

**Loi relative à la Lutte contre la  
Manipulation de l'Information  
Déclaration annuelle au CSA pour 2019  
Facebook Ireland Limited**

---

23 avril 2020

## TABLE DES MATIERES

1.	Mise en œuvre d'un mécanisme de signalement facilement accessible et visible permettant aux utilisateurs de dénoncer les fausses informations, en particulier concernant le contenu sponsorisé par les tiers (article 11, alinéa 2 de la loi Fake News) .....	6
1.1	Description du mécanisme de signalement.....	6
1.2	Méthode de traitement des signalements des utilisateurs.....	12
1.3	Indicateurs et statistiques .....	13
1.3.1	Les faux comptes.....	14
1.3.2	Concernant les fausses informations.....	17
1.3.3	Concernant les <i>spams</i> .....	22
1.4	Transparence – Sensibilisation des utilisateurs aux Standards de la communauté et aux Règles de la communauté...23	
2.	Information des utilisateurs concernant la transparence des algorithmes (article 11, 1° de l'alinéa 3 de la loi Fake News)..25	
2.1	Remarques préliminaires sur la façon dont sont utilisées les technologies automatisées, telles que l'intelligence artificielle (IA) et nos modèles d'apprentissage automatique pour lutter contre la désinformation .....	25
2.2	Identification des fausses informations sur les photos et vidéos .....	27
2.3	Plus grande transparence et meilleur contrôle .....	30
2.3.1	Activité sur Facebook .....	30
2.3.1.1	Transparence des contenus publicitaires .....	30
2.3.1.2	Transparence des contenus organiques .....	32
2.3.2	Activité en dehors de Facebook.....	34
3.	La promotion des contenus issus des entreprises de presse, agences de presse et services de communication audiovisuelle (article 11, 2° de l'alinéa 3 de la loi Fake News).....	36
4.	La lutte contre les comptes diffusant de fausses informations (article 11, 3° de l'alinéa 3 de la loi Fake News) .....	37
4.1	Contexte général.....	37
4.2	Faux comptes .....	37
5.	Fournir des informations aux utilisateurs sur la nature, l'origine et les modalités de diffusion du contenu, ainsi que sur l'identité des personnes qui payent pour parrainer du contenu d'information (article 11, 4° et 5° de l'alinéa 3 de la loi Fake News)	38
5.1	Publicité à caractère politique et thématique .....	38
6.	La promotion de l'éducation aux médias et à l'information (article 11, 6° de l'alinéa 3 de la loi Fake News).....	40
7.	Les principales mesures mises en œuvre pour lutter contre la désinformation à partir de janvier 2020.....	42
7.1	Continuer à promouvoir l'éducation aux médias et à l'information .....	42
7.2	Fournir des informations exactes et limiter la désinformation sur le COVID-19 .....	42

## **Contexte de la Déclaration**

Aux termes de la loi n° 2018-1202 du 22 décembre 2018 relative à la Lutte contre la Manipulation de l'Information (la « **loi Fake News** »), les plateformes en ligne doivent prendre toutes mesures pour faire face à la diffusion de fausses informations susceptibles d'occasionner un trouble à l'ordre public ou d'altérer la sincérité d'un scrutin.

Les plateformes en ligne doivent notamment mettre en œuvre des mesures visant à :

- mettre en place un dispositif de signalement accessible et facilement visible permettant aux utilisateurs de signaler toute fausse information, notamment en ce qui concerne les contenus sponsorisés par des tiers ;
- assurer la transparence des algorithmes ;
- promouvoir le contenu des entreprises de presse, des agences de presse et des services de communication audiovisuelle ;
- lutter contre les comptes diffusant de fausses informations ;
- informer les utilisateurs sur la nature, l'origine et les modalités de diffusion des contenus, ainsi que sur l'identité des personnes qui paient pour sponsoriser du contenu informationnel ; et
- promouvoir l'éducation aux médias et à l'information.

Ces exigences ont été clarifiées par le *Conseil supérieur de l'audiovisuel* (« **CSA** ») dans ses lignes directrices publiées en mai 2019, et, par la suite, complétées par un questionnaire en février 2020.

Conformément à l'article 11 de la loi Fake News, Facebook Ireland Limited (« **Facebook Ireland** ») a rédigé la présente déclaration annuelle (la « **Déclaration** ») afin de présenter au CSA les mesures mises en œuvre pour lutter contre la désinformation, les faux comptes et les comportements trompeurs coordonnés.

### **Pour les besoins de la lecture de la Déclaration, nous vous prions de bien vouloir noter que :**

- 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.
- 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.
- La Déclaration porte sur les mesures mises en œuvre sur le service Facebook, sauf lorsque le service Instagram est expressément cité.
- Les mesures et indicateurs cités dans la Déclaration concernent l'année 2019 (dans la mesure où la Déclaration doit couvrir les activités de l'année précédente), sauf lorsque l'année 2020 est expressément citée.

### **Informations générales supplémentaires demandées par le CSA :**

- 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)). Les services sont disponibles en français.
- 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.
- Il y avait 36 millions d’utilisateurs français actifs par mois sur Facebook en 2019. Aujourd’hui il y en a 37 millions.
- 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 d’utilisation (par exemple si des utilisateurs ont une interdiction légale d’utiliser nos services).
- Le modèle économique de Facebook est décrit dans ses Conditions d’utilisation, accessibles [ici](#). Les utilisateurs ne paient pas pour l’utilisation de ces services. Au contraire, les entreprises et les organisations nous paient pour diffuser aux utilisateurs des publicités de leurs produits et services. Les Conditions d’utilisation d’Instagram sont accessibles [là](#).
- Les revenus de Facebook France en 2018 sont accessibles au public sur le site [Infogreffe.fr](http://Infogreffe.fr). Les indicateurs pour 2019 seront disponibles dans les semaines à venir sur ce même site.

\*\*\*\*\*

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

L’année dernière, 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 malintentionnés, la limitation de la diffusion des fausses informations et de la désinformation, ainsi qu’une 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 notre plateforme. C’est pourquoi notre stratégie de lutte contre la désinformation consiste en trois phases : supprimer, limiter, informer.

- Nous supprimons le contenu qui viole nos Standards de la communauté (Facebook) ou nos Règles de la communauté (Instagram), lesquels interdisent notamment, en matière de désinformation, les faux comptes, les *spams*, les vidéos manipulées et trompeuses, la pression sur les électeurs pour ne pas voter et la désinformation pouvant contribuer à causer un dommage physique.
- Nous limitons 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 le rétrogradant dans le fil d’actualité de Facebook et dans celui d’Instagram, ainsi que dans les *stories* Instagram.
- Nous informons nos utilisateurs en leur donnant davantage de contexte sur les contenus qu’ils souhaitent lire 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 :

- considérablement augmenté le nombre de nos partenaires de vérification indépendants et offert à nos utilisateurs davantage de contrôle sur leurs informations et le contenu qu’ils voient dans leur fil d’actualités.
- 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 six mois 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*) 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.
- amélioré 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.


\*\*\*\*\*

# 1. Mise en œuvre d'un mécanisme de signalement facilement accessible et visible permettant aux utilisateurs de dénoncer les fausses informations, en particulier concernant le contenu sponsorisé par les tiers (article 11, alinéa 2 de la loi Fake News)

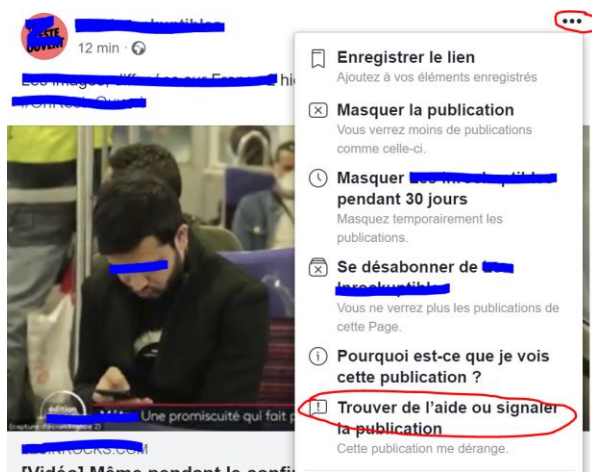
## 1.1 Description du mécanisme de signalement

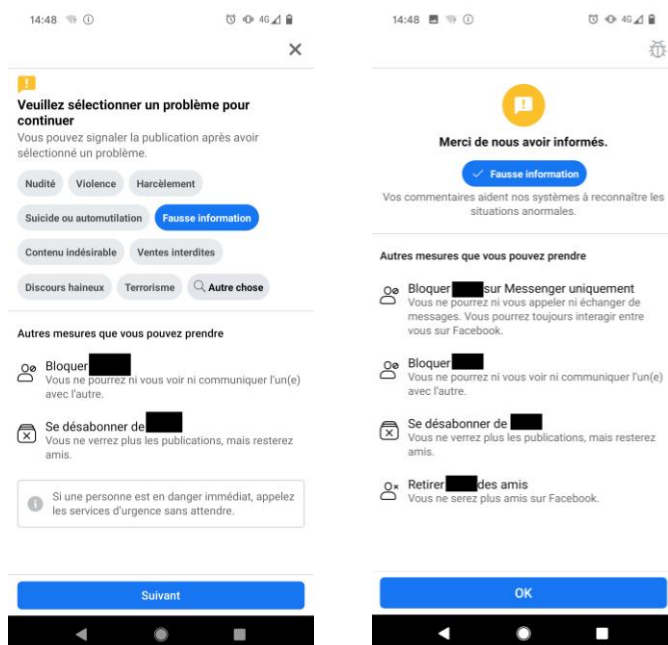
La communauté des utilisateurs Facebook joue un rôle important dans le signalement des fausses informations.

Pour signaler une publication comme une fausse information, l'utilisateur doit :

1. Cliquer sur  à côté de la publication qu'il/elle souhaite signaler comme fausse.
2. Sélectionner « Trouvez de l'aide ou signaler la publication ».
3. Cliquer sur « Fausse information ».
4. Cliquer sur « Envoyer ».


Ci-dessous des captures d'écran montrant l'expérience utilisateur :






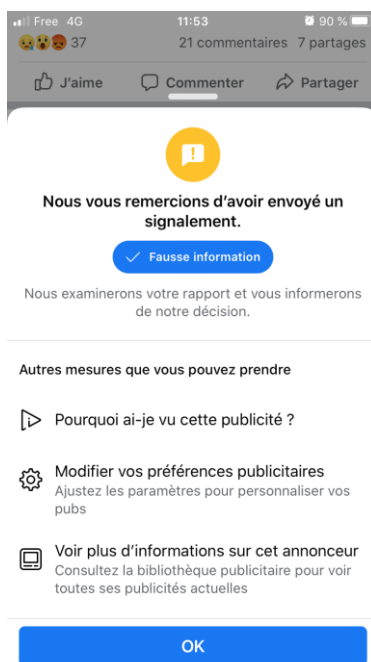
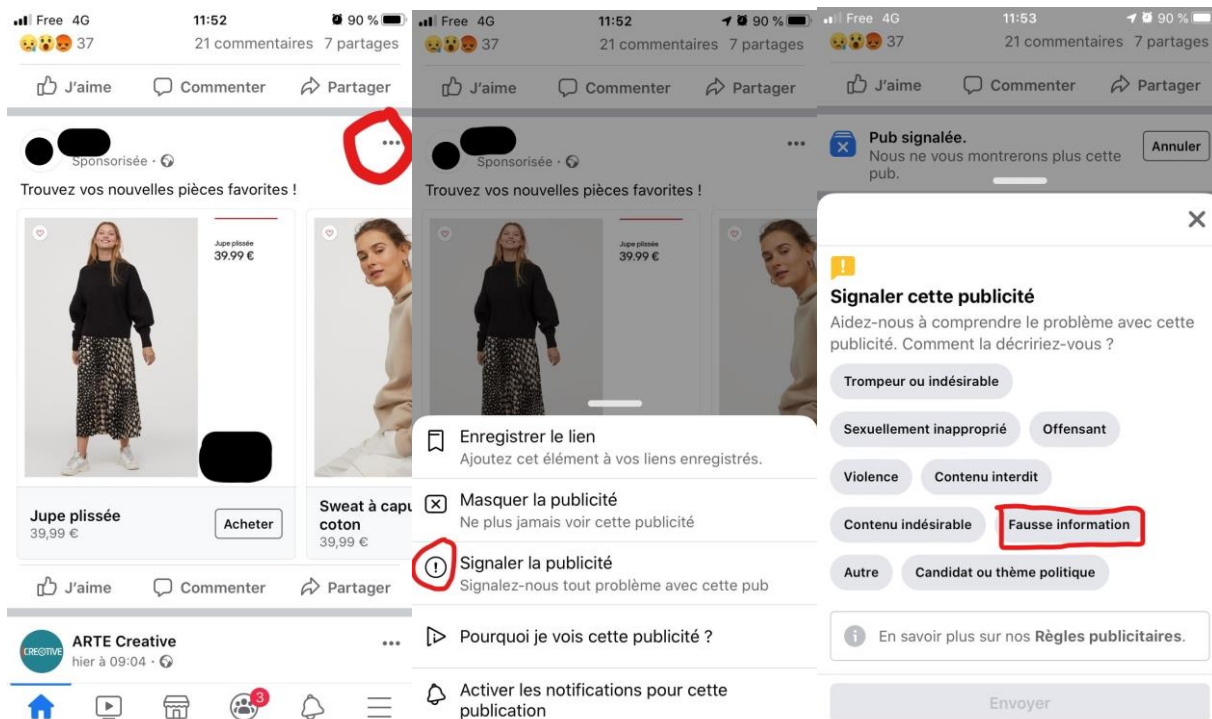
Pour signaler un courrier *spam*, la procédure est identique à celle décrite ci-avant. Au lieu de sélectionner « Fausse information » à l'étape 3, l'utilisateur sélectionnera alors « Contenu indésirable ».

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.

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 »).





Notre page d'aide intitulée « *Comment effectuer un signalement sur Facebook* » 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 sa page aide:

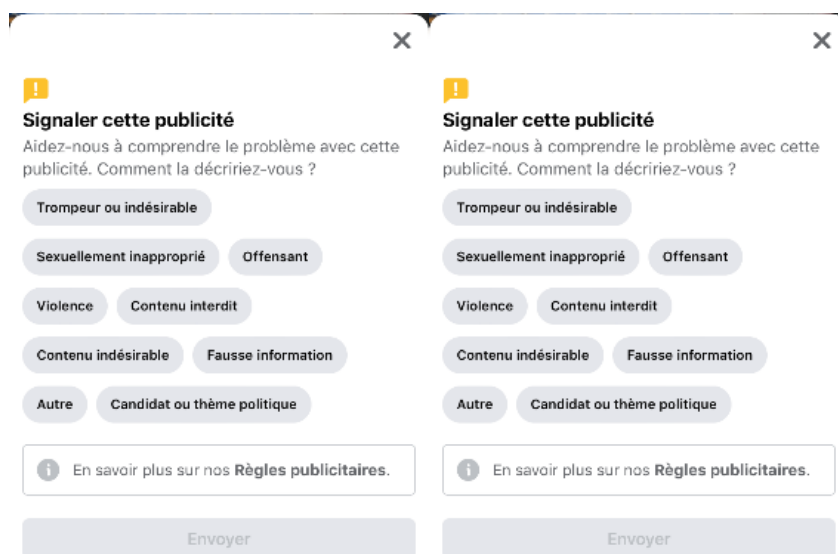
Pour signaler une fausse information sur une publication :

1. Cliquez sur **...** à côté de la publication que vous souhaitez signaler.
2. Sélectionnez **Trouver de l'aide ou signaler la publication**.
3. Cliquez sur **Fausse information**.
4. Cliquez sur **Envoyer**.

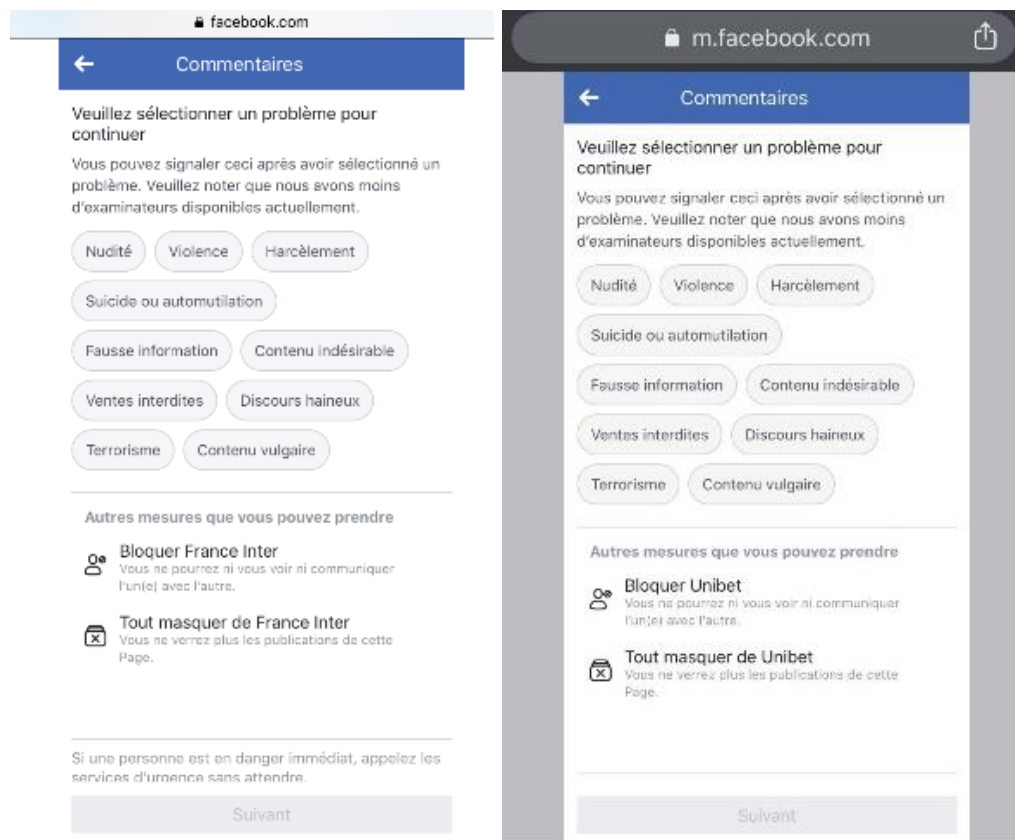
Facebook fournit les mêmes moyens pour signaler du contenu quel que soit le mode d'accès au service ou le type de contenu concerné.

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

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



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




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

**Signaler cette publicité** ✕

! Aidez-nous à comprendre le problème avec cette publicité. Comment la décrivez-vous ?

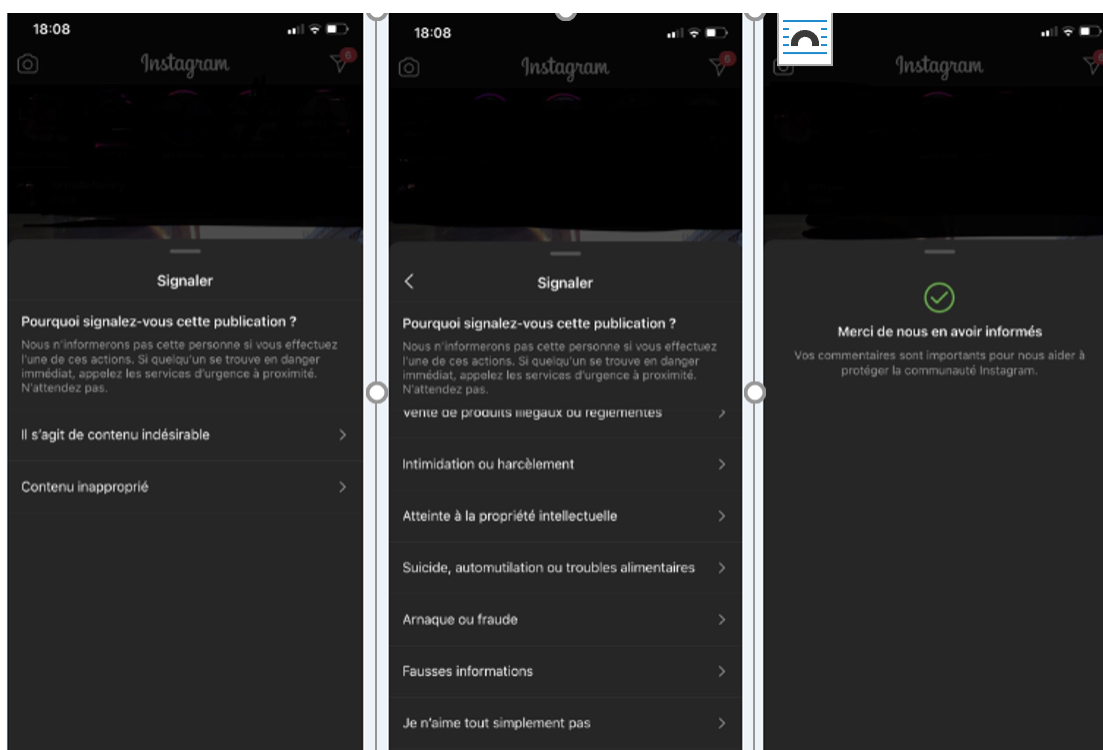
i En savoir plus sur nos [Règles publicitaires](#).

De la même façon, les utilisateurs peuvent signaler de fausses informations sur Instagram.

Lorsqu'un utilisateur voit un contenu sur Instagram susceptible, selon lui, d'être une fausse information, il peut nous le faire savoir en cliquant sur  situé à droite de chaque publication, puis sur « Signaler » puis :

-si l'utilisateur pense qu'il s'agit d'un spam, ce dernier peut cliquer sur « contenu indésirable »,

-si l'utilisateur pense qu'il s'agit d'une fausse information, ce dernier peut cliquer sur « Contenu inapproprié » puis « Fausses informations ».



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

- Capture d'écran de l'[Espace assistance de Facebook](#), qui affiche le statut des signalements de contenu :

Dans l'Espace Assistance, vous pouvez :

1. suivre l'évolution de votre signalement ;
2. consulter et répondre aux messages de l'équipe d'assistance ;
3. afficher les messages importants de votre compte.

- 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 Méthode de traitement des signalements des utilisateurs

Nous nous efforçons de répondre à tous les 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. À la fin de l'année 2019, l'équipe responsable de la sécurité sur Facebook et Instagram était constituée d'environ 35 000 personnes, dont environ 15 000 modérateurs de contenu 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 son [Rapport sur la Transparence](#), Facebook publie les [standards](#) et [directives](#) d'application utilisés par notre équipe « *Community Operations* ».

Si le contenu d'un utilisateur est supprimé en raison d'une violation de nos Standards ou Règles, l'utilisateur recevra une notification de la suppression, et pourra, dans la plupart des cas, 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.

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 décrite ci-dessus.

Nous avons mis en place un conseil de surveillance indépendant auprès duquel les utilisateurs peuvent contester les décisions prises par Facebook concernant un contenu. Initialement, toute personne en désaccord avec la décision de Facebook de suppression d'un contenu et qui a épuisé toutes les voies de recours disposera de 15 jours pour formuler un recours auprès du conseil de surveillance. Facebook pourra également soumettre directement les cas importants et difficiles au conseil. Finalement, Facebook souhaite permettre aux utilisateurs de soumettre des recours concernant le contenu qui demeure sur la plateforme également. Le conseil de surveillance pourra choisir parmi les cas admissibles soumis par Facebook ou un utilisateur et pourra déterminer si Facebook a pris la bonne décision.

**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, lesquels sont des violations de nos Standards ou Règles de la communauté sont, pour leur grande majorité, 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).
- 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 (*machine learning model*) 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 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 Indicateurs et statistiques

Comme il a été mentionné ci-dessus, Facebook associe les signalements des utilisateurs et les technologies d'apprentissage automatique avancées (*machine learning*) pour identifier et prendre des mesures concernant les contenus qui violent les Standards de la communauté (Facebook) ou les Règles de la communauté (Instagram), tout comme pour la désinformation. 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 quel que soit la manière dont ils ont été identifiés.**

### 1.3.1 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 (MAU), au niveau mondial, sur Facebook en 2019.

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



Dans les six mois compris entre le 4<sup>e</sup> trimestre 2018 et le 1<sup>er</sup> trimestre 2019, nous avons noté une forte augmentation de la création de faux comptes sur Facebook. Le nombre de faux comptes désactivés est passé d'1,2 milliard de comptes au 4<sup>e</sup> trimestre 2018 à 2,19 milliards au 1<sup>er</sup> trimestre 2019, en grande partie en raison de l'augmentation des attaques automatisées des acteurs malveillants qui tentent de créer une grande quantité de comptes en même temps. Ensuite, au 2<sup>e</sup> trimestre 2019, 1,5 milliard de faux comptes ont été désactivés et 1,7 milliard au 3<sup>e</sup> trimestre 2019. 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 avons constaté que bon nombre d'entre eux sont utilisés dans des campagnes de *spams* et sont créés pour des raisons financières.

Nous nous efforçons de supprimer les faux comptes qui abusent de nos plateformes 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 à 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.

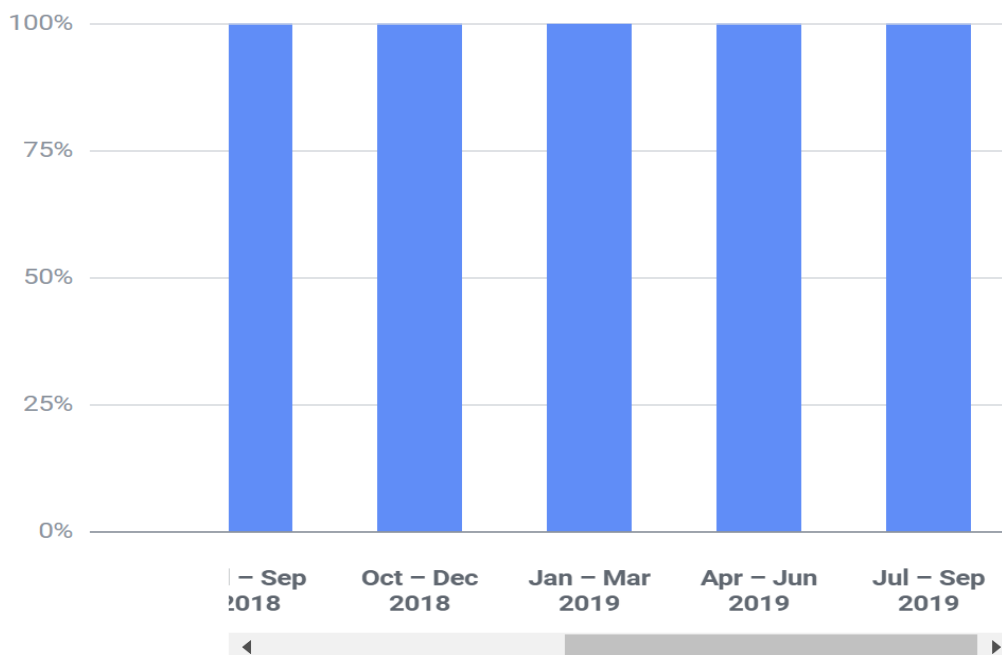
Les comptes que nous désactivons lors de leur création sont inclus dans nos statistiques sur le nombre de faux comptes supprimés. Ces statistiques peuvent varier en raison de la survenance d'attaques peu sophistiquées, ce qui est souvent le cas, à l'instar de celles que nous avons rencontrées au cours des deux derniers trimestres. 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 la plateforme. 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 99,8 % des comptes que nous supprimons ont été détectés avant tout signalement :



■ Percentage of accounts actioned that we found and flagged before users reported them

■ Percentage of accounts actioned that users reported first



### 1.3.2 Concernant les fausses informations

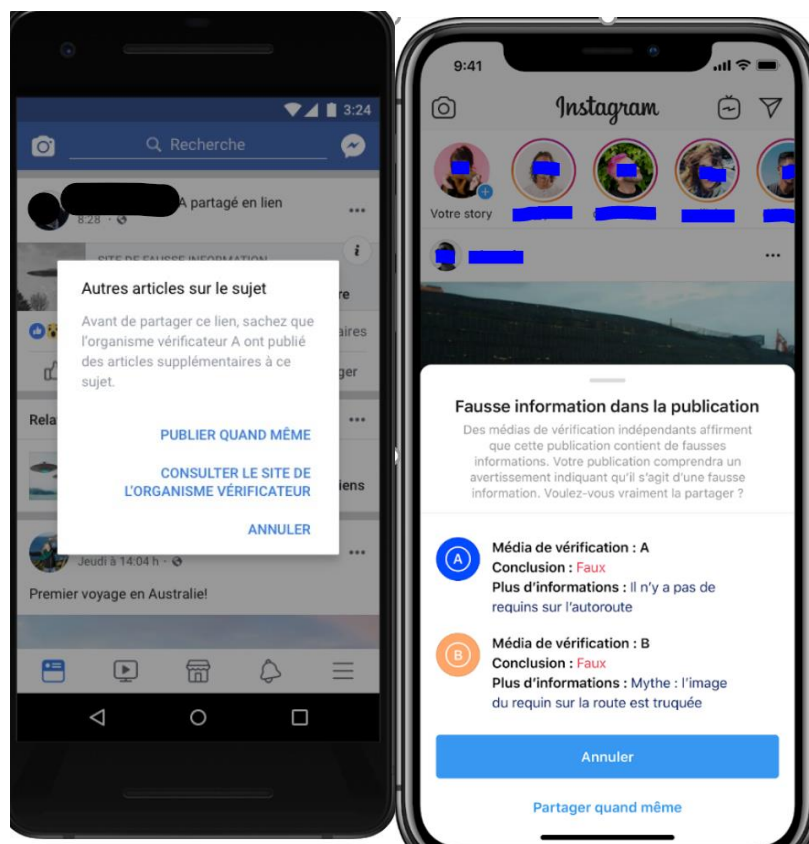
Nous considérons qu'une société technologique privée ne doit pas être l'arbitre de la vérité. C'est la raison pour laquelle nous avons des partenariats avec des experts externes et organisations tierces qui nous aident à distinguer le vrai du faux.

Nous considérons que la réduction de la diffusion de fausses informations – plutôt que leur suppression pure et simple (sauf si la fausse information en cause viole directement nos Standards ou Règles de la communauté en ce qu'elle causerait ou contribuerait à causer un dommage 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. Autrement dit, au nom de la liberté d'expression de nos utilisateurs, nous n'allons pas supprimer la fausse information publiée. En revanche, nous allons faire en sorte qu'elle apparaisse tout en bas du fil d'actualité des utilisateurs de Facebook ou d'Instagram. Cela permet de limiter le nombre de vues de ces contenus et de freiner leur viralité. Si un contenu est considéré comme faux sur Facebook, il le sera automatiquement sur Instagram (et vice versa).

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 (à gauche) et Instagram (à droite)



Trois [études](#) réalisées en 2018, par l'Université du Michigan et le journal français Le Monde, utilisant différentes méthodologies et définitions de fausses informations, ont démontré que la quantité globale de fausses informations sur Facebook a tendance à diminuer.

La [recherche](#) dirigée par Les Décodeurs (Le Monde) a analysé les interactions de média social (*social media engagement*) de Facebook, Twitter, Pinterest et Reddit sur 630 sites internet français, qu'ils ont qualifiés de « *sites peu fiables et douteux* » – le pire possible – « *sites les moins fiables* » et « *sites recommandables* ». Les Décodeurs ont jugé que les interactions de Facebook avec des « *sites peu fiables et douteux* » ont diminué de moitié en France depuis 2015. Les Décodeurs ont jugé que les interactions de Facebook avec les sites « *les moins fiables* » sur leur liste suivaient des courbes similaires de 2015 à début 2017, et étaient désormais à un niveau significativement plus faible qu'en 2015, alors que les sites « *plus recommandables* » avaient maintenu leur niveau d'interaction.

### 1.3.2.1 Suppression des fausses informations contribuant à causer un dommage physique

Concernant les fausses informations susceptibles de 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 dans le monde 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 connaissance des informations inexacts ou trompeuses pouvant contribuer à des dommages dans le monde réel. Nous avons établi des canaux pour que ces organisations nous contactent et nous transmettent de tels contenus rapidement.

### 1.3.2.2 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 récemment (en automne dernier) renforcé notre politique de lutte contre la pression sur les électeurs, soit contre les comportements visant à dissuader ou empêcher les utilisateurs de voter. Cette politique vise désormais :

- 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 (exemple : « *Si vous avez voté durant les primaires, votre vote pour l'élection générale ne sera pas comptabilisé.* ») ; et
- Les menaces de violence concernant le vote, l'inscription des électeurs ou le résultat d'une élection.

Nous supprimons ce type de contenu quel 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).

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 vérificateurs (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.

### 1.3.2.3 Vidéos truquées et trompeuses

Parfois, les utilisateurs partageant du contenu sur nos plateformes 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 augmente.

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 *deepfake*. Désormais, nous supprimerons les contenus vidéos manipulés et trompeurs s'ils remplissent les critères suivants :

- 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 *deepfake*, 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.

#### 1.3.2.4 Réduction des incitations économiques

Concernant la lutte contre les fausses informations, une des approches les plus efficaces consiste à supprimer les gains financiers pour les réseaux de désinformation. En effet, beaucoup de fausses informations sont partagées dans le but de réaliser des profits. Les *spammers* gagnent de l'argent en se faisant passer pour des éditeurs d'informations légitimes et en publiant des intox (*hoax*) pour que les utilisateurs visitent leurs sites qui sont en réalité, dans la majorité des cas, essentiellement constitués de publicités.

La diffusion des pages et domaines qui partagent de fausses informations de manière répétée sera réduite, et leur capacité à monétiser et à faire de la publicité sera suspendue. Ces récidivistes ne seront plus autorisés à faire de la publicité sur Facebook. Cette politique vise également les pages qui n'ont pas créé les fausses informations mais qui les partagent, car nous considérons qu'elles sont, tout comme les éditeurs, responsables du contenu qu'elles partagent avec leur audience. Une fois que les pages ne seront plus considérées comme des récidivistes, nous lèverons les mesures de réduction de diffusion et les autoriserons de nouveau à faire de la publicité et à monétiser. De la même façon, la visibilité des comptes Instagram partageant de fausses informations de manière répétée 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).

#### 1.3.2.5 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 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.

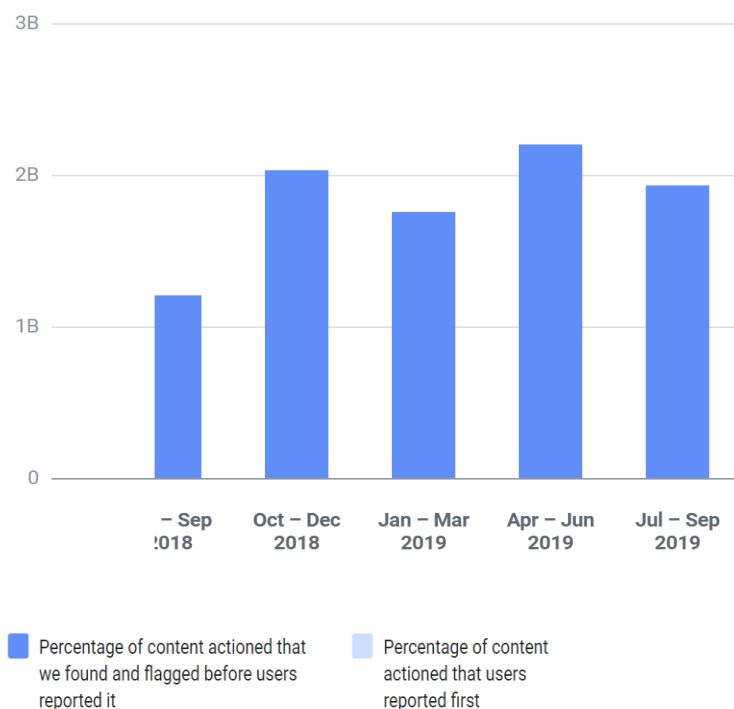
### 1.3.3 Concernant les *spams*

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 apprentissage automatique, en vue de détecter plus précisément les *spams* et de prendre toutes les mesures qui s'imposent.

Comme indiqué dans notre dernier [Rapport d'Application des Standards de la communauté](#), au cours du [4e trimestre 2018 et du 1er trimestre 2019](#), nous avons retiré 1,8 milliard de contenus qualifiés de spam (à titre de comparaison, nous avons agi sur 737 millions de contenus au 4<sup>e</sup> trimestre 2017, 836 millions au 1<sup>er</sup> trimestre 2018, 957 millions au 2<sup>e</sup> trimestre 2018 et 1,2 milliard au 3<sup>e</sup> trimestre 2018).

Au cours du 2<sup>e</sup> trimestre 2019, les chiffres réels révèlent une hausse des contenus qualifiés de spam que nous avons retirés dans la mesure où nos actions se sont renforcées, notamment en retirant des contenus spams publiés rétroactivement. Les contenus faisant l'objet de recours ont également chuté au 3<sup>e</sup> trimestre 2019 grâce à nos efforts pour assurer une meilleure cohérence dans la procédure de recours :

#### How much content did we take action on?



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, 99,9 % 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.

#### 1.4 **Transparence – Sensibilisation des utilisateurs aux Standards de la communauté et aux Règles de la communauté**

Pour s’inscrire et utiliser Facebook, un utilisateur doit accepter les Conditions d’utilisation. Ces Conditions d’utilisation 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 d’utilisation du service. Les Conditions d’utilisation intègrent les Standards de la communauté, ainsi que d’autres dispositions interdisant aux utilisateurs de publier du contenu illégal.

En particulier, l’article 3.2 des Conditions d’utilisation de Facebook (intitulé « *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) :*

*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 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, et, le cas échéant, nous pouvons prendre toute mesure à l’encontre de votre compte, pour les raisons susmentionnées. Nous pouvons également désactiver votre compte si vous violez à plusieurs reprises les droits de propriété intellectuelle de tiers. » (souligné par nos soins).*

De plus, l’article 4.2 des Conditions d’utilisation de Facebook (intitulé, « *Suspension ou résiliation 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 pourrions 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 de Facebook. Comme expliqué dans les Conditions d’utilisation, les Standards de la communauté présentent « [...] nos standards concernant le contenu que vous publiez sur Facebook, votre activité sur Facebook et autres produits Facebook ». Les Standards de la communauté interdisent également la publication de certains contenus afin de garantir la sécurité et le respect de tous sur 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 Règles de la communauté sont comprises dans les Conditions d'utilisation du service. Ces dernières interdisent aux utilisateurs de violer (ou d'aider ou d'inciter autrui à violer) les Conditions ou les politiques, notamment et en particulier les Règles de la communauté, la Politique de la plate-forme Instagram et les Règles applicables à la musique.

De même, sur Instagram, les Conditions d'utilisation prévoient que :

« Nous pouvons supprimer tout contenu ou toute information que vous partagez sur le Service si nous pensons qu'il ou elle enfreint les présentes Conditions d'utilisation ou nos politiques (y compris nos Règles de la communauté), ou si la loi nous y oblige. Nous pouvons immédiatement refuser de vous fournir ou arrêter de vous fournir tout ou une partie du Service (y compris résilier ou désactiver votre compte) si vous : enfreignez manifestement, gravement ou à plusieurs reprises les présentes Conditions d'utilisation ou nos politiques (y compris nos Règles de la communauté), si vous violez les droits de propriété intellectuelle d'autres personnes à plusieurs reprises ou si la loi nous y oblige. »

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. Elles sont rédigées et mises en œuvre par nos équipes internes formées dans une variété de domaines incluant l'application du droit, 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 actualisés et mis à la disposition du public à l'adresse suivante : <https://www.facebook.com/communitystandards/>. Les Règles de la communauté (Instagram) sont régulièrement actualisées et mis à la disposition du public à l'adresse suivante : <https://help.instagram.com/477434105621119>. Nous appliquons ces Standards et Règles, et nos conditions d'utilisation lorsque nous décidons de la suppression de contenu ou l'interdiction d'utilisateurs.



## 2. Information des utilisateurs concernant la transparence des algorithmes (article 11, 1<sup>o</sup> de l'alinéa 3 de la loi Fake News)

### 2.1 Remarques préliminaires sur la façon dont sont utilisées les technologies automatisées, telles que l'intelligence artificielle (IA) et nos modèles d'apprentissage automatique pour lutter contre la désinformation

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. 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.

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 exemples 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.

En outre, nos modèles d'apprentissage automatique nous aident 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 *clickbait*s, 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.

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

## 2.2 Identification des fausses informations sur les photos et vidéos

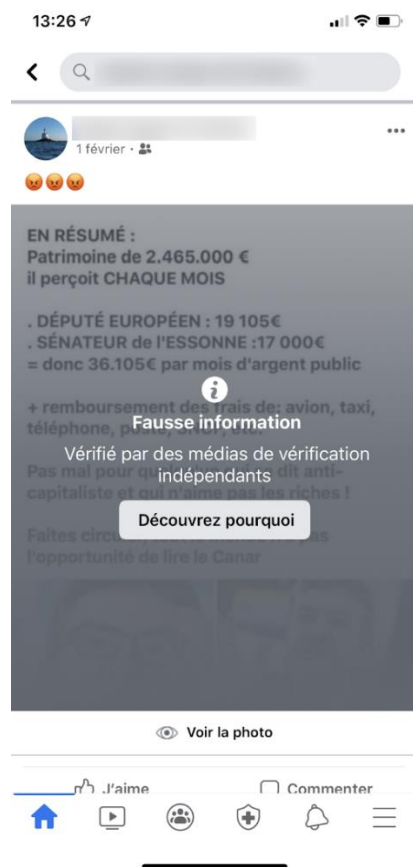
Il est essentiel de fournir aux utilisateurs davantage de contexte afin qu'ils puissent prendre des décisions réfléchies. A cet effet, nous agissons de différentes manières.

Premièrement, 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 indépendants et les encourage à bien réfléchir avant de partager un tel contenu (voir les captures d'écran insérés au 1.3.2 pour Facebook et Instagram).

Deuxièmement, lorsqu'un contenu est qualifié de faux ou partiellement faux par un partenaire de vérification indépendants, nous ajoutons des filtres bien visibles sur ce contenu, pour indiquer de façon claire à notre communauté que le contenu a été analysé comme faux ou partiellement faux. Ces filtres couvrent le 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 :



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 indépendants et des liens vers des articles provenant de sources fiables.

Troisièmement, nous affichons un 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. 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. Au printemps de cette année, nous avons étendu le bouton Contexte aux images contenant des informations ayant été vérifiées sur Facebook.

Voir capture d'écran ci-dessous :



Capture d'écran de ce qui apparaît après avoir cliqué sur le bouton « i » :

**À 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**

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

 L'ÉFIGARO.FR  
Coronavirus: Emmanuel Macron met en garde contre un déconfinement trop rapide

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

**PARTAGES**



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.

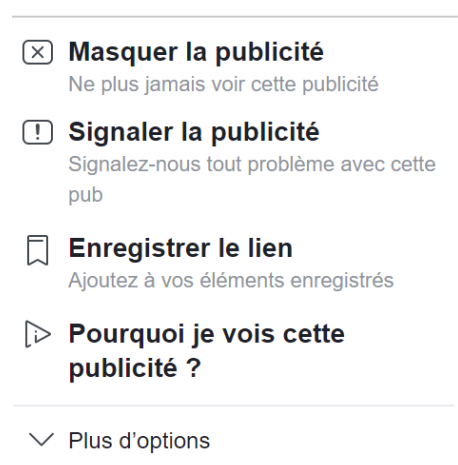
## 2.3 Plus grande transparence et meilleur contrôle

Que ce soit pour les nouveaux utilisateurs ou pour ceux qui utilisent depuis plusieurs années Facebook, 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 d'informations et de fonctionnalités, afin que nos utilisateurs puissent avoir davantage de contexte et de contrôle sur ce qu'ils voient sur Facebook.

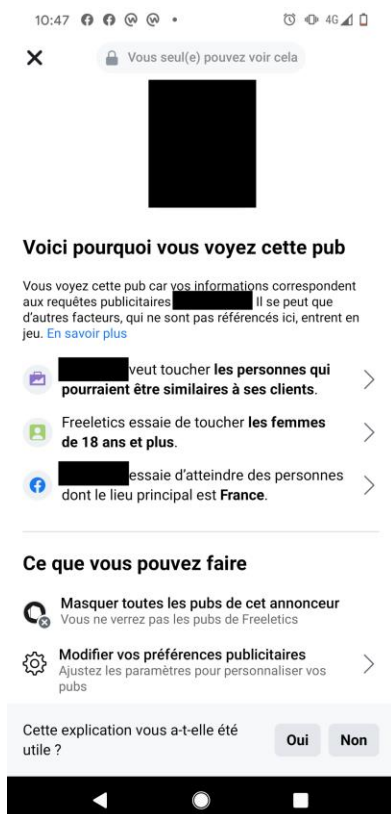
### 2.3.1 Activité sur Facebook

#### 2.3.1.1 Transparence des contenus publicitaires

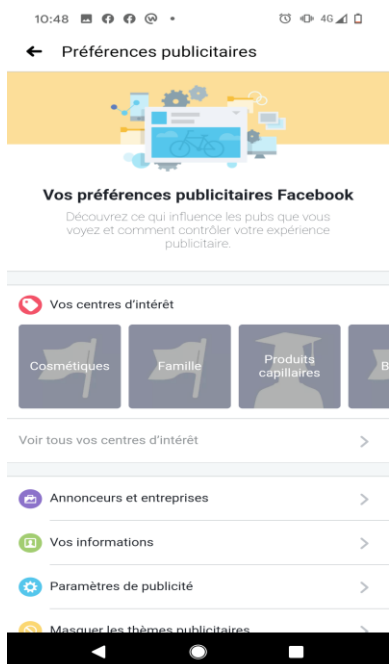
Nous avons créé un outil intitulé « *Pourquoi je vois cette publicité ?* » accessible dans le menu déroulant en cliquant sur  en haut à droite de chaque publicité (voir capture d'écran ci-dessous ce qui apparaît après avoir cliqué sur  ) :



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).




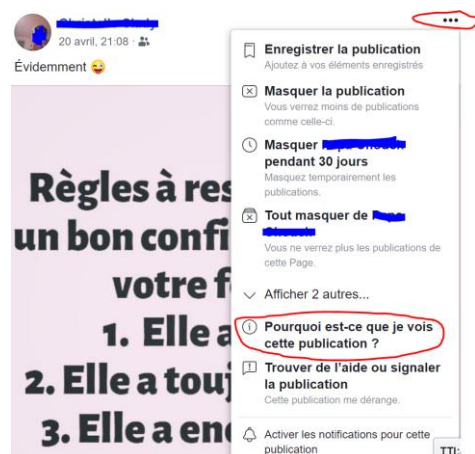
Dans la section « *Ce que vous pouvez faire* » (voir capture d'écran ci-dessus), l'utilisateur peut également choisir de masquer les publicités de cet annonceur ou modifier ses préférences publicitaires (Voir capture d'écran ci-dessous montrant ce qui apparaît lorsque l'utilisateur clique sur « *modifier ses 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.

### 2.3.1.2 Transparence des contenus organiques

De la même manière que pour les contenus publicitaires, 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 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).

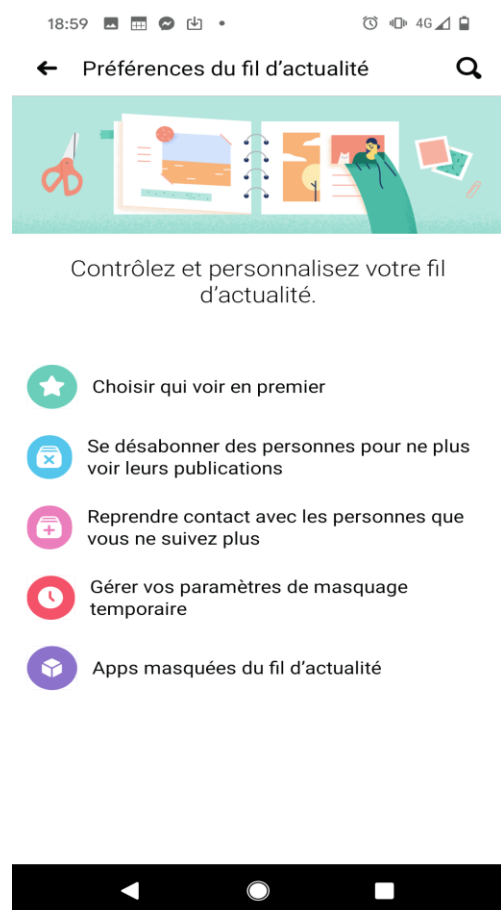


Voir la capture d'écran ci-dessous montrant ce qui apparaît après avoir cliqué sur « *Pourquoi est-ce que je vois cette publication* » :





En cliquant sur « *Modifier les préférences du 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 du fil d'actualité* ») :



L'utilisateur peut également modifier le référencement et la recommandation faite par les algorithmes du service 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) :



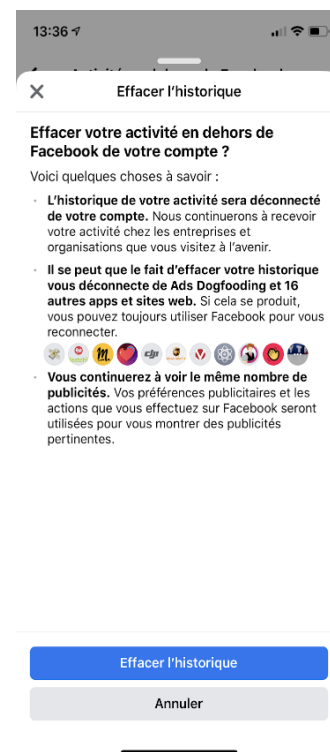
### 2.3.2 Activité en dehors de Facebook

Pour permettre aux utilisateurs de mieux comprendre et contrôler les données que les autres applications et sites internet partagent avec Facebook, les utilisateurs peuvent :

- Accéder à un résumé des informations que d'autres applications et sites internet ont envoyé à Facebook à l'aide de nos outils commerciaux en ligne, tels que Facebook Pixel ou Facebook Login ;
- Déconnecter ces informations de leur propre compte s'ils le souhaitent ; et
- Choisir de déconnecter de leur propre compte toute future activité en dehors de Facebook. Les utilisateurs peuvent choisir de le faire pour toute leur activité en dehors de Facebook, ou seulement pour des applications et sites internet spécifiques.

Depuis le 28 janvier 2020, l'outil « *Activité en dehors de Facebook* » est disponible dans le monde entier. Sur mobile (cf. captures d'écran ci-dessous), cet outil est accessible lorsque l'utilisateur clique sur « *Paramètres et confidentialité* », puis « *Paramètres* », puis « *Activités en dehors de Facebook* ».

Voir ci-dessous une copie d'écran de l'outil sur mobile :



### **3. La promotion des contenus issus des entreprises de presse, agences de presse et services de communication audiovisuelle (article 11, 2<sup>o</sup> de l'alinéa 3 de la loi Fake News)**

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. Nous commencerons bientôt à labelliser les médias qui sont sous le contrôle éditorial d'un État. Ce label figurera sur la page de ces médias et dans notre bibliothèque publicitaire et nous prévoyons également de faire de même sur Instagram.

Nous maintiendrons le plus haut niveau de transparence pour ces pages car elles conjuguent le pouvoir d'influence de l'opinion des médias avec le soutien stratégique d'un État.

Nous développons notre propre définition et nos propres règles pour les médias contrôlés par un État avec la collaboration 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 académiques, des organisations à but non lucratif et des organisations internationales de premier plan dans ce domaine, incluant Reporters sans frontières, le Centre international pour l'assistance internationale des Médias, (*Center for International Media Assistance*), le Centre européen de journalisme (*European Journalism Center*), le projet de *Oxford Internet Institute's Project* sur la propagande informatique (*Oxford Internet Institute's Project on Computational Propaganda*), le Centre pour les médias, les données et la société (*Center for Media, Data and Society*) à l'université d'Europe centrale, le Conseil de l'Europe, l'UNESCO et d'autres organisations.

Il est important de souligner que notre politique fait volontairement la distinction entre les médias contrôlés par un État et les médias publics, que nous définissons comme toute entité 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 État.

## 4. La lutte contre les comptes diffusant de fausses informations (article 11, 3° de l'alinéa 3 de la loi Fake News)

### 4.1 Contexte général

À ce jour, notre programme de vérification des informations compte plus de 55 partenaires dans plus de 45 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 en Espagne, en Pologne, au Portugal, en Grèce, en Irlande du Nord, en Croatie et en Lituanie. Notre programme inclut désormais 23 partenaires vérifiant des informations dans 17 langues européennes : croate, danois, néerlandais, anglais, français, allemand, grec, italien, letton, lituanien, norvégien, polonais, portugais, russe, espagnol et suédois.

Facebook a choisi de confier la vérification des informations publiées sur sa plateforme à des médias spécialisés et indépendants. Pour conclure un tel partenariat avec Facebook, un média doit être membre de l'*International Fact-Checking Network* (IFCN). Pour devenir membre de l'IFCN, le dossier de candidature est examiné par des tiers indépendants et l'adhésion du média doit être renouvelée chaque année.

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

### 4.2 Faux comptes

La prévention de l'apparition des faux comptes est un moyen de mettre fin aux abus sur nos plateformes. Nous concentrons nos efforts en luttant contre les faux comptes afin de prévenir tout dommage et d'éviter de prendre des mesures par erreur concernant des vrais comptes.

Voir la section 1.3.1 pour les indicateurs relatifs aux faux comptes et la méthodologie utilisée pour lutter contre ces faux comptes.

## 5. Fournir des informations aux utilisateurs sur la nature, l'origine et les modalités de diffusion du contenu, ainsi que sur l'identité des personnes qui payent pour parrainer du contenu d'information (article 11, 4<sup>o</sup> et 5<sup>o</sup> de l'alinéa 3 de la loi Fake News)

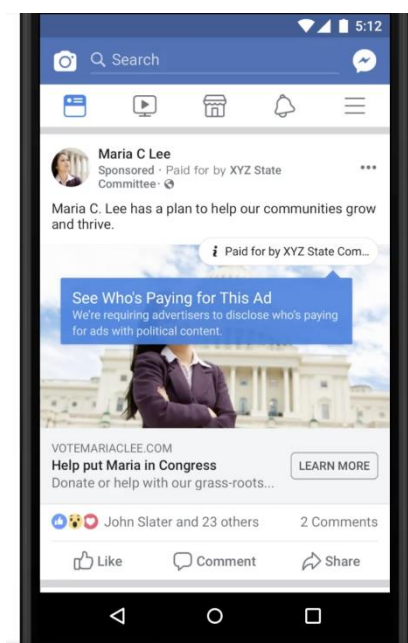
### 5.1 Publicité à caractère politique et thématique

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 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* ». 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.



- **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 à caractère 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 politiques 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 à caractère 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 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](#).

## 6. La promotion de l'éducation aux médias et à l'information (article 11, 6° de l'alinéa 3 de la loi Fake News)

Une partie de notre travail pour combattre la diffusion de fausses informations consiste à aider les utilisateurs à les identifier par eux-mêmes. C'est la raison pour laquelle nous avons conclu des partenariats avec des organisations et experts en matière d'éducation aux médias. Nous souhaitons aider les utilisateurs à prendre des décisions éclairées sur les informations qu'ils lisent et leur permettre de choisir ce qu'ils souhaitent lire, croire et partager. Nous faisons ceci en partenariat avec des organisations locales, en parrainant des programmes d'éducation aux médias et en partageant des conseils pour aider à identifier les fausses informations.

**Pour les élections européennes**, nous avons soutenu nos partenaires de vérification indépendants au sein de l'Union Européenne pour le lancement d'une campagne d'éducation aux médias lancée dans les 28 États membres et nous leur avons fourni des conseils sur comment éradiquer les fausses informations. Nous continuerons à mener cette campagne au sein de l'Union Européenne à l'occasion des prochaines élections. Selon les informations fournies par nos agences partenaires – et qui n'ont pas été vérifiées indépendamment par Facebook – cette campagne a eu plus de 100 millions de vues au sein de l'UE, et nos partenaires continuent à mener cette campagne là où de prochaines élections auront lieu.

**Nous avons également conclu un partenariat avec l'Association Européenne pour les intérêts des Téléspectateurs (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 sessions de formation à l'éducation aux médias et une session de question-réponse sur l'éducation aux médias.

**Concernant les élections municipales françaises**, nous avons engagé des discussions avec des entreprises et acteurs offrant des services aux élus locaux sur la manière de garantir l'intégrité des élections municipales françaises en organisant un panel intitulé « *Comment protéger l'intégrité des élections municipales 2020 - l'exemple de Facebook* » au salon des maires et des collectivités Locales le 20 novembre 2019.

**Enfin, nous avons conclu un partenariat avec Freeformers et plus de 20 ONG nationales et organismes de formation** pour offrir un programme de formation aux compétences numériques 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 s'assurer de leur sécurité en ligne.

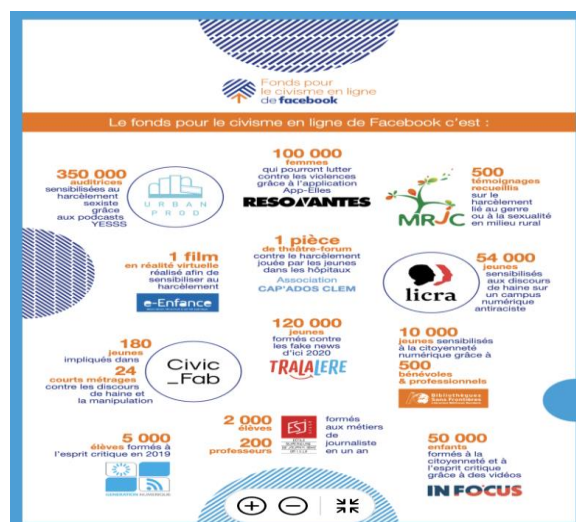


**Ressources sur le numérique :** L'automne dernier, nous avons lancé les Ressources sur le numérique (*Digital Literacy Library* - DLL) dans 45 langues dans le monde entier. Les DLL sont une collection de cours prêts à l'emploi fournis par l'équipe Jeunesse et Média (*Youth and Media*) du *Berkman Klein 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 :** En mai dernier, 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.

**Le Fonds pour le Civisme en ligne :** ce fonds vise à soutenir une douzaine de projets français de lutte contre les discours haineux, le cyber-harcèlement et la promotion de l'éducation aux médias grâce à des campagnes nationales hors ligne, en leur accordant des subventions comprises entre 20 000 et 200 000 euros. Un panel d'experts français comprenant des intellectuels, psychologues, influenceurs et représentants du Gouvernement est responsable de la sélection des projets.



## 7. Les principales mesures mises en œuvre pour lutter contre la désinformation à partir de janvier 2020

### 7.1 Continuer à promouvoir l'éducation aux médias et à l'information

- Le Fonds pour le Civisme en ligne

Comme indiqué ci-dessus ce Fonds a notamment été mis en place pour soutenir des projets œuvrant à la promotion du contre discours, du développement à l'esprit critique ou de toute initiative aidant à la promotion du bon usage du numérique. Ce Fonds sera renouvelé en 2020 vous pouvez retrouver l'ensemble des informations [ici](#).

- 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 seront accessibles à tous les intervenants, écoles et universités français à la fin de la première année. Dans les mois à venir, l'ESJ de Lille annoncera des partenariats avec d'autres laboratoires de recherche étrangers basés au Canada, en Belgique et dans d'autres pays afin de mettre en commun et d'intensifier les efforts et le travail de recherche dirigé par la chaire universitaire.

### 7.2 Fournir des informations exactes et limiter la désinformation sur le COVID-19

Facebook soutient les efforts de la communauté mondiale de la santé publique pour préserver la santé des personnes et assurer leur bonne information durant la crise de santé publique du Coronavirus.

- **Mettre à disposition des individus des informations exactes et des ressources utiles des autorités sanitaires pertinentes**

Notre grande priorité est de nous assurer que chacun dispose d'informations exactes. Nous mettons en avant les informations officielles et vérifiées telles que celles du site internet [www.gouvernement.fr](http://www.gouvernement.fr) en étroite collaboration avec le service d'information du Gouvernement, l'OMS et les autorités sanitaires locales. Par exemple, lorsque vous tapez « Coronavirus » dans la barre de recherche, les sites internet du gouvernement apparaîtront en premier. En outre, nous avons facilité la mise en place des robots de discussion ou encore « chatsbot » pour faciliter l'accès des citoyens aux informations vérifiées des autorités publiques et sanitaires françaises.

Nous avons également fourni un module d'information reprenant les informations officielles dans le fil d'actualité de Facebook pour qu'elles soient vues par tous les utilisateurs. Enfin, nous avons créé un centre d'information sur le coronavirus en haut du fil d'actualité pour fournir un espace central où les personnes peuvent voir les dernières informations (mises à jour en temps réel par les autorités sanitaires nationales et les organisations internationales) ainsi que les ressources et recommandations aux utilisateurs pour rester en bonne santé et soutenir leurs familles et communautés. A titre de résultat nous avons redirigé jusqu'à présent près de 2 milliards d'utilisateurs dans le monde vers des informations émanant des autorités de santé publique, à travers notre centre d'information Covid-19. Plus de 350 millions d'utilisateurs ont cliqué sur nos messages pédagogiques pour en savoir plus sur Facebook et Instagram.

Nous affichons également des messages dans le fil d'actualité des personnes qui ont aimé, réagi ou commenté des informations nuisibles liées au Covid-19, que nous avons depuis supprimées. Ces messages renvoient nos utilisateurs vers les contre-vérités autour du Covid-19 qui ont été déconstruites par l'Organisation Mondiale de la Santé. Par cette démarche, nous souhaitons permettre aux utilisateurs qui ont peut-être interagi avec ces fausses informations, de voir la vérité émanant d'une source faisant autorité, en l'occurrence l'Organisation Mondiale de la Santé.

- La suppression sur Facebook et Instagram des fausses informations sur le COVID-19 qui peuvent entraîner des dommages physiques

Alors que les autorités sanitaires mondiales publient de nouvelles directives et mises en garde concernant le COVID-19, nous continuons à limiter la diffusion de fausses informations sur le virus sur nos plateformes.

Nous supprimons le contenu qui contient de fausses allégations ou théories du complot qui ont été dénoncées par les principales organisations sanitaires mondiales et les autorités sanitaires locales et qui pourraient entraîner des dommages pour les personnes qui y croiraient. Nous menons ces actions conformément à nos politiques existantes de suppression du contenu pouvant causer un dommage physique. Ainsi au cours du mois de mars 2020, nous avons supprimé des centaines de milliers de contenus liés au Covid-19 qui ne respectaient pas nos standards relatifs aux fausses informations.

Nous nous concentrons sur les déclarations qui visent à dissuader le recours à un traitement ou de prendre les précautions appropriées. Cela comprend les affirmations liées aux faux traitements ou moyens de prévention – par exemple prétendre que la consommation d'eau de javel guérirait le coronavirus – les affirmations qui créent une confusion quant aux ressources sanitaires ou services d'utilité publique disponibles, ou à l'emplacement ou la gravité de l'épidémie. Nous bloquons ou limitons également les hashtags utilisés pour diffuser des fausses informations sur Instagram et réalisons des recherches proactives afin d'identifier et de supprimer ce type de contenu autant que possible.

Notre approche est dynamique et évoluera avec les nouvelles orientations en matière de santé publique - par exemple, nous supprimons désormais les affirmations selon lesquelles la distanciation sociale ou physique ne fonctionne pas.

- Soutien de nos partenaires de vérification indépendants.

Toutes les fausses informations en matière de santé qui ne relèvent pas d'une opinion, d'une satire ou d'un discours (non revu) d'une personnalité politique peuvent être vérifiées conformément à notre politique applicable.

Nos partenaires de vérification des informations indépendants en France continuent à examiner le contenu et à démentir les affirmations fausses concernant le coronavirus. A titre d'exemple, nous avons affiché des avertissements sur plus de 40 millions de publications Facebook, sur la base de 4 000 évaluations de nos partenaires *fact-checkers*. Lorsque les utilisateurs voient ces étiquettes de mise en garde, dans plus de 95% des cas, ils ne vont pas plus loin et ne consultent pas le contenu.

De plus, nous avons conclu un partenariat avec l'*International Fact-Checking Network (IFCN)* du *Poynter Institute* afin de lancer un programme d'une subvention d'1 million de dollars pour soutenir les partenaires

de vérification indépendants dans leur travail concernant le COVID-19. En plus de fournir un financement important permettant à nos partenaires de maintenir ou d'augmenter leur capacité durant cette période, les subventions soutiendront également les propositions telles que la traduction des vérifications des informations de la langue d'origine vers d'autres langues, la production de contenus multimédias (vidéos, infographie, podcasts) sur le COVID-19 et la collaboration avec des experts de la santé pour assurer la prise en compte de la dimension scientifique.

Enfin, en Avril 2020, nous avons mis en avant sur la page de Facebook France et dans les fils d'actualité d'un grand nombre d'utilisateurs en France des vidéos pédagogiques sur les fausses informations relatives au Covid-19, réalisée par les journalistes de l'AFP. Chaque épisode d'une durée d'environ trois minutes retrace l'ADN d'une fausse information, de sa création à sa diffusion, et explique le travail des vérificateurs d'informations de l'AFP. Les équipes y analysent une fausse information liée au Covid-19, expliquent comment elle fonctionne et donnent des conseils pour détecter les fausses informations et aider les gens à mieux comprendre les informations qu'ils voient. Par exemple, dans le premier épisode, un vérificateur d'informations de l'AFP nous explique pourquoi l'Institut Pasteur n'a pas créé le Covid-19.

Vous pouvez consulter toutes nos actions dans le cadre de la lutte contre le Coronavirus, qui sont régulièrement actualisées, [ici](#) pour Facebook et [là](#) (en anglais) pour Instagram.