifier le contenu à examiner par eux-mêmes.



# FACEBOOK

- **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),

Le nombre de contenus faisant l'objet d'une intervention est indépendant des partenaires de vérification indépendants. Comme indiqué à la Question 1.4, nous nous concentrons sur le nombre de contenus retirés car violant les Standards de la communauté (Facebook) ou Règles de la communauté (Instagram) indépendamment de la façon dont ils ont été identifiés. Cependant, nous vous renvoyons aux indicateurs et statistiques relatifs aux *spams* et aux faux comptes (voir Question 1.4) ainsi qu'aux tendances générales concernant les avertissements (voir Questions générales).

- la nature des **liens juridiques** unissant l'opérateur aux *fact-checkers* tiers,

Il est renvoyé à l'IFCN.

- la nature des **contreparties** qui leur sont offertes.

Il est renvoyé à la Question 3.3 concernant le budget.

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

Nous faisons appel à des partenaires de vérification indépendants (voir ci-dessus).

- 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 ;

# FACEBOOK

- 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 ;

Les réponses ci-dessous concernent les quatre tirets ci-dessus.

Comme expliqué Question 1.3.5, si un partenaire évalue un contenu comme Faux, Modifié, Partiellement faux ou Contexte manquant, nous prenons une ou plusieurs des mesures suivantes :

- Distribution réduite : Nous affichons ce contenu plus bas dans le fil d'actualité, ce qui réduit considérablement sa diffusion. Et sur Instagram, nous le supprimons d'Explorer et des pages de hashtags, et le déclassons dans le fil ainsi que dans les Stories.
- Avertissement de partage : Lorsque quelqu'un essaie de partager une publication évaluée par un partenaire de vérification indépendant, nous lui envoyons une information sous forme de pop-up, pour qu'il puisse décider lui-même de ce qu'il doit lire, croire et partager.
- Notifications de partage : Si quelqu'un a partagé une actualité que les partenaires de vérification indépendants déterminent comme fausse plus tard, nous lui indiquons qu'il existe d'autres rapports concernant ce contenu.
- Étiquettes de désinformation : Nous appliquons une étiquette claire et visible sur le contenu qui a été discrédité par les partenaires de vérification indépendants, et présentons leurs articles de vérification des informations pour plus de contexte.
- Suppression des avantages pour les récidivistes : Lorsque des Pages ou des sites web partagent de façon répétée du contenu évalué comme faux par nos partenaires de vérification indépendants, la diffusion globale de leurs publications est réduite, et ils perdront la capacité de faire de la publicité ou de monétiser pendant un certain temps.

Les contenus classés « Faux » ou « Modifié » constituent la plus grave des fausses informations. Ainsi, nous réduirons considérablement la distribution de ces publications et leur attribuons nos étiquettes d'avertissement les plus fortes.

Le contenu évalué « Partiellement faux » comprend certaines inexactitudes factuelles. En conséquence, nous réduisons la distribution de ce contenu, mais moins que s'il était « Faux » ou

# FACEBOOK

« Modifié ». Pour le contenu qui est « Contexte manquant », nous nous concentrerons sur l’affichage de plus d’informations de nos partenaires de *fact-checking*. Pour les deux options d’évaluation, nous introduirons une étiquette d’avertissement plus légère pour refléter plus précisément les évaluations des partenaires de vérification indépendants.

- si l’intervention des *fact-checkers* est directement **visible sur le service** ;

Il est renvoyé à la Question 3.7.

- 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.

Nous avons mis en place des **voies de recours internes** à la disposition des éditeurs des contenus qui ont été analysés comme contenant ou susceptibles de contenir de fausses informations.

Les éditeurs peuvent [soumettre une correction pour un contenu qu’ils ont créé ou contester l’évaluation](#) reçue pour ce contenu. Les éditeurs peuvent contacter directement les partenaires de vérification indépendants (1) s’ils ont corrigé le contenu évalué ou (2) s’ils jugent incorrecte l’évaluation réalisée par le partenaire de vérification indépendant est inexacte. Les partenaires de vérification indépendants sont chargés d’évaluer la validité de chaque correction.

Si un partenaire de vérification indépendant décide de modifier une évaluation sur la base d’une correction ou d’une contestation, la rétrogradation du contenu sera annulée, les refus publicitaires associés pourront être annulés et la sanction contre la page ou le domaine devenant un récidiviste sera levée. La diffusion de la page ou du domaine peut mettre quelques jours pour revenir à la normale.

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) ;

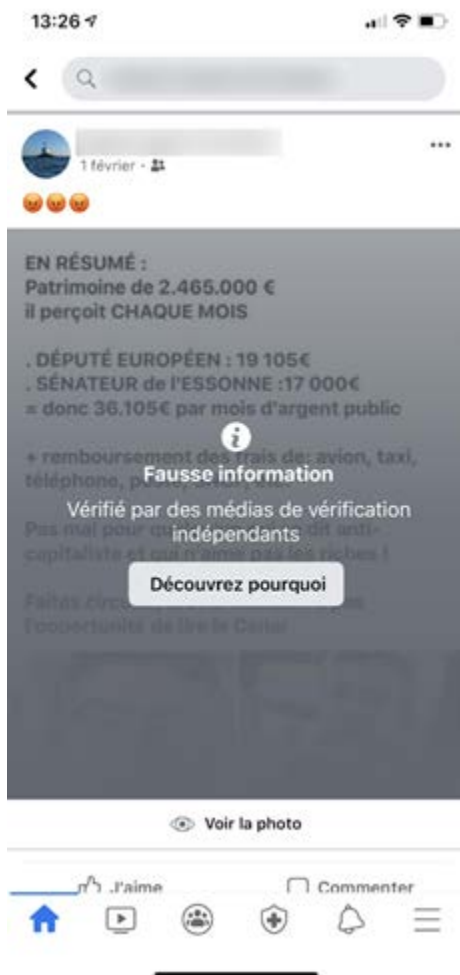
Lorsqu’un contenu est qualifié de faux ou partiellement faux par un partenaire de vérification indépendant, nous ajoutons des filtres bien visibles sur ce contenu. Ces filtres couvrent le

# FACEBOOK

contenu lui-même, que celui-ci soit partagé dans le fil d'actualité, les Stories ou par message et les utilisateurs doivent cliquer dessus pour découvrir le contenu.

Voir ci-dessous pour exemples les captures d'écran sur Facebook et Instagram.

Facebook:



# FACEBOOK

Instagram:



En cliquant sur l'option « Découvrez pourquoi » (cf. captures d'écran ci-dessus), les utilisateurs accèdent à un article du partenaire de vérification indépendant expliquant pourquoi le contenu a été qualifié de faux ou partiellement faux et fournissant des informations supplémentaires, notamment sur le contexte dans lequel s'inscrit le contenu litigieux. Ces filtres incluent notamment des liens vers l'évaluation du partenaire de vérification indépendant et des liens vers des articles provenant de sources fiables.

Nous affichons également le bouton Contexte d'un « i » dans le coin inférieur droit d'un article, afin de permettre aux utilisateurs d'en savoir plus sur les sites internet et éditeurs qui publient sur Facebook. Voir Question 3.1 à cet égard.

# FACEBOOK

S'agissant des éléments de contexte sur Instagram, par exemple, nous partageons plus d'informations concernant les comptes rassemblant un grand nombre de personnes. En permettant une plus grande transparence au sujet de ces comptes, les utilisateurs peuvent mieux évaluer le contenu qu'ils voient et savoir d'où il vient. L'option « *A propos de ce compte* » permet ainsi aux utilisateurs d'obtenir des informations précises à propos d'un compte, notamment : la date de création du compte, le pays dans lequel se situe le titulaire du compte, les publicités diffusées et les anciens noms d'utilisateurs utilisés au cours de la dernière année. Cette fonctionnalité est conçue pour fournir à nos utilisateurs plus d'informations concernant les comptes qu'ils suivent afin qu'ils puissent effectuer leurs propres choix, de façon éclairée, s'agissant des contenus qu'ils lisent, en lesquels ils croient et qu'ils partagent.

- de l'ensemble des utilisateurs en rendant visible, sur le service, les interventions sur un contenu donné ou un compte liées au *fact checking*.

Lorsqu'une personne est sur le point de partager une information qualifiée de fausse ou partiellement fausse par un partenaire de vérification indépendant, une fenêtre pop-up apparaît pour les informer que le contenu a été jugé comme faux par un partenaire de vérification indépendant et l'encourage à bien réfléchir avant de partager un tel contenu (voir les captures d'écran insérés en Question 1.4 pour Facebook et Instagram).

## **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.

Nous n'avons pas identifié de freins spécifiques.

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

De tels processus peuvent évoluer selon le scénario d'utilisation.

# FACEBOOK

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.

Dans le bilan de l'application et de l'effectivité des mesures mises en œuvre par les opérateurs en 2019, le CSA a formulé les recommandations suivantes concernant la promotion des contenus issues d'entreprises et d'agences de presse et services de communication audiovisuelle (EAP-SCA) (reproduites en *italique*) :

- *Il salue et encourage les pratiques consistant non seulement à **renseigner l'utilisateur** sur le nom et le logo de l'organisme publiant un contenu, mais aussi à lui fournir des éléments complémentaires d'information et de contexte sur les contenus des EAP-SCA mis en avant.*

Nous avons mis en œuvre des procédures de labellisation pour fournir plus de contexte et de transparence aux utilisateurs (voir Questions 3.1 et 3.2).

- *Lorsque l'identification des EAP-SCA n'est pas réalisée par la plateforme, il serait utile au CSA et aux utilisateurs de savoir si celle-ci s'assure de la fiabilité de cette identité et le cas échéant, selon quel processus.*

Il est renvoyé aux Questions 3.1 et 3.2 concernant le News Page Index (NPI). Pour être enregistrée sur le NPI, une page doit être associée à un domaine vérifié. La vérification de domaine permet aux propriétaires de pages de revendiquer la propriété de leur domaine. Cette propriété permet à l'éditeur de contrôler les éditions de ses liens et autres contenus afin d'empêcher l'utilisation abusive de son domaine et d'empêcher les acteurs malveillants de diffuser des fausses informations.

- *Le CSA recommande également aux plateformes qui identifient des sources de contenus comme étant « fiables » ou « d'autorité » de faire preuve de transparence à l'égard des utilisateurs quant aux critères qu'elles mobilisent et qui les conduisent à cette caractérisation.*

A titre d'exemple, en ce qui concerne la COVID-19, nous avons renforcé nos partenariats avec l'OMS et les autres institutions de référence pour accroître la fiabilité de l'information (voir Questions générales). En outre, nous travaillons au niveau national et de manière renforcée depuis la crise sanitaire, avec le Système d'Information du Gouvernement (SIG).

# FACEBOOK

- *Le CSA encourage les **démarches de labellisation des EAP-SCA** et estime utile que les utilisateurs des plateformes les connaissent.*

Nous avons mis en œuvre des procédures labellisation pour fournir plus de contexte et de transparence aux utilisateurs (voir Questions 3.1 et 3.2).

- *De manière générale, le CSA encourage fortement le développement de partenariats entre les plateformes et les fact-checkers.*

Comme détaillé ci-dessus et réitéré par le CSA dans son bilan, nous avons développé des partenariats solides avec les partenaires de vérification indépendants (voir Questions 3.3 à 3.7).

- *Parmi les initiatives existantes, le système mis en place par Facebook relève d'une approche particulièrement intéressante. Un espace dédié permet en effet à des tiers vérificateurs de confiance, strictement sélectionnés, de se saisir des contenus circulant sur les services de la plateforme, de vérifier leur véracité et d'intervenir le cas échéant pour expliquer dans quelle mesure tel contenu est susceptible de contenir une fausse information. Il permet d'assurer la diffusion d'une information fiable et indépendante tout en ne faisant pas reposer sur la seule plateforme la responsabilité d'apprécier la véracité d'une information. De manière générale, le CSA recommande aux opérateurs qui auraient recours à ce type de dispositifs de l'assortir d'une procédure d'urgence permettant de répondre de façon adaptée à une situation exceptionnelle (ex. : événement capital ayant un caractère d'urgence).*

Comme expliqué dans les Questions générales, nous avons étendu notre partenariat de vérification des informations. Ce programme a prouvé son efficacité. À titre d'exemple, au cours du seul mois d'avril 2020, nous avons labellisé environ 50 millions de contenus liés à la COVID-19 sur Facebook, sur la base d'environ 7 500 articles de nos partenaires de vérification indépendants. Lorsque les utilisateurs voient ces avertissements, ils ne visualisent pas le contenu d'origine dans 95 % des cas.

Nous n'avons pas de dispositif d'urgence, comme précisé Question 1.3.2. Néanmoins, nous travaillons en étroite collaboration avec le régulateur et les autorités françaises (par exemple à travers le « Groupe de Contact Permanent » dirigé par le ministère de l'Intérieur) et avons un point de contact dédié pour répondre aux requêtes urgentes liées aux situations de crise.



# FACEBOOK

- *La fonction ClaimReview, utilisée par Bing, Facebook, Google et YouTube et qui permet à l'éditeur d'apposer le libellé « fact-checking » au contenu de fact-checking qu'il édite, est également un outil intéressant. Elle repose sur un moindre contrôle à l'entrée et une plus grande ouverture (pas de processus de sélection), ce qui permet une utilisation plus large mais induit davantage d'incertitudes quant à l'identité de la source d'où émane le contenu balisé.*

Nous utilisons la fonctionnalité ClaimReview pour permettre aux éditeurs d'appliquer l'étiquette « fact-checking ». Nous avons encouragé l'adoption d'outils permettant de présenter et de mieux comprendre la vérification des informations sur internet.

- *Enfin, le module NewsGuard, en partenariat avec Bing, qui indique au moyen d'un code-couleur la fiabilité d'une source d'information, contribue également à la bonne information des utilisateurs. Il mériterait d'être disponible par défaut plutôt qu'optionnel.*

Nous avons mis en œuvre des procédures labellisation pour fournir plus de contexte et de transparence aux utilisateurs (voir Questions 3.1 et 3.2).

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.

Nous appuyons le travail de la communauté mondiale de santé publique visant à maintenir les gens en sécurité et informés pendant la crise de santé publique causée par la COVID-19. Nous nous efforçons également de faire face aux impacts à long terme de cette pandémie en soutenant les industries dans le besoin et en facilitant la recherche et l'offre d'aide dans leurs communautés. Par exemple, le Centre d'informations sur le coronavirus, présenté en haut du fil d'actualité, comprend des mises à jour en temps réel provenant des autorités sanitaires nationales et des organisations mondiales telles que l'OMS, ainsi que des articles, des vidéos et des publications utiles sur la distanciation sociale et la prévention de la propagation de la COVID-19. Voir également Question 6.4 concernant la campagne lancée avec l'AFP en avril 2020, qui a fait plus de 9 millions de vues.

# FACEBOOK

## 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.*

### **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.

**Catégories** : il existe 4 types généraux de désinformation :

- faux comptes ;
- spams ;
- fausses informations contribuant à causer un dommage physique ;
- fausses informations contribuant à la pression sur les électeurs ; et
- Vidéos manipulées et trompeuses.

Il est renvoyé à la Question 1.4 pour plus d'informations sur ces catégories ainsi qu'à la Question 2.9 concernant les comportements inauthentiques coordonnées qui utilisent nos applications pour tenter de manipuler le débat public.

# FACEBOOK

**Origine** : nous prenons des mesures en fonction du comportement que nous observons sur notre plateforme.

**Moyens de diffusion** : humains (acteurs mal intentionnés) et automatisés.

**Nature de l'objectif poursuivi par leurs émetteurs** : Nous avons constaté qu'un grand nombre de fausses informations sont partagées dans le but de réaliser des profits. Ces *spammeurs* gagnent de l'argent en se faisant passer pour des éditeurs d'informations légitimes et en publiant des canulars (*hoax*) pour que les utilisateurs visitent leurs sites qui sont en réalité, dans la majorité des cas, essentiellement constitués de publicités. C'est pourquoi l'une des approches les plus efficaces consiste à supprimer les incitations économiques pour les réseaux de désinformation. Voir Question 5.1 à cet égard.

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 ;

Il est renvoyé à la Question 2.9 ci-dessus concernant les comportements inauthentiques coordonnés.

- 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.

Afin de lutter plus efficacement contre la désinformation, nous prenons des mesures à l'encontre des pages et des domaines qui partagent ou publient de manière répétée du contenu qualifié de faux ou modifié. Ces pages et domaines verront leur distribution réduite, perdront leur capacité à monétiser et à faire de la publicité, et perdront leur capacité à s'enregistrer en tant que page d'actualités sur Facebook. Si une page d'actualité enregistrée partage de façon répétée de fausses informations, son enregistrement comme telle lui sera retiré. Au fil du temps, les pages

# FACEBOOK

et les domaines peuvent retrouver leur distribution et leur capacité à monétiser et à faire de la publicité s'ils cessent de partager de fausses informations.

De la même façon, la visibilité des comptes Instagram partageant de fausses informations de manière répétées et de leurs publications sera réduite, par exemple ils n'apparaîtront plus dans la section Explorer et les pages de hashtag (recherche d'informations via un hashtag).

## ***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 :

- o 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 ;
- o de la lutte contre l'accélération et la viralité des contenus contenant de fausses informations.

### 4.3.1 Détecter

Il est renvoyé aux Questions 1.4 et 2.9 concernant notre stratégie de détection des pratiques de diffusion massive.

### 4.3.2 Faire obstacle

Il est renvoyé aux Questions 1.4 et 2.9 concernant notre stratégie pour faire obstacle aux pratiques de diffusion massive.

Nous avons également mis en place des outils à la disposition des utilisateurs pour faire obstacle à ces pratiques :

# FACEBOOK

- Contrôler qui peut publier et commenter : les administrateurs de pages peuvent choisir de désactiver ou d'examiner les publications de visiteurs. Ces publications resteront non visibles jusqu'à ce qu'elles soient revues. Par ailleurs, nous avons lancé de nouvelles fonctionnalités sur Facebook permettant aux utilisateurs de décider qui peut commenter les publications : tout le monde, les personnes identifiées par la page ou les personnes que les pages suivent. Sur Instagram, les utilisateurs peuvent autoriser les commentaires de tout le monde, ou en limiter la portée.
- Contrôler l'affichage de certains mots sur la page et bannir certains utilisateurs : avec le filtre à injures et les outils modération de page, les administrateurs de pages peuvent automatiquement bloquer des termes spécifiques de leurs publications, stories et vidéos en direct. Sur Instagram, les comptes peuvent activer un filtre par mots-clés pour masquer les commentaires contenant des termes, des phrases ou même des émojis spécifiques qu'ils souhaitent éviter.
- Protéger la communauté : tout commentaire qui semble inapproprié peut être masqué et rendu de nouveau visible facilement. Sur les publications d'une page, tout commentaire peut être supprimé. Les contenus abusifs ou les *spams* peuvent également être signalés en utilisant le lien « Trouver de l'aide ou signaler ».

Enfin, nous avons mis au point un outil d'analyse de publications publiques sur Facebook - [CrowdTangle](#) - qui permet de suivre, d'analyser et de signaler facilement des contenus publics sur les médias sociaux. En outre, nous avons ouvert un canal dédié pour que le CSA puisse faire des demandes spécifiques et puisse avoir plus de visibilité sur le contenu le plus viral partagé sur notre service.

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 ;

Il est renvoyé aux Questions 1.4 et 2.9.

# FACEBOOK

- les **origines de cette détection** (moyens technologiques, signalements par des tiers, signalements en interne) et leur répartition ;

Il est renvoyé aux Questions 1.3 et 1.4 concernant la procédure de traitement des signalements et les indicateurs et statistiques.

- 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) ;

Il est renvoyé aux Questions 1.4 et 2.9.

- les **moyens financiers et humains** mis en œuvre par l'opérateur pour lutter contre les comptes propageant massivement de fausses informations ;

Il est renvoyé aux Questions 1.3 et 1.4 concernant la procédure de traitement des signalements et les indicateurs et statistiques.

- 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 information.

Concernant les mesures prises, nous vous renvoyons à la Question 4.1 et 5.a.

## ***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) de comptes 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 ;

# FACEBOOK

- les modalités de consultation et de téléchargement de **ces informations par l'utilisateur**.

Il est renvoyé à notre CSER publiquement accessible à l'adresse <https://transparency.facebook.com/community-standards-enforcement> et les rapports sur les comportements inauthentiques coordonnés publiquement accessibles à l'adresse <https://about.fb.com/news/tag/coordinated-inauthentic-behavior/>.

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 ;

Les informations destinées aux utilisateurs concernant les risques encourus en cas de création de contenus violant nos Standards de la Communauté (décrits dans la section IV relative à « l'intégrité et l'authenticité », couvrant l'intégrité du compte et l'identité authentique, le contenu indésirable, la cybersécurité, le comportement trompeur, les fausses informations, les médias manipulés et la commémoration) sont fournies dans nos Conditions de service - voir ci-dessous pour plus d'informations à cet égard.

- 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 un souci de transparence, les utilisateurs ont connaissance des Standards de la Communauté (Facebook) et des Règles de la Communauté (Instagram). Ces informations sont fournies par écrit, sont claires et facilement accessibles.

Pour s'inscrire et utiliser le service Facebook, tout utilisateur doit accepter les Conditions de Service. Les Conditions de service confèrent expressément à Facebook Ireland le droit de supprimer du contenu et de mettre un terme à tout ou partie du service Facebook pour les utilisateurs qui violent les Conditions de service de manière grave ou répétée. Les Conditions de

# FACEBOOK

service renvoient également les Standards de la Communauté, ainsi que d'autres dispositions interdisant aux utilisateurs de publier du contenu illicite.

En particulier, la section 3.2 des Conditions de Service (intitulée « *Ce que vous pouvez partager et faire sur Facebook* ») prévoit :

*« Nous voulons que les gens utilisent Facebook pour s'exprimer et partager du contenu qui leur importe, mais pas au détriment de la sécurité et du bien-être des autres ou de l'intégrité de notre communauté. Vous acceptez donc de ne pas agir de la manière décrite ci-dessous (ni d'encourager ou de soutenir les autres à le faire) :*

1. *Vous ne pouvez pas utiliser nos Produits pour faire ou partager quoi que ce soit :*
  - *qui enfreint les présentes Conditions, nos [Standards de la communauté](#) et les [autres conditions générales et règlements](#) applicables à votre utilisation de Facebook ;*
  - *qui est illégal, trompeur, discriminant ou frauduleux ;*
  - *qui enfreint ou viole des droits tiers, y compris des droits de propriété intellectuelle.*

*[...] Nous pouvons supprimer ou bloquer le contenu qui enfreint ces dispositions. »*

*En outre, la section 4.2 des Conditions (intitulée « *Suspension ou suppression d'un compte* ») prévoit :*

*« Nous voulons faire de Facebook un lieu à la fois chaleureux et sécurisé, où chacun peut s'exprimer et partager ses opinions et ses idées.*

*S'il s'avère que vous avez [manifestement, gravement ou à maintes reprises enfreint nos Conditions ou nos Règlements](#), et notamment nos [Standards de la communauté](#), nous pourrons suspendre ou désactiver définitivement l'accès à votre compte.* » (souligné par nos soins).

*Les Standards de la communauté prennent en compte la grande diversité des utilisations que l'on peut faire du service Facebook. Comme expliqué dans les Conditions de service, les Standards de la communauté présentent "[...] nos standards quant au contenu que vous publiez sur Facebook et à vos activités sur Facebook et sur d'autres produits Facebook". Les Standards de la communauté interdisent également la publication de certains contenus afin de garantir la*



# FACEBOOK

*sécurité et le respect de tous sur le service Facebook. Les types de contenus que nous interdisons sont clairement présentés dans nos Standards de la communauté :*

*« Nous sommes conscients de l'importance pour Facebook d'offrir un espace incitant les gens à communiquer, et nous prenons très au sérieux notre rôle afin d'écarter les abus de notre service. C'est la raison pour laquelle nous avons créé un ensemble de Standards de la communauté qui expliquent ce qui est autorisé, ou non, sur Facebook. Nos politiques sont fondées sur les retours de notre communauté et l'avis de spécialistes dans des domaines comme les technologies, la sécurité publique et les droits de l'homme. Pour nous assurer que tout le monde est respecté, nous nous efforçons de créer des politiques qui prennent en compte les différents points de vue et croyances, surtout ceux des personnes et communautés marginalisées ou négligées. »*

De même, sur Instagram, les Conditions d'utilisation renvoient aux Règles de la communauté. Les Conditions d'utilisation interdisent aux utilisateurs d'Instagram de violer (ou d'aider ou d'inciter autrui à violer) les Conditions ou nos politiques, en ce compris notamment les Règles de la Communauté, la Politique de Plate-forme Instagram et les Règles applicables à la musique.

En outre, sur Instagram, les Conditions d'utilisation prévoient :

*« Nous pouvons supprimer tout contenu ou toute information que vous partagez sur le Service si nous estimons qu'il y a infraction aux présentes Conditions d'utilisation ou à nos politiques (y compris nos [Règles de la communauté Instagram](#)), ou si la loi nous y oblige. Nous pouvons immédiatement refuser de vous fournir le Service ou arrêter de vous fournir le Service en tout ou en partie (et notamment en mettant fin ou en désactivant votre accès aux Produits Facebook et aux Produits des Entités Facebook) afin de protéger notre communauté ou nos services, ou si vous générez un risque ou une responsabilité légale pour nous, enfreignez les présentes Conditions d'utilisation ou nos politiques (notamment nos [Règles de la Communauté Instagram](#)), si vous violez de manière répétée les droits de propriété intellectuelle d'autres personnes ou si la loi nous y oblige. »*

Nos politiques sont fondées sur les retours de la communauté et l'avis de spécialistes dans des domaines tels que la technologie, la sécurité publique et les droits de l'homme. Elles sont rédigées et mises en œuvre par nos équipes internes dont l'expérience couvre un large éventail de domaines, notamment l'application de la loi, les droits de l'homme, l'aide sociale, la lutte contre le terrorisme, le monde universitaire et le droit. Les Standards de la communauté sont régulièrement mis à jour et sont accessibles au public à l'adresse suivante :

# FACEBOOK

<https://www.facebook.com/communitystandards/>. Les Règles de la communauté (Instagram) sont régulièrement mises à jour et mises à la disposition du public à l'adresse suivante : <https://help.instagram.com/477434105621119>. Nous appliquons ces Standards et Règles, ainsi que nos Conditions, lorsque nous décidons de la suppression de contenus ou l'interdiction d'utilisateurs.

## **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.

Il est renvoyé aux Questions 1.4 (en particulier "Vidéos manipulées et trompeuses") et 2.5. Nous ne faisons pas une telle distinction entre les *deepfakes* et les autres contenus dans notre CSER.

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**.

Il est renvoyé à la Question 1.4.

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 son bilan de l'application et de l'effectivité des mesures mises en œuvre par les opérateurs en 2019, le CSA a formulé les recommandations suivantes relatives à la lutte contre les comptes propageant massivement de fausses informations (reproduites en *italique*) :

# FACEBOOK

- *Il demande aux opérateurs de lui fournir, à l'avenir, le **détail de leur procédure de détection et de traitement des comptes** propageant massivement de fausses informations.*

Nous avons fourni au CSA toutes ces informations : voir Questions 1.4, 2.9, 4.1 et 4.3.

- *Il souhaite obtenir des informations sur les **recettes publicitaires**, même minimales, générées par les comptes ayant propagé ou susceptibles de propager massivement de fausses informations et qui n'ont pas été détectés et désactivés dès leur création.*

En période électorale, nous avons mis en place des mesures favorisant la transparence des campagnes publicitaires grâce à notre outil Bibliothèque publicitaire (voir Question 5.10).

- *Dans un souci de pédagogie et de transparence à l'égard des utilisateurs, le CSA encourage les opérateurs à **développer l'information aux utilisateurs** sur les mesures de lutte contre de tels comptes.*

Les utilisateurs sont également pleinement informés de nos procédures. Nous fournissons en effet de nombreux détails sur nos procédures de détection et de traitement des comptes responsables de la propagation massive de fausses informations (voir Questions 4.3 et 4.6).

- 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.

Ces mesures renvoient à nos développements sur la manière dont nous avons augmenté nos efforts pour supprimer les fausses déclarations sur Facebook et Instagram concernant la COVID-19, les vaccins COVID-19 et les vaccins en général pendant la pandémie (voir Questions générales).

# FACEBOOK

## **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**

*Il est apparu au Conseil, à l'aune de ses travaux et de l'examen des déclarations concernant l'exercice 2019, qu'un des points cruciaux de la lutte contre la diffusion des fausses informations était l'action sur les liens entre flux financiers et fausses informations. En effet, ces dernières peuvent être, d'une façon ou d'une autre, source de revenus pour différents types d'acteurs ; elles peuvent également bénéficier, directement ou non, des mécanismes de diffusion des contenus commerciaux. En conséquence, ces liens peuvent être un facteur, d'une part, d'augmentation de la viralité des fausses informations et, d'autre part, d'incitation à les diffuser. Si ces phénomènes dépassent le seul cadre des services de plateformes en ligne, ces derniers peuvent y participer. Ainsi, le Conseil a élaboré les questions suivantes avec son Comité d'experts sur la désinformation en ligne afin d'interroger les opérateurs sur l'ampleur réelle de ces liens ainsi que sur les moyens de lutte mis en œuvre en la matière.*

Comme indiqué ci-dessus, lorsqu'il s'agit de lutter contre les fausses informations, une des approches les plus efficaces consiste à supprimer les incitations économiques pour les réseaux de désinformation.

Facebook vise à s'attaquer aux incitations économiques derrière les fausses informations en agissant sur les liens entre flux financiers et fausses informations. Facebook peut déclasser ou désactiver la possibilité de faire de la publicité pour les acteurs qui diffusent des fausses informations pour des raisons financières. Nous vous renvoyons à notre politique concernant les récidivistes à la Question 4.2.

Nous interdisons également diverses formes de comportements trompeurs sur notre service, en prenant systématiquement des mesures contre les [comportements inauthentiques \(IB\)](#). Comme développé dans nos Standards de la communauté, le comportement inauthentique/trompeur est une tentative de tromper les gens ou Facebook sur la popularité d'un contenu, le but d'une communauté (c'est-à-dire les groupes, les pages, les événements) ou l'identité des personnes la

# FACEBOOK

composant. Cela inclut également les comportements conçus pour tromper Facebook et échapper aux contrôles et aux limites que nous posons pour l'utilisation de nos plateformes.

Le comportement inauthentique est principalement axé sur l'augmentation de la viralité d'un contenu et est souvent motivé par des raisons financières. Ce n'est pas nouveau. Les gens essaient depuis toujours d'utiliser des techniques non authentiques pour gagner de l'argent. Récemment, nous avons vu des acteurs malveillants à travers le monde utiliser les questions politiques et sociales américaines pour attirer les utilisateurs vers des sites tiers à Facebook composés quasiment essentiellement de publicités ou ayant pour objet la vente de marchandises. Nous les avons également vus utiliser des tactiques similaires pour tirer parti de sujets importants dans d'autres régions du monde. À première vue, ils peuvent être confondus avec des opérations d'influence politiquement motivées, alors qu'en fait ce sont des acteurs malveillants qui utilisent simplement des thèmes politiques comme une forme de *spam* ou des pièges à clics. Dans notre rapport d'octobre 2020, nous avons partagé des exemples de comportements inauthentiques trouvés au cours des derniers mois et de la façon dont nous les avons traités.

Étant donné qu'une partie du contenu partagé par les acteurs à l'origine de comportements inauthentiques ne viole pas en lui-même nos règles, nous prenons des mesures basées sur le comportement trompeur et non sur le contenu. Dans notre rapport d'[octobre 2020](#), nous avons donné plusieurs exemples de sanctions que nous mettons en œuvre, qui peuvent consister en des avertissements allant de la réduction de la distribution d'un contenu à l'interdiction des acteurs à l'origine de comportements inauthentiques de notre service. Le rapport décrit également comment nos politiques ont évolué au fil du temps pour rester à jour sur l'évolution des comportements trompeurs. Ces comportements vont continuer d'évoluer, et nous continuerons de nous adapter.

# FACEBOOK

## ***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. [7]

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 définitions des annonces publicitaires (annonces créées via des outils spécifiques), des contenus sponsorisés (contenus organiques dont la visibilité est augmentée contre rémunération) et des contenus d'utilisateurs en partenariat (contenus organiques créés par un utilisateur en partenariat avec un tiers contre rémunération) sont pertinentes pour Facebook/Instagram et également adaptées à notre modèle puisque nous utilisons les catégories similaires suivantes :

- Publicités : les campagnes publicitaires sont mises en place via un outil publicitaire dédié, le Gestionnaire de publicités, qui est le point de départ pour diffuser des publicités sur Facebook, Instagram, Messenger ou Audience Network. C'est un outil tout-en-un pour créer des publicités, gérer l'endroit et le moment de leur diffusion et suivre les performances des campagnes. Les opérateurs peuvent y créer de nouvelles campagnes publicitaires, les vérifier et les ajuster et accéder à des rapports publicitaires en temps réel. Pour plus d'informations, voir : <https://www.facebook.com/business/tools/ads-manager> and <https://www.facebook.com/business/help/280594309504695?id=2030300680608329>
- Contenus sponsorisés : nous appelons les contenus sponsorisés « Publications boostées ». Ce sont des annonces que les opérateurs créent à partir de publications sur leur page Facebook ou leur profil Instagram. « Booster » une publication peut les aider à générer plus de réactions, à augmenter les partages et commentaires. La visibilité d'une publication peut augmenter contre rémunération. Pour plus d'informations, voir : <https://www.facebook.com/business/help/240208966080581?id=352109282177656>

# FACEBOOK

- Contenus en partenariat : ce que nous appelons « contenu de marque » est un contenu de créateur ou d'éditeur qui met en avant ou est influencé par un partenaire commercial dans le cadre d'un échange de valeur. Nous avons récemment lancé de nouveaux outils pour le contenu de marque qui obligent les créateurs à ajouter une étiquette indiquant « *Partenariat rémunéré avec* », afin de montrer clairement que le contenu est le résultat d'un partenariat. Pour plus d'informations, voir : <https://www.facebook.com/business/help/788160621327601?id=1912903575666924>

## ***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 ?

Le placement des espaces publicitaires sur Facebook repose sur une combinaison de systèmes d'enchères et d'algorithmes en temps réel. Le placement peut être automatique ou manuel.

Il existe 2 principaux types de publicités payantes sur Facebook :

- « Achat couverture et répétition » : l'achat couverture et répétition permet de planifier et d'acheter des campagnes à l'avance, avec une diffusion publicitaire prévisible et plus de contrôle sur les paramètres de répétition. Les publicités peuvent être diffusées sur Facebook et Instagram. En outre, les placements Audience Network sont disponibles pour les objectifs suivants: Notoriété de la marque, Couverture, Vues de vidéo et Interaction avec la publication.
- « Achat enchère » : l'achat Enchère offre davantage de choix, d'efficacité et de flexibilité, avec des résultats moins prévisibles. Les publicités peuvent être diffusées sur Facebook, Instagram, Messenger et Audience Network.

# FACEBOOK

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.).

Le système de diffusion des publicités de Facebook utilise les enchères et l'apprentissage automatique pour déterminer à qui, quand et où nous montrons vos publicités. Ces processus fonctionnent ensemble pour optimiser la valeur à la fois pour les utilisateurs et pour les entreprises.

Afin de diffuser des publicités pertinentes, chaque fois qu'une occasion de montrer une publicité se présente, une enchère est lancée pour déterminer quelle publicité montrer. Comme nous voulons que chaque utilisateur voit des publicités pertinentes, les enchères prennent en considération les prévisions de pertinence de chaque publicité pour la personne concernée. Cela signifie que les publicités ayant une pertinence élevée peuvent remporter les enchères à moindre coût. Les prévisions de pertinence estiment la probabilité que cette personne considère la publicité comme étant de qualité et effectue l'action désirée par l'annonceur. Les diagnostics de pertinence des publicités aident à déterminer si des publicités que vous avez diffusées étaient adaptées à l'audience ciblée.

Notre volonté est que chaque annonceur atteigne ses objectifs à moindre coût. Le système de diffusion utilise également l'apprentissage automatique pour améliorer les performances de chaque publicité. Les événements d'optimisation révèlent dans quelle mesure une stratégie d'enchères a permis à un annonceur d'atteindre ses objectifs, selon ses critères actuels d'optimisation. Les performances sont optimisées pour minimiser le coût de chaque publicité par événement d'optimisation. Chaque fois qu'une publicité est affichée, les prédictions de pertinence du système de diffusion deviennent plus précises. En conséquence, plus une publicité est diffusée, plus le système de diffusion détermine où, quand et auprès de qui diffuser la publicité pour minimiser le coût par événement d'optimisation. Nous appelons la période où le système de diffusion a encore beaucoup à apprendre sur une publicité la « phase d'apprentissage ».

Parfois, la publicité ou l'ensemble de publicités obtenant les meilleurs résultats n'est pas l'ensemble de publicités qui a obtenu le coût le plus bas par événement d'optimisation. En effet, le système de diffusion des publicités utilise des prévisions de futures performances pour



# FACEBOOK

déterminer où les diffuser ensuite (et non les performances passées de chaque ensemble de publicités). C'est la raison pour laquelle nous recommandons souvent l'utilisation des placements automatiques et les audiences larges afin d'offrir au système de diffusion davantage de flexibilité afin de trouver des conversions à moindre coût.

Les [placements automatiques](#) nous permettent d'obtenir les meilleurs résultats issus de tous les placements par défaut. Parce que nous pouvons choisir des résultats parmi la plus large gamme possible, les placements automatiques permettent généralement d'optimiser l'utilisation du budget des annonceurs et aide à contrôler les coûts.

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 dispositif de signalement pour les publicités et les contenus sponsorisés comprend, comme le dispositif de signalement pour les autres contenus, la possibilité de signaler de fausses informations (autre que le contenu trompeur ou indésirable) (Voir Question 1.1.)

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**.

Il est renvoyé à la Question 1 pour les quatre tirets ci-dessus.

# FACEBOOK

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 ;

Nous avons un processus de vérification des annonces publicitaires (voir Question 5.4). Afin de confirmer qu'un compte Business Manager appartient à une entreprise ou une organisation réelle, Facebook peut demander aux annonceurs de compléter un [processus de vérification](#), qui vérifie leur identité et leur existence en tant que personne morale.

Ce système de vérification est renforcé en ce qui concerne les publicités portant sur un enjeu social, électoral ou politique. La transparence est une priorité sur de tels enjeux, c'est pourquoi la [bibliothèque publicitaire](#) affiche des informations supplémentaires concernant ces publicités, comme qui a financé la publicité, une fourchette des dépenses impliquées et la couverture de la publicité pour plusieurs profils démographiques. Nous stockons ces publicités dans la bibliothèque pendant 7 ans. Voir Question 5.10.

- 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 ;

Afin de détecter et de lutter contre la publicité en faveur des sites internet spécialisés dans les fausses informations, nous comptons sur nos partenaires de vérification indépendants. Pour plus de détails sur le fonctionnement de notre dispositif de vérification des informations, voir Questions 3.3 à 3.7.

De plus, nous prenons des mesures contre les récidivistes (voir Question 4.3).

- 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 ;

Nous collaborons régulièrement avec les gouvernements et les autorités judiciaires et de police, les chercheurs en cybersécurité et d'autres entreprises technologiques pour détecter les

# FACEBOOK

menaces émergentes. Nous disposons également d'un canal dédié avec la DGCCRF pour accroître la coopération. Enfin, nous nous sommes mis à la disposition de la Sous-direction de lutte contre la cybercriminalité et nous avons accepté avec enthousiasme de recevoir tout signal/renseignement concernant les escroqueries, les publicités, les fausses informations sur la COVID19 (aucune remontée majeure reçue à ce stade).

## 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.

Nous interdisons les publicités comportant des déclarations démenties par les partenaires de vérification indépendants ou, dans certaines circonstances, des déclarations démenties par des organisations possédant une expertise particulière. Les annonceurs qui publient de manière répétée des informations jugées fausses peuvent se voir imposer des restrictions quant à leur capacité à faire de la publicité sur Facebook. Plus globalement, nous nous engageons à assurer la transparence de la publicité et avons mis en place des mesures visant à rendre difficile l'utilisation de notre plateforme pour tirer profit de fausses informations. Les partenaires de vérification indépendants peuvent vérifier toutes les publicités de notre bibliothèque publicitaire, et nous rejetons les publicités qu'ils jugent fausses, partiellement fausses, modifiées ou sans contexte.

# FACEBOOK

5.7 L'opérateur détaillera les **mesures de sécurité des marques** mises en place et expliquera comment elles ont été conçues. Il énumérera notamment les éléments suivants :

- **liste des mesures[8] :**

- Désactivation du placement : possibilité de contrôler les plateformes et placements dans lesquels les publicités apparaissent;
- Liste d'autorisation d'éditeurs : possibilité de sélectionner les éditeurs sur lesquels l'annonceur souhaite que ses annonces apparaissent et d'empêcher qu'elles apparaissent ailleurs sur ces placements ;
- Exclusions LIVE Stream : possibilité de désactiver l'affichage en direct sur Facebook d'un groupe sélectionné de partenaires approuvés ;
- Listes de blocage : possibilité de bloquer la diffusion de publicités sur des pages, des sites internet et/ou des applications spécifiques ;
- Listes d'autorisation de contenus : possibilité de travailler avec un tiers de confiance pour examiner et personnaliser les ensembles vidéo adaptés à la marque pour exécuter des campagnes en direct ;
- Liste des éditeurs : possibilité de consulter la liste complète des éditeurs sur lesquels les annonces peuvent être diffusées ;
- Filtres d'inventaire : possibilité d'empêcher la diffusion des publicités aux côtés de contenus sensibles ;
- Exclusion de flux : possibilité de choisir l'exclusion au niveau du contenu parmi 4 sujets différents : Actualités, politique, jeux et contenu religieux/spirituel ; et
- Rapports de diffusion : possibilité d'afficher, de filtrer et de télécharger des rapports pour voir où les publicités sont réellement apparues.

# FACEBOOK

- **moyens humains et automatisés** dédiés à l'exécution de ces mesures ;

- Technologie : Technologie IA pour l'application à grande échelle ;
- Examen humain : ~ 15 000 modérateurs, 24 heures sur 24, 7 jours sur 7 et dans plus de 50 langues ;
- Contribution d'experts : partenariats avec des universitaires et des experts en la matière.

Pour plus de détails voir Question 1.4

- appréciation de l'**impact** de ces mesures;

Les rapports de diffusion listés ci-dessus montrent que les préférences définies dans les outils ont fonctionné de manière globale.

- existence d'une **information des annonceurs** lorsque leurs contenus se sont trouvés accolés à de fausses informations et, le cas échéant, de **mécanismes de compensation ou de recours** ;

Conformément à nos Règles publicitaires, nous n'autorisons pas les annonceurs à diffuser des publicités incluant du contenu qui a été démenti par des partenaires de vérification indépendants. Si une publicité est désapprouvée car elle contient de fausses informations, nous pouvons en informer les annonceurs de deux manières : (i) une notification d'administrateur de page, qui spécifie le partenaire de vérification indépendant qui a évalué leur contenu ; et (ii) une notification de désapprobation de publicité dans leur interface de création de publicités (par ex, le Gestionnaire de publicités).

Les annonceurs peuvent contacter directement le partenaire de vérification indépendant qui a évalué leur contenu pour apporter une correction ou contester une évaluation du contenu créé par l'annonceur. Dans les cas où les publicités sont rejetées parce qu'elles contiennent les mêmes informations que celles que les partenaires de vérification indépendants ont démenties, les annonceurs peuvent demander que la correspondance soit réexaminée par Facebook via la notification de désapprobation de publicités. La véracité du contenu ne peut pas être contestée via ce canal, et ces correspondances ne peuvent pas être contestées auprès des partenaires de vérification indépendants.

# FACEBOOK

Si la correction ou la contestation d'un annonceur est réussie, la rétrogradation de l'annonce associée sera levée, à condition que nous n'ayons pas identifié d'autres violations de nos Règles publicitaires.

En outre, nous avons mis en place des procédures de réclamation à la disposition des annonceurs en cas de désactivation des publicités, qui s'appliquent généralement si les annonces violent nos Règles publicitaires et Conditions de service, mais ces procédures de réclamations ne sont pas spécifiques aux publicités liées à de fausses informations. Si un placement, une offre ou un espace publicitaire a été bloqué, les utilisateurs peuvent envoyer une réclamation via le Gestionnaire de monétisation. S'ils disposent de la possibilité de faire une réclamation, ils pourront le faire en suivant [certaines étapes](#).

Si la publicité d'un annonceur a été [refusée](#) parce qu'elle n'était pas entièrement conforme à nos Règles publicitaires, nous lui enverrons un e-mail contenant plus d'informations. Cela peut aider les annonceurs à décider s'ils préfèrent modifier leur publicité pour se conformer à nos règles ou demander un nouvel examen.

Si un annonceur ne peut pas modifier sa publicité ou pense que son refus est injustifié (au regard de nos Règles publicitaires), il peut demander un nouvel examen sur la page Qualité du compte. Nous examinerons son annonce et déterminerons si elle respecte les Règles publicitaires. Si la décision initiale concernant sa publicité est annulée et que ses campagnes sont actives, ses publicités commenceront à être diffusées. Si nous annulons la décision, le statut de sa publicité sera le même que celui de sa campagne.

Pour plus d'informations, veuillez consulter <https://www.facebook.com/business/help/968402779873601?helpref=search&sr=6&query=advertiser%20appeal#>

- modalités de participation de l'opérateur dans tout mécanisme de **coopération** avec les marques et l'interprofession publicitaire [9].

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

Nous avons participé aux travaux de coopération suivants :

**GARM** : The Global Alliance for Responsible Media (GARM) de la Fédération mondiale des annonceurs (*World Federation of Advertisers*) afin de s'aligner sur les normes et les définitions

# FACEBOOK

de sécurité de la marque, l'éducation à la mise à l'échelle, les outils et systèmes communs et la supervision indépendante de l'industrie.

**TAG** : Le 10 septembre 2020, le Trustworthy Accountability Group (TAG), un programme mondial de publicité numérique, a annoncé Facebook comme l'un des premiers groupes de sociétés certifiées Sécurité des marques par le TAG.

Enfin, le 23 septembre 2020, Facebook a reçu l'**IAB UK Gold Standard 2.0**.

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.

Il est renvoyé à la Question 5.6.

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,

- o si celles-ci peuvent être filtrées spécifiquement pour la France,
- o si l'utilisateur peut les consulter et les télécharger,
- o si cette base a évolué depuis 2019.

Nous disposons de trois bases de données principales qui sont toutes publiques, qui peuvent être filtrées spécifiquement pour la France et qui sont en constante évolution. Veuillez trouver une description de ce que ces bases de données contiennent ci-dessous :

# FACEBOOK

- **[La bibliothèque publicitaire](#)**. La bibliothèque publicitaire (*Ad Library*) assure la transparence de la publicité en offrant une collection complète et consultable de toutes les publicités actives et en cours de diffusion sur Facebook et Instagram. Elle inclut les publicités non politiques. Toutefois, les publicités portant sur un enjeu social, électoral ou politique sont archivées et conservées dans la bibliothèque publicitaire pendant sept ans. La bibliothèque publicitaire permet de sélectionner des publicités selon le nombre de fois qu'elles sont apparues à l'écran, de dépenses, d'âge et de genre ciblés pour chaque publicité. Les utilisateurs peuvent signaler les publicités qui devraient être accompagnées de la mention « *Payé par* » et qui ne le seraient pas, ou qui violent d'autres règles publicitaires de Facebook, directement depuis la bibliothèque. Tout le monde peut consulter la bibliothèque, avec ou sans compte Facebook.
- **L'API de la bibliothèque publicitaire**. L'interface de programme d'application (*API*) de la bibliothèque publicitaire permet aux régulateurs, journalistes, groupes de contrôle citoyen et autres personnes d'analyser les annonces ayant un enjeu social, électoral ou politique et d'aider à responsabiliser les annonceurs et Facebook. Les utilisateurs peuvent effectuer des recherches personnalisées par mot clé sur les publicités stockées dans la bibliothèque publicitaire. Les résultats de l'API comprennent des données relatives aux créations et aux performances publicitaires archivées dans la bibliothèque publicitaire. Depuis son lancement européen en mars 2019, nous avons amélioré notre API afin que les personnes puissent facilement accéder aux annonces depuis un pays donné et faire des requêtes sur des annonceurs spécifiques. Nous nous efforçons de faciliter l'accès par la programmation aux images, vidéos publicitaires et aux annonces récemment publiées.
- **Le rapport de la bibliothèque publicitaire**. Le rapport de la bibliothèque publicitaire (*Ad Library Report*) est un rapport d'informations synthétique concernant les publicités sur des questions sociales, électorales ou politiques collectées dans la bibliothèque publicitaire. Il fournit des statistiques détaillées concernant les annonces dans la bibliothèque publicitaire. Par exemple, le rapport montre le nombre total de publicités et le montant total dépensé pour les publicités dans la bibliothèque publicitaire par pays, tout comme le montant total dépensé par annonceur, le montant total dépensé par annonceur par jour et les mots-clés les plus recherchés au cours de la semaine passée. Les utilisateurs peuvent consulter, filtrer et télécharger les données dans un fichier CSV. Le rapport de la bibliothèque publicitaire est utile pour les personnes souhaitant comprendre de manière globale l'importance de l'activité dans la librairie publicitaire



# FACEBOOK

depuis son lancement, alors que la bibliothèque publicitaire va plus servir à chercher à analyser des annonces spécifiques. Vous pouvez consulter ce rapport pour la France [ici](#).

Pour plus de détails sur les mises à jour ici <https://about.fb.com/news/2020/01/political-ads/> ,  
<https://about.fb.com/news/2019/06/offering-greater-transparency/> et  
<https://about.fb.com/news/2019/01/european-parliament-elections/>.

## **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 ?

Le contenu de marque suit les mêmes règles et normes décrites ci-dessus que tout autre contenu sur le service.

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

Il est renvoyé à la Question 5.1.

**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**

*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 d'information se rattachant à un débat d'intérêt général 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 tels contenus au titre des mesures complémentaires prévues par l'article 11 de la loi du 22 décembre 2018.*

# FACEBOOK

## 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 ?

Oui, les publicités et les contenus sponsorisés se distinguent des autres contenus car ils sont étiquetés « sponsorisés ».

De plus, pour accroître la transparence des annonceurs, les publicités d'enjeu social, électoral et politique identifiées sur Facebook et Instagram doivent être clairement labellisées avec la mention « *Payé par* » (voir capture d'écran ci-dessous).

- 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 ?

Il est renvoyé aux Questions 2.7 (en particulier « *Activité sur Facebook* ») et 5.4.

Nous pensons que plus de transparence entraîne, au fil du temps, plus de responsabilités – pas seulement pour Facebook mais aussi pour les annonceurs. C'est la raison pour laquelle nous continuons à introduire des outils permettant aux élus, aux candidats à une fonction électorale et aux organisations souhaitant influencer l'opinion publique de révéler davantage d'informations sur les publicités qu'ils diffusent et les personnes à l'origine de celles-ci. Nous avons déployé nos outils de transparence dans le monde entier pour les annonceurs qui souhaitent placer des publicités sur des questions liées à la politique mais aussi, plus généralement, à la société et l'intérêt général.

Au sein de l'Union Européenne, nous avons lancé les outils de transparence des publicités fin mars 2019. Ces outils constituent une part essentielle de nos efforts en matière d'intégrité des élections et comprennent :

- **La confirmation d'identité et les autorisations.** Afin de prévenir les abus et les interférences étrangères, les administrateurs des pages doivent fournir un document

# FACEBOOK

d'identification pouvant être vérifié/confirmé. Les administrateurs ne peuvent être autorisés que dans un seul pays et doivent fournir un document d'identité pour le pays qu'ils souhaitent cibler pour la publicité.

- **Le marquage des publicités politiques et d'intérêt général.** Pour augmenter la transparence de l'annonceur, ce type de publicités sur Facebook et Instagram doivent être clairement identifiées par une mention « *Payé par* » (voir capture d'écran ci-dessous). Cette mention comporte plusieurs éléments qui sont vérifiés avant d'être approuvés – le nom de la page ou de l'organisation ; le site internet ; l'adresse électronique ; le numéro de téléphone ; et l'adresse. Ces informations sont incluses dans chaque publicité et restent publiquement accessibles dans notre bibliothèque publicitaire.



Nous avons également mis à jour notre bibliothèque publicitaire pour augmenter le niveau de transparence qu'elle offre aux gens et leur donner plus de contrôle sur les publicités qu'ils voient (Voir notre présentation de la Bibliothèque publicitaire à la Question 5.10).

# FACEBOOK

- 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) ?

Oui, il est renvoyé à la Question 2.7 (en particulier « *Activité sur Facebook* »).

- 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) ?

Oui, il est renvoyé à la Question 2.7 (en particulier « *Activité sur Facebook* »).

- 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 ? Comment l'utilisateur peut-il accéder à ces descriptions ?

Facebook a défini un ensemble de politiques applicables aux communications commerciales et promotionnelles accessibles ici : [Règles publicitaires \(facebook.com\)](https://www.facebook.com/policies). Facebook interdit toutes publicités contenant des revendications réfutées par des partenaires de vérification indépendants ou, dans certaines circonstances, par des organismes spécialisés. Les annonceurs qui publient à plusieurs reprises des informations jugées fausses peuvent faire l'objet de restrictions publicitaires sur Facebook.

- 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 ?

Si le contenu n'a pas été supprimé, cela signifie qu'il n'y a pas violation de nos Standards de la communauté. Néanmoins, en labellisant les contenus, nous appelons les utilisateurs à la vigilance (voir détails et capture d'écran dans la Question 3.7).

Facebook encourage également ses utilisateurs à la vigilance par le biais de (i) son outil spécifique « *Pourquoi je vois cette publicité ?* » et (ii) les Préférences publicitaires, que nous avons introduits il y a plus de quatre ans pour offrir davantage de transparence et de contrôle aux utilisateurs sur les publicités auxquelles ils ont accès.

# FACEBOOK

- *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 ?*

Nos utilisateurs ont la possibilité de masquer les publicités et les contenus sponsorisés, mais aussi de paramétrer leurs préférences publicitaires. Ils peuvent en particulier : (i) choisir [de voir moins de publicités portant sur certains sujets](#) comme l'alcool, la parentalité, les animaux de compagnie ou un enjeu social, électoral ou politique ; (ii) choisir de ne pas voir d'annonces d'un annonceur en particulier ; et (iii) choisir de masquer une annonce à laquelle ils ont été exposés afin de ne pas la revoir.

Ils sont informés de cette possibilité sur leurs Préférences publicitaires. Ces paramètres sont disponibles pour les utilisateurs sur leur profil Facebook, les pages, etc.

## ***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 ;

Il est renvoyé à la Question 5.10 concernant la bibliothèque publicitaire et à la Question 5.6 concernant le système de vérification pour les tirets ci-dessus.

# FACEBOOK

- si ces contenus font l'objet de **restrictions ou règles particulières** (ex : restriction de ciblage) ;

Nous appliquons des restrictions de ciblage à tous les types de publicités. Par exemple, nous avons déjà supprimé des milliers de catégories de ciblage liées à des catégories protégées telles que l'origine raciale, l'origine ethnique, l'orientation sexuelle et la religion. Des règles strictes de confidentialité des données au sein de l'Union Européenne imposent également certaines restrictions de ciblage, telles que les données à caractère personnel concernant les « opinions politiques » des personnes (pour plus de détails, veuillez consulter : [https://about.fb.com/de/wp-content/uploads/sites/10/2020/09/Facebook\\_Response\\_European\\_Democrac\\_Action\\_Plan\\_2020.09.15.pdf](https://about.fb.com/de/wp-content/uploads/sites/10/2020/09/Facebook_Response_European_Democrac_Action_Plan_2020.09.15.pdf))

Vous trouverez plus d'informations sur ce qui précède ici :

- <https://www.facebook.com/business/news/reviewing-targeting-to-ensure-advertising-is-safe-and-civil>;
- <https://www.facebook.com/business/news/keeping-advertising-safe-and-civil>;
- <https://www.facebook.com/business/news/update-to-facebook-ads-targeting-categories/>

- si la procédure de signalement des contenus d'informations se rattachant à un débat d'intérêt général diffère de celle des autres communications commerciales ;

Le dispositif de signalement des publicités ne diffère pas du dispositif général de signalement, sauf que les utilisateurs peuvent également signaler des contenus trompeurs et indésirables (voir Question 1.1).

- les différences de traitement de ce type de communications commerciales avec le reste des communications commerciales.

Nous ne faisons aucune différence.

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.

# FACEBOOK

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 tels que listées à la question 5.10.

Il est renvoyé à la Question 5.10 concernant la bibliothèque publicitaire.

## **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.

Dans le contexte de la crise sanitaire causée par la COVID-19, nous travaillons avec des organismes de santé et des dirigeants communautaires pour mener des [campagnes sur notre plateforme](#) afin de promouvoir des informations faisant autorité sur les vaccins contre la COVID-19 et encourager les gens à se faire vacciner. Nous donnons 120 millions de dollars américains en crédits publicitaires pour aider les ministères de la Santé, les ONG et les agences des Nations Unies à atteindre des milliards de personnes dans le monde avec le vaccin contre la COVID-19 et l'information sur la santé préventive.

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.

Dans son bilan de l'application et de l'effectivité des mesures mises en œuvres par les opérateurs de 2019, le CSA a formulé les recommandations suivantes concernant l'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 (reproduit en *italique*) :

- *Le CSA encourage les plateformes à continuer à mettre en place une **labellisation claire et directement accessible** des contenus promus, aisément visible par l'utilisateur et précisant directement l'identité de l'annonceur.*

# FACEBOOK

Nous avons mis en œuvre plusieurs mesures pour permettre une labellisation claire et directement accessible des contenus promus (voir Question 5.2).

- Dans un souci de clarté et de transparence pour l'utilisateur, il est essentiel que les politiques publicitaires, notamment en matière de publicités trompeuses ou inappropriées, soient accessibles **en français** et comprennent une **partie relative à la manipulation de l'information**. Un renvoi vers ces documents pourrait être envisagé dans différentes situations (par exemple, à partir du contenu).

Nos Règles publicitaires sont disponibles en français et comprennent une section sur la désinformation (voir Question 5.13).

- *Les outils de transparence et de responsabilisation des utilisateurs sont à encourager, tels que ceux **leur permettant de comprendre pourquoi ils ont été ciblés ou de paramétrer leurs préférences publicitaires**. Les fonctionnalités de paramétrage doivent toutefois être accompagnées d'une information claire et facilement accessible permettant d'en saisir pleinement les implications .*

Les utilisateurs ont toutes les informations nécessaires pour comprendre pourquoi ils sont ciblés et comment ils peuvent définir leurs préférences publicitaires. Ces paramètres sont accompagnés d'informations permettant aux utilisateurs de comprendre pleinement les implications encourues. Voir Question 2.7.

- *Le CSA demande aux opérateurs, à l'avenir, un effort supplémentaire de transparence à son égard s'agissant des **mesures permettant aux contenus sponsorisés de gagner en visibilité**. La compréhension de ces mécanismes est essentielle pour lutter contre les fausses informations.*

Nous partageons de nombreuses informations sur la façon dont les contenus sponsorisés gagnent en visibilité, notamment dans nos pages d'aide pour les entreprises (voir Question 5.4).

- *Il appelle les plateformes à adopter des **mesures relatives à la sécurité des marques et à la démonétisation des pages** propageant des fausses informations afin d'empêcher les créateurs de ces contenus d'en tirer des ressources financières. Il appelle également l'attention des annonceurs sur les effets négatifs de l'association de leur marque et de leur image à un contenu propageant des fausses informations.*



# FACEBOOK

Nous avons mis en œuvre différentes mesures pour la sécurité des marques (voir Question 5.8). Nous supprimons déjà les incitations économiques pour les récidivistes (voir Question 5.1).

- *Afin d'éviter une situation dans laquelle un annonceur serait empêché sur une plateforme mais pas sur une autre, il recommande **un dialogue entre les opérateurs** permettant d'œuvrer à une plus grande harmonisation des approches retenues en matière de promotion de contenus d'information se rattachant à un débat d'intérêt général.*

Nous sommes d'accord avec une telle approche, c'est pourquoi nous participons à des travaux de coopération (voir Question 5.8).

- *Il invite à cette fin les opérateurs à faire en sorte que leurs bibliothèques publicitaires, notamment pour les contenus promus d'information se rattachant à un débat d'intérêt général, **soient accessibles sans difficulté** et faciles d'utilisation. Elles doivent répondre à un degré de transparence accru, par exemple en fournissant des informations détaillées sur l'identité de l'annonceur, les montants dépensés, le public visé et le public touché, le nombre de publications ou encore en permettant de faire des recherches complètes.*

Nous confirmons que nos bibliothèques publicitaires sont toutes facilement accessibles et conviviales (voir Question 5.10).

## 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 des sources d'informations** qui sont proposées sur le service;

Il est renvoyé aux Questions 3 et 5 pour une description de nos outils permettant aux utilisateurs d'identifier la fiabilité des informations et des sources d'informations. De plus, nous avons développé les Ressources sur le numérique et le Portail pour les jeunes utilisateurs de Facebook.

**Ressources sur le numérique** : nous avons lancé les Ressources sur le numérique (*Digital Literacy Library* - DLL) dans 45 langues dans le monde entier. Les DLL sont une collection de cours prêts à

# FACEBOOK

l'emploi fournis par l'équipe Jeunesse et Média (*Youth and Media*) du *Berkman Klein Center for Internet & Society* de l'université d'Harvard, qui vise à aider les jeunes âgés de 11 à 18 ans à faire preuve d'esprit critique et à partager de manière intelligente du contenu en ligne. Les cours abordent des thèmes tels que la vie privée et la réputation, l'exploration de l'identité, la sécurité, la sûreté, le bien-être et bien d'autres. Cette bibliothèque est une ressource pour les éducateurs qui souhaitent aborder la question de la culture du numérique et aider ces jeunes à acquérir les compétences nécessaires pour profiter en toute sécurité de la technologie numérique.

**Portail pour les jeunes utilisateurs de Facebook :** nous avons lancé notre portail pour les jeunes (*Facebook Youth Portal*), un lieu central pour les jeunes, qui comprend :

- **Un volet éducatif :** des informations afin de tirer le meilleur profit de produits tels que les pages, groupes, événements et profil, tout en préservant sa sécurité. Egalement, des informations sur les types de données collectées par Facebook et comment celles-ci sont utilisées.
- **L'avis d'autres jeunes :** l'avis de jeunes adolescents du monde entier sur comment ils utilisent la technologie de manière innovante et créative.
- **Les moyens de contrôler leur expérience :** des conseils sur des sujets tels que la sécurité, le signalement de contenu et sur comment choisir qui peut voir ce que vous partagez.
- **Des conseils :** notamment sur ce que vous devez faire si vous avez besoin de faire une « pause » des réseaux sociaux, et sur comment profiter au mieux d'Internet.

Plus généralement, comme expliqué dans cette Déclaration, nous avons élargi nos partenariats avec des organisations locales, parrainé des programmes d'éducation aux médias et partagé des conseils pour aider à repérer les fausses informations.

- 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 ;

Programme de vérification des informations. Voir Questions 3.3 à 3.7.

# FACEBOOK

**Continuer à combattre les organisations et les mouvements liés à la violence.** En août 2020, nous avons élargi notre [politique sur les individus et organismes dangereux](#) afin de nous attaquer aux mouvements sociaux militarisés et aux réseaux de complots violents, comme QAnon. Ces réseaux peuvent tirer parti de certaines crises ou certains événements critiques, tels que la pandémie de la COVID-19, pour faire progresser leurs objectifs, de sorte que les retirer de nos services était une étape essentielle pour assurer la sécurité de nos services pour notre communauté.

De plus, lorsqu'un utilisateur cherche des termes liés à QAnon sur Facebook et Instagram, nous le redirigeons vers des ressources crédibles du [Global Network on Extremism and Technology](#) (GNET), le réseau de recherche universitaire du Forum Internet mondial contre le terrorisme. Il s'agit du dernier élargissement de notre initiative Redirect pour aider à lutter contre l'extrémisme violent et nous orientons les gens vers des ressources qui peuvent les aider à s'informer sur les réalités de QAnon et ses liens avec la violence et les dommages du monde réel. Alors que nous continuons à étudier l'impact de notre application contre QAnon, nous travaillons en partenariat avec GNET pour évaluer l'impact de cette initiative Redirect, et nous continuons à réévaluer la liste des termes qui, lorsqu'ils seront recherchés sur notre service, devraient diriger les gens vers ces ressources.

**Nous avons conclu un partenariat avec Freeformers et plus de 20 ONG nationales et organismes de formation** pour offrir un programme de formation portant sur le numérique à 75 000 citoyens au sein de sept pays d'Europe : l'Allemagne, la France, l'Italie, l'Espagne, la Pologne, le Royaume-Uni et l'Irlande. Cette série d'ateliers a été conçue pour développer la confiance et les compétences nécessaires aux individus dans une économie numérique et pour les inciter à vouloir en savoir plus, y compris dans le domaine de l'éducation aux médias pour assurer leur sécurité en ligne.

- 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) ;

**Nous avons également conclu un partenariat avec l'Association Européenne pour les intérêts des spectateurs EAVI** (*European Association for Viewers Interests*) pour participer au Yo!Fest (un événement pour les jeunes organisé par le Forum européen de la jeunesse) pour dispenser deux

# FACEBOOK

sessions de formation à l'éducation aux médias et une session de questions-réponses sur l'éducation aux médias.

**En 2020, nous avons lancé une campagne d'éducation aux médias contre les fausses informations liées à la COVID-19. La campagne « Trois questions pour aider à éradiquer les fake news »,** est apparue sur les publicités Facebook et a dirigé les gens vers un site internet dédié avec un panel complet de conseils, [www.stampoutfalsenews.com](http://www.stampoutfalsenews.com). Ces annonces demandaient aux gens de questionner les informations qu'ils voient sur les publications en se posant les questions suivantes : 1) D'où vient l'information ? S'il n'y a pas de source, cherchez-en une; 2) Qu'est-ce qu'il manque ? Lisez toute l'histoire, pas seulement les gros titres. 3) Quel est mon ressenti ? Les gens qui font des fake news essaient de manipuler les sentiments.

Parallèlement à la campagne, nous avons mené une étude pour examiner l'efficacité de notre approche. Nous avons divisé l'audience totale de la campagne en groupes de test et de contrôle. Grâce à ce test, nous avons constaté une hausse positive de 3,4 points chez les utilisateurs se souvenant activement d'avoir vu les annonces et une hausse de 1,8 points chez les utilisateurs signalant un changement de comportement après avoir interagi avec la campagne. Le pourcentage indique la différence entre le groupe de test et celui de contrôle. Nous utiliserons ces résultats pour éclairer les futures campagnes d'éducation aux médias en France.

**Le Fonds pour le Civisme en ligne :** ce Fonds a été mis en place pour soutenir les comportements responsables sur internet, lutter contre les discours haineux et l'intimidation, et favoriser l'esprit critique et l'éducation aux médias. 1 millions d'euros a été donné à environ 20 initiatives françaises pour les soutenir dans leur action à l'éducation à travers la France. Vous pouvez retrouver l'ensemble des informations [ici](#).

- **des soutiens à des initiatives indépendantes émanant de journalistes et de chercheurs destinées à mieux comprendre et mesurer le phénomène de la désinformation.**

**Chaire universitaire d'éducation aux médias** Facebook a lancé, en janvier 2020, une chaire universitaire nationale sur l'éducation aux médias en partenariat avec l'École supérieure de journalisme de Lille (ESJ). Elle vise à créer des outils et programmes d'éducation sur le numérique dont l'efficacité et l'impact auprès des jeunes ont été approuvés et testés. Ces outils sont accessibles à tous les intervenants, écoles et universités français. L'ESJ de Lille a conclu des partenariats avec d'autres laboratoires de recherche étrangers basés au Canada, en Belgique et

# FACEBOOK

dans d'autres pays afin de mettre en commun et d'intensifier les efforts et le travail de recherche dirigé par la chaire universitaire

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.

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

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 ?) ;

**La plateforme Facebook Open Research & Transparency (FORT) :** Cette plateforme fournit aux universitaires et aux chercheurs indépendants les outils et les données dont ils ont besoin pour étudier l'impact de Facebook sur le monde, en mettant l'accent sur les élections, la démocratie et le bien-être. Depuis le 1er février 2021, les chercheurs ont accès à des informations ciblées pour plus de 1,65 million de publicités sur les enjeux sociaux, électoraux et politiques via la plateforme FORT. Nous avons créé cet outil pour permettre aux chercheurs universitaires d'étudier l'impact des produits Facebook sur les élections, et avons inclus des mesures pour protéger la vie privée des gens et garder la plateforme sécurisée. Cet ensemble de données

# FACEBOOK

comprend les publications diffusées au cours de la période de trois mois précédant le jour du scrutin, du 3 août au 3 novembre 2020. Il s'agit notamment d'un ensemble de données sur le ciblage des annonces (les informations de ciblage sélectionnées par les annonceurs qui diffusent des publicités sur les enjeux sociaux, électoraux et politiques, mais qui, par mesure de protection de la vie privée, excluent les annonces de moins de 100 impressions) et d'un ensemble de données sur la bibliothèque publicitaire (les publicités sur les enjeux sociaux, électoraux et politiques qui font partie de la bibliothèque publicitaires) afin que les chercheurs puissent analyser les annonces et les informations de ciblage dans le même environnement.

**Changement climatique** : Facebook a fait appel à des experts de l'Université George Mason, du Yale Climate Change Communications Program et de l'Université de Cambridge pour mettre en œuvre une nouvelle fonctionnalité du Centre d'information sur le climat afin de mettre en évidence les mythes sur les changements climatiques. Pour plus d'informations sur nos évolutions pour relever le défi du changement climatique, voir: <https://about.fb.com/news/2020/09/stepping-up-the-fight-against-climate-change/>.

Voir également la **Chaire universitaire d'éducation aux médias** décrite à la Question 6.1.

- 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.

**Partenariat sur les attitudes politiques lors des élections américaines de 2020** : Facebook a annoncé le 31 août 2020 un nouveau [partenariat de recherche](#) pour mieux comprendre l'impact de Facebook et d'Instagram sur les attitudes et comportements politiques ayant eu un impact lors des élections américaines de 2020, dans la continuation de l'initiative que nous avons lancée en 2018. Ce partenariat examinera l'impact sur la façon dont les gens interagissent avec nos produits, y compris le contenu partagé dans le fil d'actualité et sur Instagram, et le rôle de fonctionnalités telles que les systèmes de classement de contenus. Il s'agit d'un partenariat entre des chercheurs de Facebook et des universitaires externes indépendants. En externe, le projet est dirigé par les professeurs [Talia Stroud](#) et [Joshua A. Tucker](#), deux universitaires indépendants qui assurent la présidence des comités Social Science One. Les professeurs Tucker et Stroud ont sélectionné 15 chercheurs supplémentaires pour travailler avec eux sur ce projet, sur la base de leur expertise.

# FACEBOOK

- 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**.

Le 21 décembre 2020, nous avons [soumis des observations](#) à l'Observatoire européen des médias numériques (*European Digital Media Observatory*) (EDMO) en réponse à son appel à commentaires du 24 novembre concernant l'accès aux données détenues par les plateformes numériques aux fins de la recherche scientifique sociale. L'EDMO prévoit de former un groupe de travail en 2021 pour élaborer un code de conduite conformément à l'Article 40 du Règlement général sur la protection des données (RGPD) afin de faciliter le partage responsable des données, y compris des données sur les plateformes numériques, à des fins de recherche en sciences sociales. Vous pouvez lire nos commentaires [ici](#). Avec nos commentaires, nous espérons donner un aperçu de certains des défis que nous avons rencontrés et des solutions potentielles pour partager les données aux fins de la recherche en sciences sociales d'une manière qui protège la vie privée des gens.

- 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**.

Comme développé ci-dessus, nous avons conclu plusieurs partenariats avec le milieu de la recherche.

## **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.

Dans son bilan de l'application et de l'effectivité des mesures mises en œuvre par les opérateurs en 2019, le CSA a formulé les recommandations suivantes relatives à l'éducation aux médias et à l'information (EMI) (reproduit en *italique*) :

# FACEBOOK

- Il invite les opérateurs à développer des **approches multidimensionnelles** en matière d'éducation aux médias et à l'information, à destination de tous les publics (8-12 ans, 12-15 ans, 16-18 ans, 19-25 ans, 26-40 ans, 40-65 ans et + de 65 ans, parents, enseignants, etc.)

Comme développé ci-dessus, nous avons pris plusieurs mesures pour soutenir l'EMI auprès des jeunes publics, par exemple, nous avons lancé les Ressources sur le numérique et le Portail pour les jeunes utilisateurs de Facebook (voir Question 6.1 ci-dessus).

- Il les encourage à nouer des **partenariats pluriannuels** en matière d'EMI avec des institutions reconnues et des acteurs majeurs du monde de l'éducation (institutions publiques, associations, etc.).

Nous avons déjà commencé à nouer des partenariats pluriannuels avec des acteurs majeurs dans le monde de l'éducation, tels que le Berkman Klein Center for Internet & Society de l'Université de Harvard.

- Il les incite à **évaluer l'impact** de leurs actions sur le comportement des utilisateurs

Nous nous assurons déjà que les jeunes utilisateurs comprennent l'impact de leurs actions et comportements sur Facebook. Par exemple, le Portail pour les jeunes utilisateurs de Facebook comprend notamment une section sur la façon de « *Partager en toute sécurité* » (voir [ici](#)).

- Le CSA les encourage vivement à bâtir avec le monde de la recherche des **protocoles opérationnels de partage de données** et d'exploitation transparente de ces dernières.

Nous avons créé plusieurs outils permettant au monde de la recherche de mieux comprendre le rôle de Facebook dans notre société (par exemple, la plateforme FORT ou la Chaire d'éducation aux médias). Plus particulièrement en ce qui concerne le partage et l'utilisation des données, nous avons soumis des commentaires à l'Observatoire européen des médias numériques concernant l'accès aux données détenues par les plateformes numériques aux fins de la recherche scientifique sociale. Voir Question 6.2.



# FACEBOOK

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.

**Campagne d'éducation aux médias Covid-19** : Comme mentionné à la Question 6.1, nous avons mené une campagne de 4 semaines au cours de l'été 2020 pour éduquer et informer les gens sur la façon de détecter les fausses informations potentielles. La campagne a touché 16 millions d'utilisateurs en France.

Par ailleurs, en avril 2020, nous avons mis en avant sur la page [Facebook France](#) et dans les fils d'actualité d'un grand nombre d'utilisateurs français des vidéos éducatives sur les fausses informations sur la COVID-19, produites par des journalistes de l'AFP. Chaque épisode dure environ trois minutes et retrace l'ADN d'une fausse actualité, de sa création à sa diffusion, et explique le travail des partenaires de vérification de l'AFP. Les équipes analysent une fausse information liée à la COVID-19, expliquent comment cela fonctionne et donnent des conseils sur la façon de détecter les fausses informations et d'aider les gens à mieux comprendre les informations qu'ils voient. Par exemple, dans le premier épisode, un partenaire de vérification de l'AFP explique pourquoi l'Institut Pasteur n'a pas créé la COVID-19. Vous pouvez consulter toutes nos actions dans le cadre de la lutte contre la COVID-19, régulièrement mises à jour [ici](#) pour Facebook et [ici](#) pour Instagram.

« **Tous ensemble contre les fausses informations sur le COVID-19** » : En partenariat avec l'OMS et nos partenaires de vérification indépendants, nous avons lancé « Tous ensemble contre les fausses informations sur le COVID-19 », une campagne publicitaire d'éducation aux médias pour aider les utilisateurs à détecter la désinformation sur les vaccins contre la COVID-19. La campagne a orienté les utilisateurs de Facebook en France vers le lien du ministère de la Santé. <https://www.facebook.com/combating-misinfo>

# FACEBOOK

Nom de la campagne, de l'outil ou de l'action de soutien (indiquer l'URL)	Descripti on 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	Budge t 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 mesures (si aucune, préciser les difficultés rencontrés)
<b>Les Ressources sur le numérique</b>	Voir Question 6.1	N/A	Leçons disponibles sur le service Facebook	Jeunes entre 11-18 ans	monde	Automne 2018 / en cours	N/A	Site internet Facebook dédié	N/A	N/A

# FACEBOOK

Nom de la campagne, de l'outil ou de l'action de soutien (indiquer l'URL)	Descripti on 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 e t 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 mesures (si aucune, préciser les difficultés rencontrés)
<b>Portail pour les jeunes utilisateurs de Facebook</b>	Voir Question 6.1	N/A	Service offert par Facebook	Jeunes	monde	Mai 2018 / en cours	N/A	Site internet Facebook dédié	N/A	N/A

# FACEBOOK

<b>Partenariat de vérification des informations</b>	Voir Questions 3.3 à 3.7	Nous avons conclu un partenariat avec plus de 80 partenaires dans le monde, dans plus de 60 langues.	N/A	Tous les utilisateurs	monde	2016 / en cours	N/A	N/A	Voir Questions générales	N/A
<b>Continuer à combattre les mouvements et les organisations liés à la violence</b>	Voir Question 6.1	GNET	Contenu actionné par Facebook	Tous les utilisateurs	monde	Août 2020	N/A	N/A	N/A	N/A
<b>Programme de formation aux</b>	Voir Question 6.1	Freeformers et plus de 20 ONG nationales et	Contribution à	Tous les utilisateurs	France, Allemagne, Italie, Espagne,	2018 - 2020	N/A	Communiqués de presse	N/A	N/A

# FACEBOOK

Nom de la campagne, de l'outil ou de l'action de soutien (indiquer l'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	Budget 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 mesures (si aucune, préciser les difficultés rencontrés)
compétences numériques		organismes de formation	l'organisation des ateliers		Pologne, le Royaume-Uni et l'Irlande					

# FACEBOOK

Nom de la campagne, de l'outil ou de l'action de soutien (indiquer l'URL)	Descripti on 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 e t 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 mesures (si aucune, préciser les difficultés rencontrés)
Yo!Fest	Voir Question 6.1	EAVI ( <i>European Association for Viewers Interests</i> )	Sessions de formations dispensées par Facebook	Jeunes	EU	2020	N/A	Site internet Yo!Fest	N/A	N/A

# FACEBOOK

<p><b>Trois questions pour aider à éradiquer les fake news</b></p>	<p>Voir Question 6.1</p>	<p>Full Fact  Correctiv, Maldita, Pagella Politica</p>	<p>Facebook a lancé la campagne</p>	<p>N/A</p>	<p>EU, Middle East, Afrique</p>	<p>2020</p>	<p>N/A</p>	<p>N/A</p>	<p>16 millions d'utilisateurs Facebook en France</p>	<p>Hausse positive de 3,4 points chez les utilisateurs se souvenant activement avoir vu les annonces et hausse de 1,8 points chez les utilisateurs signalant un changement de comportement après avoir visionné la campagne</p>
--	--------------------------	--	-------------------------------------	------------	---------------------------------	-------------	------------	------------	--	---

# FACEBOOK

<b>Fonds pour le civisme en ligne</b>	Voir Question 6.1	ESJ LilleCivic Fab Tralalère Génération Numérique Bibliothèque sans frontière In Focus BondyBlog Dysturb Signes de Sens Unis Cité	Soutien financier Soutien à la communication Soutien méthodologique	ESJ Lille : 2000 étudiants + 300 enseignants Civic Fab : 180 jeunes Tralalère : 120 000 jeunes Génération Numérique : 13 000 jeunes de 9 à 13 ans Bibliothèque sans frontière : 10 000 jeunes + 500 bénévoles In Focus : 50 000 jeunes scolarisés	Toutes les régions françaises	De 2019 à aujourd'hui	2 millions euros	Communiqués de presse, sites internet dédiés, communication globale sur les réseaux sociaux, événements de lancement pour présenter les projets aux médias, aux universitaires et aux acteurs politiques français.	ESJ Lille : 2000 étudiants + 300 enseignants Civic Fab : 180 jeunes Tralalère : 120 000 jeunes Génération Numérique : 13 000 jeunes de 9 à 13 ans Bibliothèque sans frontière : 10 000 jeunes + 500 bénévoles	N/A
---------------------------------------	-------------------	---	---	--	-------------------------------	-----------------------	------------------	--	---	-----



# FACEBOOK

				<p>BondyBlog : +300 000 jeunes de 12 à 25 ans</p> <p>Dysturb : 12 experts médias /</p> <p>12 ambassadeurs influenceurs / 15 écoles en ZEP</p> <p>Signes de Sens : programme pilote pour les jeunes en situation de handicap, et les jeunes de 9 à 13 ans</p>					<p>In Focus : 50 000 jeunes scolarisés</p> <p>BondyBlog : +300 000 jeunes de 12 à 25 ans</p> <p>Dysturb : 12 experts médias /</p> <p>12 ambassade urs influenceur s / 15 écoles en ZEP</p> <p>Signes de Sens : programme pilote pour</p>
--	--	--	--	--	--	--	--	--	--

# FACEBOOK

				Unis Cité : 5000 jeunes dans +150 écoles					les jeunes en situation de handicap, et les jeunes de 9 à 13 ans  Unis Cité : 5000 jeunes dans +150 écoles	
<b>Chaire universitaire d'éducation aux médias</b>	Voir Question 6.1.	Ecole Supérieure de Journalisme de Lille	Soutien financier	Créer des outils et des programmes d'alphabétisation numérique dont l'efficacité et l'impact auprès des jeunes ont été prouvés et	France	Depuis février 2020	Appro x 140 0 00 euros	Relations avec les médias, événement de lancement à l'ESJ Lille, communication sur les réseaux sociaux.	N/A	N/A

# FACEBOOK

				<p>testés, qui seront rendus accessibles à tous les acteurs français, écoles et universités en France à la fin de l'année 1.</p> <p>Mieux comprendre et évaluer les attentes des nouvelles générations pour repenser les modèles éditoriaux des médias.</p>					
--	--	--	--	---	--	--	--	--	--