

# THÉMA



**Lutte contre la diffusion de  
fausses informations sur  
les plateformes en ligne :**  
*bilan de l'application et de l'effectivité des mesures  
mises en œuvres par les opérateurs en 2019*

Juillet 2020



## Introduction – cadre et méthode

*Le titre III de la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information prévoit un devoir de coopération dans la lutte contre la diffusion de fausses informations pour les opérateurs de plateforme en ligne<sup>1</sup> dépassant un seuil de connexion de 5 millions d'utilisateurs uniques par mois, par plateforme, calculé sur la base de la dernière année civile<sup>2</sup>.*

*Dans cet objectif, ils sont tenus de mettre en œuvre un dispositif de signalement de telles informations accessible et visible. Ils doivent également prendre des mesures complémentaires qui peuvent relever de la transparence des algorithmes, de la promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle, de la lutte contre les comptes propageant massivement de fausses informations, d'information des utilisateurs sur les contenus sponsorisés d'information se rattachant à un débat d'intérêt général et de l'éducation aux médias et à l'information.*

*La loi prévoit que ces mesures et les moyens que les opérateurs concernés y consacrent soient rendus publics et fassent l'objet d'une déclaration annuelle au CSA. Elle charge celui-ci de publier un bilan périodique de leur application et de leur effectivité, en vue duquel il peut recueillir les informations nécessaires à son élaboration auprès des opérateurs concernés<sup>3</sup>. Elle lui donne également la compétence d'adresser des recommandations aux opérateurs concernés visant à améliorer la lutte contre la diffusion de fausses informations. A cette fin, le Conseil a publié une recommandation le 15 mai 2019.*

*En vue du bilan de la première année d'exercice, 2019, un questionnaire basé sur cette recommandation a été adressé aux opérateurs et mis en ligne sur le site du CSA en février 2020<sup>4</sup>. Le délai de remise de la déclaration, prévu le 31 mars 2020 conformément à la recommandation du CSA, a été reporté au 23 avril 2020 afin de tenir compte des difficultés liées aux circonstances sanitaires. A cette date, le CSA a reçu dix déclarations ; une onzième, celle de Unify (pour Doctissimo et Aufeminin.com<sup>5</sup>), lui a été adressée le 25 juin 2020.*

*À l'occasion de la publication du présent bilan et comme il l'avait annoncé dans son questionnaire, le CSA rend publiques les déclarations sur son site internet (sauf éléments couverts par le secret des affaires). Les résumés des mesures prises présentés dans le présent bilan ne sauraient refléter la richesse des informations fournies dans les déclarations des opérateurs,*

---

<sup>1</sup> Tels que définis en ces termes à l'article L 111-7 du Code de la consommation : « I.- Est qualifiée d'opérateur de plateforme en ligne toute personne physique ou morale proposant, à titre professionnel, de manière rémunérée ou non, un service de communication au public en ligne reposant sur :

1° Le classement ou le référencement, au moyen d'algorithmes informatiques, de contenus, de biens ou de services proposés ou mis en ligne par des tiers ;

2° Ou la mise en relation de plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service. (...) ».

<sup>2</sup> Seuil fixé par l'article 1 du Décret n° 2019-297 du 10 avril 2019 relatif aux obligations d'information des opérateurs de plateforme en ligne assurant la promotion de contenus d'information se rattachant à un débat d'intérêt général.

<sup>3</sup> Les dispositions légales applicables sont rappelées en annexe 1.

<sup>4</sup> Questionnaire disponible en annexe 2.

<sup>5</sup> Les services Doctissimo et Aufeminin sont édités par des sociétés différentes appartenant au groupe TF1 qui a regroupé ses activités numériques sous la marque Unify.



*auxquelles le CSA invite le lecteur à se reporter pour prendre la pleine mesure des moyens mis en œuvre.*

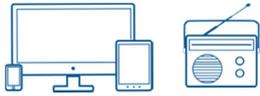
*L'instruction des déclarations, en mai et juin 2020, a notamment amené à demander des compléments et d'éclaircissements aux opérateurs, auxquels certains ont répondu. Les données issues de ces échanges ultérieurs à la déclaration sont mentionnées comme telles dans le présent bilan.*

*Le comité d'experts sur la désinformation en ligne auprès du CSA<sup>6</sup> s'est réuni le 29 juin 2020 afin de fournir son analyse, de mettre en perspective les informations reçues et de réfléchir à des pistes de recommandations. Leur contribution a permis de nourrir la réflexion du CSA, dont les conclusions du présent bilan relèvent toutefois de la seule responsabilité.*

*Plusieurs plateformes ont mentionné les dispositifs mis en œuvre pour lutter contre la diffusion de fausses informations dans le cadre de la crise du COVID-19. Ces éléments sont abordés dans le présent bilan, mais non pris en compte dans l'analyse qui concerne l'application et l'effectivité des mesures mises en place au cours de l'année 2019. En outre, tous les opérateurs ne les ayant pas mentionnés, leur présentation n'a pas vocation à être exhaustive.*

---

<sup>6</sup> Voir composition en annexe 3.



## Synthèse

Les opérateurs de plateforme en ligne soumis au devoir de coopération prévu par le **titre III de la loi du 22 décembre 2018** relative à la lutte contre la manipulation de l'information se sont livrés pour la première année à l'exercice de déclaration au CSA des moyens mis en œuvre, en 2019, afin de **lutter contre la diffusion de fausses informations**.

En vertu de l'article 11 de cette loi, les opérateurs sont tenus de mettre en œuvre un dispositif de signalement ainsi que des mesures complémentaires qui peuvent relever de la transparence des algorithmes, de la promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle (ci-après EAP-SCA), de la lutte contre les comptes propageant massivement de fausses informations, de l'information des utilisateurs sur les contenus sponsorisés d'information se rattachant à un débat d'intérêt général et de l'éducation aux médias et à l'information (EMI). Dans sa **recommandation du 15 mai 2019**, le CSA a publié des lignes directrices concernant les moyens de mise en œuvre de ces dispositifs.

Le présent document dresse un bilan de l'application et de l'effectivité des mesures mises en œuvre par les opérateurs, conformément aux dispositions du 4<sup>ème</sup> alinéa de l'article 17-2 de la loi du 30 septembre 1986 relative à la liberté de communication résultant de la loi du 22 décembre 2018 précitée.

Le calendrier initialement fixé par le Conseil pour ce premier exercice a été modifié du fait de la crise sanitaire de la COVID-19. Le délai de remise des déclarations a été repoussé au **23 avril 2020**<sup>7</sup>, date à laquelle la majorité des opérateurs ont effectivement adressé ce document au CSA, en s'appuyant notamment sur le questionnaire publié par ce dernier en février 2020. En outre, plusieurs opérateurs ont souhaité communiquer au Conseil des informations complémentaires relatives aux dispositifs spécifiques mis en œuvre dans le contexte de la crise sanitaire. Le présent bilan est fondé sur l'ensemble de ces informations, ainsi que sur les échanges complémentaires menés entre les services du Conseil et les opérateurs et sur les apports du comité d'experts sur la désinformation en ligne.

**Onze** opérateurs ont adressé une déclaration concernant, au total, **seize** services. Il s'agit de services de classement et référencement de contenus, biens ou services mis en ligne par des tiers et de services de mise en relation de plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service, ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service. Dans son analyse, le CSA a tenu compte de la diversité des plateformes, tant au regard de leur activité et de leurs modèles économiques que de leurs fréquentation et de leur chiffre d'affaires.

Le CSA tient à saluer le **degré élevé de coopération des opérateurs** et la qualité du dialogue noué avec eux. **La plupart d'entre eux ont à cet égard désigné un interlocuteur référent**, conformément aux dispositions de l'article 13 de la loi du 22 décembre 2018.

---

<sup>7</sup> Au lieu du 31 mars comme prévu à l'article 7 de la recommandation du 15 mai 2019.



La quantité et la qualité globales des informations déclarées est notable. Toutefois, le niveau de précision des réponses est hétérogène. En règle générale, **peu d'informations ont été fournies sur les moyens humains et financiers déployés pour lutter contre la manipulation de l'information et sur l'intelligibilité<sup>8</sup> des algorithmes**. Bien que **le secret des affaires** ait pu légitimement être invoqué dans certains cas, le CSA rappelle qu'il avait donné la possibilité aux opérateurs de lui signaler la confidentialité des éléments qu'ils ne souhaitaient pas voir publiés.

Ce défaut d'informations obère la capacité d'appréciation par le CSA des dispositifs mis en place. Le Conseil, confiant en leur **esprit de responsabilité**, demande donc aux opérateurs de fournir à l'avenir une déclaration plus complète, notamment au regard de son questionnaire.

A l'issue de ce travail d'analyse, le CSA constate d'une manière générale que les opérateurs se sont saisis de l'enjeu de la lutte contre la diffusion de fausses informations, en suivant plusieurs de ses recommandations ; néanmoins, ces mesures doivent encore être approfondies.

## 1. Le dispositif de signalement de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un scrutin

Les informations relatives aux moyens mis en œuvre autour du dispositif de signalement de fausses informations ont été particulièrement développées dans les déclarations.

Tous les opérateurs, à une exception près<sup>9</sup>, ont mis en place un dispositif de signalement, à l'ergonomie variable. La majorité de ces outils sont à proximité immédiate du contenu, mais il est parfois nécessaire de remplir un formulaire, ce qui peut rendre la démarche plus contraignante. En outre, **si certaines plateformes ont ajouté une catégorie spécifique permettant de signaler un contenu pour « fausse information », d'autres ne l'ont pas prévu** et rattachent le cas échéant ces signalements à la catégorie « autres ». Deux arguments sont avancés pour justifier ce choix : d'une part, certains opérateurs ne s'estiment pas légitimes pour qualifier un contenu de « fausse information » ; d'autre part, la présence d'une catégorie « fausse information » susciterait un afflux de signalements qui ne seraient pas factuellement justifiés.

---

<sup>8</sup> L'intelligibilité est entendue ici comme « explicabilité » des algorithmes, notion qui renvoie à la publicité d'éléments relatifs à la finalité des traitements pour lesquels les algorithmes ont été programmés, afin d'informer les utilisateurs auxquels ces traitements s'appliquent. Les règles définissant ces traitements, les principales caractéristiques de leur mise en œuvre, les données traitées et leurs sources, les paramètres de traitement et leur pondération, les opérations effectuées par le traitement et, le cas échéant, le degré et mode de contribution à la décision de, par exemple de proposer un contenu plutôt qu'un autre, font partie des éléments communicables en vertu de ce principe. Celui-ci a notamment été cité dans le cadre des relations administrations – usagers par le décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique, pris en application de l'article L. 311-3-1 du code des relations entre le public et l'administration, tel que modifié par l'article 4 de la loi du 7 octobre 2016 pour une République numérique.

<sup>9</sup> Wikipédia, eu égard à son mode de fonctionnement collaboratif.



Un opérateur externalise l'instruction de ces signalements à des organismes de *fact-checking*, tandis que les autres les traitent grâce à des équipes internes composées de modérateurs. Cette instruction peut en outre être en partie automatisée. Le CSA souligne toutefois que **les informations transmises quant aux modalités de la procédure d'instruction et aux moyens humains restent insuffisantes et ne lui permettent pas d'apprécier pleinement la portée des moyens mis en œuvre pour répondre aux objectifs de la loi.**

Les opérateurs font état de plusieurs types de mesures susceptibles d'être prises à l'encontre des contenus signalés, **de la réduction de leur visibilité à leur retrait, voire à des actions en direction des comptes.** L'information des utilisateurs et l'existence de voies de recours restent disparates selon les acteurs.

Enfin, **le CSA regrette la rareté des données chiffrées** communiquées par les opérateurs sur les signalements et leur traitement, même s'il prend note des difficultés pratiques mises en avant par les opérateurs pour les réunir. Ces données sont en effet essentielles à l'analyse quantitative de l'efficacité de leurs dispositifs.

#### **Le CSA formule les recommandations suivantes :**

1. 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.
2. 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).
3. 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.
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 des 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.
5. 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.
6. **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 utilisateurs concernés.

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

## 2. Transparence des algorithmes

À deux exceptions près, les opérateurs utilisent des algorithmes afin **d'ordonner, référencer et sélectionner** les contenus, même si cette tripartition n'a pas été reprise expressément dans leurs déclarations. Ils font aussi état de l'utilisation d'algorithmes afin de lutter contre les comptes propageant massivement des fausses informations, détecter et traiter les fausses informations et augmenter ou réduire la visibilité de certains contenus.

Les utilisateurs peuvent, sur certains services, personnaliser les contenus qui leur sont proposés et obtenir des informations sur les algorithmes utilisés, mais cette transparence est limitée.

Les réponses des opérateurs diffèrent peu des informations déjà accessibles sur leurs sites et ne permettent pas au CSA de réaliser une analyse exhaustive. Malgré la nécessaire confidentialité de certaines informations eu égard aux modèles d'affaires des opérateurs qui reposent en grande partie sur la performance de leurs algorithmes, le CSA regrette **le manque de précisions sur l'intelligibilité de ces algorithmes et la non-exhaustivité des informations transmises**. À l'heure où les moyens automatisés occupent une place croissante dans le traitement des fausses informations, la compréhension des moyens mis en œuvre passe par une transparence accrue vis-à-vis du régulateur, sans pour autant porter atteinte au secret des affaires.



### Le CSA formule les recommandations suivantes :

1. Dans le cadre des futures déclarations, il souhaite que les opérateurs lui communiquent davantage d'éléments sur **l'intelligibilité et la redevabilité**<sup>10</sup>, de leurs algorithmes d'organisation, de sélection et d'ordonnancement 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. Le CSA est conscient que le modèle d'affaires de certaines plateformes repose en grande partie sur la performance de leurs algorithmes. Pour autant, il estime qu'il leur est possible de fournir des éléments lui permettant de comprendre les principes de fonctionnement et les objectifs de certains d'entre eux sans dévoiler leur modèle d'affaires.
2. **Le CSA recommande une transparence accrue en matière de fonctionnement des algorithmes vis-à-vis des utilisateurs.** Ces informations sont à ce stade fournies au sein de rubriques denses tels que les centres d'aide ou les politiques de confidentialité, ce qui n'en facilite pas toujours l'accessibilité et la compréhension. D'une manière générale, il considère que **les opérateurs devraient être proactifs pour proposer ces informations** nécessaires à la compréhension et à la bonne utilisation de leurs services.
3. **Le CSA recommande également une transparence accrue en matière de capacité de réglage des algorithmes par les utilisateurs de ces plateformes.** Les réglages contextuels<sup>11</sup>, qu'offrent déjà plusieurs opérateurs, permettent notamment d'améliorer ce niveau de transparence. 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.

Le Conseil envisage de faire **évoluer ces deux points dans le questionnaire** qu'il adressera aux plateformes dans la perspective de l'élaboration du bilan d'application des mesures mises en place en 2020.

4. Enfin, ainsi que l'ont souligné les membres du comité d'experts sur la désinformation en ligne, 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

<sup>10</sup> La redevabilité des algorithmes se définit comme « *l'attribution explicite des responsabilités impliquées par le fonctionnement des algorithmes* » selon la Commission nationale de l'informatique et des libertés (CNIL), définition citée page 51 de son rapport *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, paru en septembre 2017. Il s'agit de connaître, pour des traitements algorithmiques donnés, les personnes responsables de leur définition, de leur mise en œuvre et de la prise de décision qui en découle, ainsi que leurs étapes et degrés d'intervention par rapport à la machine. Ceci est d'autant plus important que les algorithmes développés récemment reposent sur une part d'apprentissage (« *machine learning* ») dont le détail des résultats par exemple en matière de classement des signaux échappent aux concepteurs mêmes de ces modèles.

<sup>11</sup> Réglage ou outil contextuel : fonctionnalité accessible à l'utilisateur directement depuis un contenu qui lui est proposé lors de son utilisation du service cité, par exemple en cliquant dessus. Ce type de fonctions se distingue des réglages ou outils dits généraux, accessibles au sein de pages d'aide, de sections de réglages ou encore de politiques de confidentialité.



sanitaires liées à la COVID-19, qui ont conduit à un usage accru des systèmes automatisés. 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.

### 3. Promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle (EAP-SCA)

La lutte contre les fausses informations passe également par la **valorisation de contenus fiables**. La plupart des opérateurs identifient ces contenus par un signe visuel et certains permettent à l'utilisateur d'obtenir plus d'informations sur leur provenance. Certains opérateurs procèdent, **de façon algorithmique**, à une **remontée des contenus considérés comme issus de sources fiables et/ou à une dégradation de la visibilité de ceux susceptibles de ne pas l'être**.

En revanche, si **l'importance des fact-checkers** dans la lutte contre les fausses informations est globalement reconnue, les opérateurs y font appel et mettent en avant leur travail de façon variable. Certains ont recours aux *fact-checkers* directement ou par l'intermédiaire d'un module téléchargeable, tandis que d'autres exercent une vérification de l'information au sein de leurs équipes. D'autres encore invoquent la nature de leur service ou la mise en place d'une politique éditoriale pour justifier leur choix de ne pas y recourir.

#### Le CSA formule les recommandations suivantes :

1. 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.
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.
3. 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.
4. Le CSA encourage les **démarches de labellisation des EAP-SCA** et estime utile que les utilisateurs des plateformes puissent en avoir connaissance.
5. De manière générale, le CSA encourage fortement **le développement de partenariats entre les plateformes et des fact-checkers (tiers vérificateurs de**



**l'information).**

6. 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).
7. 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.
8. 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.

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

La notion de « compte propageant massivement des fausses informations » recouvre des **réalités variables selon les opérateurs**. Sont considérés comme tels les faux comptes, les comptes dont les comportements contreviennent aux règles internes de la plateforme en matière de propagation massif de fausses informations ou encore ceux pratiquant la multipublication de contenus.

Ces comptes sont détectés grâce des moyens automatisés et/ou humains. Plusieurs types de mesures peuvent être prises à leur encontre, de **la suppression du contenu au blocage temporaire, voire définitif du compte**. Certains opérateurs font état d'une vigilance accrue vis-à-vis des **comptes étatiques**, mesure parfois documentée avec précision dans les déclarations.

Les utilisateurs peuvent accéder aux informations dans les centres d'aide et politiques de confidentialité mais ne sont pas toujours informés des conséquences de la création de tels comptes.



Si les opérateurs soulignent la difficulté de recenser ces comptes, le CSA regrette l'hétérogénéité des données chiffrées déclarées.

**Le CSA formule les recommandations suivantes :**

1. 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.
2. 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.
3. 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.

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

Plusieurs opérateurs n'ont pas abordé la partie du questionnaire relative à cet objectif au motif qu'ils interdisent les contenus se rattachant à un débat d'intérêt général. Néanmoins, les questions portaient sur l'ensemble des contenus sponsorisés<sup>12</sup>.

L'ensemble des acteurs concernés identifie les contenus sponsorisés et certains permettent aux utilisateurs de comprendre pourquoi ils sont exposés à ces contenus. Si certains opérateurs disposent d'un outil de signalement propre aux contenus sponsorisés, la plupart repose sur le dispositif de signalement général.

Les opérateurs ont mis en place **des politiques publicitaires**, qui peuvent consister en la validation des contenus publicitaires ou en l'interdiction de certains contenus. Peu d'informations ont été fournies quant à **la sécurité des marques** et aux mesures permettant aux contenus sponsorisés de gagner en visibilité.

L'approche des contenus d'information se rattachant à un débat d'intérêt général diffère selon les opérateurs. Certains interdisent leur promotion soit de façon permanente, soit en période électorale ; d'autres procèdent à un contrôle particulier des annonceurs concernés, réduisent leurs possibilités de ciblage publicitaire ou encore les identifient spécifiquement.

<sup>12</sup> La notion de contenus sponsorisés est entendue ici comme la publicité au sens large, englobant la publicité « classique » et la promotion des contenus se rattachant à un débat d'intérêt général.



Plusieurs opérateurs ont mis en place des **bibliothèques publicitaires** permettant d'accéder à tout ou partie des contenus sponsorisés.

**Le CSA formule les recommandations suivantes :**

1. Le CSA encourage les plateformes à continuer de mettre en place une **labellisation claire et directement accessible** des contenus promus, aisément visible par l'utilisateur et précisant distinctement l'identité de l'annonceur.
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).
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.



## 6. Education aux médias et à l'information

Une partie des opérateurs mènent des **actions d'EMI**, corollaires des mesures précitées. C'est un enjeu majeur pour ces services, qui exposent leurs publics à des contenus riches et disparates, que de donner les clés de leur compréhension.

Certaines actions sont développées **en partenariat avec des institutions publiques et des associations**. Elles permettent aux publics tant de maîtriser l'utilisation des plateformes que de développer leur esprit critique.

Les opérateurs de plateformes en ligne apportent aussi un soutien financier à plusieurs types d'initiatives et certains font état de liens avec le monde de la recherche, essentiellement par des aides financières.

### Le CSA formule les recommandations suivantes :

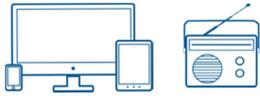
1. Il invite les opérateurs à développer des **approches multidimensionnelles** en matière d'éducation aux médias et à l'information, à destination de tous les publics (8-12 ans, 12-15 ans, 16-18 ans, 19-25 ans, 26-40 ans, 40-65 ans et + de 65 ans, parents, enseignants, etc.)
2. Il les encourage à nouer des **partenariats pluriannuels** en matière d'EMI avec des institutions reconnues et des acteurs majeurs du monde de l'éducation (institutions publiques, associations, etc.).
3. Il les incite à **évaluer l'impact** de leurs actions sur le comportement des utilisateurs.
4. 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.

Enfin, plusieurs opérateurs de plateforme en ligne ont fait état de **mesures complémentaires** à l'instar de mesures de lutte contre les *deepfakes*, de protection des processus électoraux, de fonctions de recherche avancée et de restrictions de la fonctionnalité « *live* ».



## SOMMAIRE

<b>Introduction – cadre et méthode</b> .....	2
<b>Synthèse</b> .....	4
<b>1. Les plateformes ayant fait l'objet d'une déclaration</b> .....	16
<b>1.1. Présentation des plateformes au vu des déclarations reçues</b> .....	16
<b>1.2. Périmètre des services concernés</b> .....	17
<b>2. Le devoir de coopération</b> .....	19
<b>2.1. Les opérateurs ont adopté une démarche coopérative vis-à-vis du CSA mais certains sujets ont été très peu traités dans leurs déclarations</b> .....	19
<b>2.2. La désignation d'un interlocuteur référent</b> .....	20
<b>3. Moyens mis en œuvre par les plateformes</b> .....	21
<b>3.1. Le dispositif de signalement de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un scrutin</b> .....	21
3.1.1. Typologie des mesures prises par les opérateurs.....	22
3.1.2. Analyse des mesures prises par les opérateurs.....	27
3.1.3. Recommandations du CSA.....	32
<b>3.2. Transparence des algorithmes</b> .....	33
3.2.1. Typologie des mesures prises par les opérateurs.....	34
3.2.2. Analyse des mesures prises par les opérateurs.....	37
3.2.3. Recommandations du CSA.....	39
<b>3.3. Promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle</b> .....	40
3.3.1. Typologie des mesures prises par les opérateurs.....	40
3.3.2. Analyse des mesures prises par les opérateurs.....	45
3.3.3. Recommandations du CSA.....	48
<b>3.4. Lutte contre les comptes propageant massivement de fausses informations</b> .....	49
3.4.1. Typologie des mesures prises par les opérateurs.....	49
3.4.2. Analyse des mesures prises par les opérateurs.....	55
3.4.3. Recommandations du CSA.....	57
<b>3.5. Information des utilisateurs sur la nature, l'origine, les modalités de diffusion des contenus et l'identité des personnes versant des rémunérations en contrepartie de la promotion des contenus d'information</b> .....	58



3.5.1. Typologie des mesures prises par les opérateurs.....	58
3.5.2. Analyse des mesures prises par les opérateurs.....	62
3.5.3. Recommandations du CSA.....	65
<b>3.6. Education aux médias et à l'information .....</b>	<b>66</b>
3.6.1. Typologie des mesures prises par les opérateurs.....	66
3.6.2. Analyse des mesures prises par les opérateurs.....	67
3.6.3. Recommandations du CSA.....	69
<b>3.7. Moyens complémentaires mis en œuvre .....</b>	<b>70</b>
<b>Conclusion .....</b>	<b>72</b>
<b>ANNEXES .....</b>	<b>73</b>
<b>Annexe 1 : Dispositions légales et réglementaires .....</b>	<b>73</b>
<b>Annexe 2 : Questionnaire du CSA aux opérateurs .....</b>	<b>75</b>
<b>Annexe 3 : Composition du comité d'experts sur la désinformation en ligne.....</b>	<b>87</b>



## 1. Les plateformes ayant fait l'objet d'une déclaration

### 1.1. Présentation des plateformes au vu des déclarations reçues

Les services ayant fait l'objet d'une déclaration au CSA répondent aux deux types d'activités visées par la loi<sup>13</sup> :

- classement ou référencement de contenus, biens ou services mis en ligne par des tiers : *Bing, Google, Yahoo Portal, Yahoo Search* ;
- mise en relation de plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service : *Aufeminin, Dailymotion, Doctissimo, Facebook, Google, Instagram, Jeuxvideo.com, LinkedIn, Microsoft Ads, Snapchat, Twitter, Wikipédia, Yahoo Portal, YouTube*.

**Ces plateformes ont un spectre d'activité large.** Plusieurs jouent en priorité un rôle de réseau social, généraliste (*Snapchat, Instagram, Facebook*) ou spécialisé (*LinkedIn*). D'autres exercent des fonctions de moteur de recherche (*Bing, Google, Yahoo Search*). Certaines permettent le partage de contenus vidéo et audio à destination de tous les utilisateurs (*YouTube, Dailymotion*) ou offrent des fonctionnalités de forum en ligne (*Jeuxvideo.com, Doctissimo, Aufeminin*). Enfin, certaines référencent des contenus d'information mis en ligne par des tiers (*Google, Yahoo Portal*), ou ont une vocation encyclopédique (*Wikipédia*).

**Les services de plateformes sont tous disponibles en France**<sup>14</sup>. Certains opérateurs proposent des versions spécifiques pour le territoire français (*Facebook, Yahoo, Snapchat, Google*). En termes d'établissement, *Dailymotion, Webedia (Jeuxvideo.com)* et *Unify (Doctissimo et Aufeminin)* sont des entreprises françaises. Les autres plateformes sont opérées par des sociétés ou association établies en Irlande (*Facebook Ireland Limited, LinkedIn Ireland Unlimited Company, Google Ireland Limited, Twitter International Company, Verizon Media EMEA Limited (Yahoo)*) ou aux Etats-Unis (*Microsoft Corporation, Wikimedia Foundation Inc., Snap Inc.*), dont certaines ont des bureaux en France (notamment *Facebook, Microsoft, Twitter, Snapchat, Google*).

**Ce sont des acteurs de tailles très différentes.** Leur nombre moyen d'utilisateurs mensuels en France serait, selon les chiffres déclarés, le suivant : 36 millions pour *Facebook*, 33 millions pour *Dailymotion*, 20 millions pour *Snapchat*, 8,5 millions pour *Doctissimo*, 6,5 millions pour *Jeuxvideo.com*, 5,5 millions pour *Aufeminin*. Un opérateur n'a pas fourni de chiffre en la matière, invoquant leur caractère confidentiel et deux autres les ont communiqués au CSA en lui demandant qu'il ne rende pas publiques ces informations, invoquant le secret des affaires (*LinkedIn*) ou la confidentialité de ces informations (*Yahoo*).

**Leur modèle économique** est basé pour tout ou partie sur les revenus publicitaires, à l'exception de *Wikipédia*. Certaines plateformes ont communiqué leur chiffre d'affaires (*Snapchat, Twitter, Unify*). D'autres n'avaient pas encore arrêté leurs comptes au moment de

<sup>13</sup> Voir note 1.

<sup>14</sup> Précision de lecture : ci-après, sauf mention contraire, Facebook désigne les réponses pour les services Facebook et Instagram ; Google, les réponses pour les services Google et YouTube ; Yahoo, les réponses pour les services Yahoo Search et Yahoo Actualités ; Microsoft, les réponses pour les services Microsoft Ads et Bing ; Unify les réponses pour Doctissimo et Aufeminin.



la déclaration (*Facebook, Jeuxvideo.com*). Enfin, d'autres n'ont pas souhaité communiquer ces informations (*Microsoft, Google, Dailymotion*) ou les ont communiquées de façon confidentielle (*LinkedIn, Yahoo*).

Les données communiquées relatives au nombre de visiteurs, au modèle économique des plateformes et à leur chiffre d'affaires permettent au CSA de mieux tenir compte de leurs spécificités dans son appréciation des mesures mises en œuvre pour lutter contre la diffusion de fausses informations. Le CSA souhaite que les opérateurs les lui fournissent de façon complète à l'avenir.

## 1.2. Périmètre des services concernés

La première année d'application de la loi du 22 décembre 2018 a été l'occasion pour le CSA de s'intéresser à la question du périmètre des services tenus de lui adresser une déclaration au titre du devoir de coopération prévu par le titre III de cette loi.

Le législateur a fait le choix de rendre applicable le devoir de coopération à ceux des opérateurs de plateformes en ligne définis par l'article L. 111-7 du code de la consommation qui dépassent un seuil de connexions fixé par décret.

Le code de la consommation définit l'opérateur de plateforme en ligne comme : « *toute personne physique ou morale proposant, à titre professionnel, de manière rémunérée ou non, un service de communication au public en ligne reposant sur : 1° Le classement ou le référencement, au moyen d'algorithmes informatiques, de contenus, de biens ou de services proposés ou mis en ligne par des tiers ; 2° Ou la mise en relation de plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service* ».

Le décret n°2019-297 du 10 avril 2019 a quant à lui fixé à cinq millions de visiteurs uniques par mois le nombre de connexions sur le territoire français qui déclenche les obligations de coopération d'un opérateur de plateforme en ligne.

**Onze opérateurs ont fait parvenir au CSA une déclaration pour l'exercice 2019**, présentant les actions menées en termes de lutte contre la manipulation de l'information sur un ou plusieurs de leurs services. Au total, **seize services** ont ainsi été déclarés : *Dailymotion, Doctissimo, AuFeminin, Instagram, Facebook, YouTube, Google Search, Jeuxvideo.com, LinkedIn, Bing, Microsoft Advertising, Snapchat, Twitter, Wikipédia, Yahoo Portal et Yahoo Search*.

Le CSA a été amené à s'interroger sur le périmètre des services qui ont fait l'objet d'une déclaration à ce titre.

La question s'est par exemple posée de l'inclusion dans le champ de la régulation des services d'achat de publicités sur des moteurs de recherches (sites de *Search Engine Marketing*) et de celles des magasins d'applications.



En outre, certains opérateurs n'ont pas présenté de mesures au titre de l'article 11 de la loi du 22 décembre 2018 en ce qui concerne certaines parties de leur service. Il en va ainsi d'espaces de partage de vidéos ou de « places de marchés » intégrés à certains réseaux sociaux<sup>15</sup> pour lesquels il n'a pas été indiqué les modalités de mises en œuvre de dispositifs de signalement de fausses informations ni de mesures complémentaires pouvant porter notamment sur la transparence des algorithmes ou la promotion de contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle.

D'autres opérateurs ont estimé qu'ils n'avaient pas à présenter certaines mesures mises en œuvre sur leur service dès lors que celui-ci fait appel aux technologies d'entreprises tierces sur lesquelles ils n'ont pas de contrôle. Le cas rencontré ici est celui de certains moteurs de recherches qui utilisent des outils d'indexation tiers.

En se déclarant et détaillant les mesures prises sur les services qu'ils offrent, certains opérateurs ont enfin remis en question l'application de la loi du 22 décembre 2018 à leurs services de portail d'informations ou de « curation d'actualités », estimant que ces derniers fonctionnent à partir d'une sélection de sources journalistiques impliquant au moins en partie une intervention humaine et ne représentant que peu de risques en termes de manipulation de l'information.

Tout en relevant que les onze opérateurs lui ayant adressé une déclaration figurent parmi les opérateurs de plateforme en ligne les plus concernés par le phénomène des fausses informations, **le CSA souligne que le législateur n'a pas restreint la catégorie des acteurs concernés à ceux dont l'activité serait considérée comme propice à la diffusion de fausses informations** et appelle tous les opérateurs de plateformes en ligne visées par l'article L. 111-7 du code de la consommation qui dépassent le seuil de connexion fixé par décret à respecter les obligations prévues au titre III de la loi.

La généralité des termes utilisés par le législateur pour définir les opérateurs concernés et l'absence de critères d'exclusion invitent par ailleurs à considérer que, sous réserve des seuils applicables et des spécificités propres à certains services qui ne correspondraient pas à la définition du code de la consommation, un service d'achat de publicités sur des moteurs de recherches, un magasin d'applications, un espace de partage de vidéos ou une place de marché mettent dans la plupart des cas « *en relation (...) plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service* ». Il en va de même des portails d'information lorsque ces derniers reposent sur la mise à disposition de contenus de personnes physiques ou morales tierces. Certains de ces services, notamment les magasins d'applications, peuvent au reste répondre à l'activité de « *classement ou [...] référencement, au moyen d'algorithmes informatiques, de contenus, de biens ou de services proposés ou mis en ligne par des tiers* ».

Il revient enfin à chacun des opérateurs concernés, dès qu'ils dépassent le seuil du nombre de connexions sur le territoire français prévu par le décret du 10 avril 2019, et ce indifféremment du fait qu'ils fassent appel à des technologies de sociétés tierces, de s'assurer que le service mis à disposition propose, d'une part, un dispositif facilement accessible et visible permettant aux utilisateurs de signaler des fausses informations et,

<sup>15</sup> Notamment l'espace Marketplace de [Facebook](#).



d'autre part, des mesures complémentaires dont l'article 11 de la loi du 22 décembre 2018 fournit une liste illustrative.

## 2. Le devoir de coopération

### 2.1. Les opérateurs ont adopté une démarche coopérative vis-à-vis du CSA mais certains sujets ont été très peu traités dans leurs déclarations

**Le CSA souhaite souligner que le degré de coopération des opérateurs s'est révélé, de manière générale, élevé.**

A l'exception d'*Unify*<sup>16</sup>, les opérateurs ont renvoyé des déclarations complètes dans les temps impartis. Ils ont tous répondu (certes, de façon non exhaustive) aux questions complémentaires du CSA, et plusieurs d'entre eux (*Facebook, Twitter, Google, Snapchat, Dailymotion, Jeuxvideo.com*) ont sollicité des réunions avec les services du CSA à cette fin.

Les déclarations sont, de manière générale, riches en informations. Elles s'attachent, pour la majeure partie d'entre elles, à fournir des éléments sur les différents thèmes visés par le questionnaire et découlant de la loi et de la recommandation du CSA. La plupart des opérateurs ont répondu sur l'ensemble des thèmes les concernant et ont fait œuvre de pédagogie. En revanche, **le niveau de précision des réponses est hétérogène**, certains opérateurs ayant fait l'effort de fournir des informations très détaillées et illustrées en réponse à certaines questions posées. D'autres, comme *Verizon Media*, ont signalé qu'ils n'entendaient pas répondre à plusieurs parties du questionnaire.

On note en outre que *Verizon Media* a adressé ses réponses en anglais, contrairement aux autres opérateurs qui ont tous fait l'effort de répondre en français, même lorsqu'ils n'ont pas d'établissement ou de bureau en France ou que leurs équipes techniques et juridiques en charge de la déclaration ne sont pas francophones.

Certains opérateurs ont suivi la trame du questionnaire, ce qui facilite l'instruction des déclarations. En revanche, pour fournir certaines informations, la plupart des opérateurs ont effectué des renvois vers des sites internet (site institutionnel, centre d'aide, politique de confidentialité ou conditions générales d'utilisation (CGU)), ce qui permet certes d'enrichir les déclarations sans les surcharger mais peut amener à rendre l'information moins accessible.

En règle générale, **les informations relatives aux moyens mis en œuvre autour du dispositif de signalement de fausses informations ont été particulièrement développées** par les opérateurs.

Plusieurs opérateurs ont estimé que le modèle de leur plateforme les prémunissait, de fait, du risque de diffusion de fausses informations et à ce titre, n'ont pas fourni toutes les

---

<sup>16</sup> L'opérateur a souligné un contexte difficile et une restriction de ses moyens humains du fait des conditions de travail liées au COVID-19.



réponses au questionnaire attendues. Eu égard aux enjeux de la lutte contre la diffusion de fausses informations en ligne et dans la mesure où ces acteurs sont soumis aux obligations de la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information, le CSA s'interroge vivement sur la pertinence de cet argument.

En règle générale, **peu d'informations ont été données s'agissant des moyens humains et financiers consacrés à la lutte contre les fausses informations**, malgré les demandes du CSA dans son questionnaire. En particulier, la question de la formation et de la langue de travail des modérateurs (notamment les équipes traitant les signalements) n'a été abordée que par un nombre restreint d'acteurs (*Jeuxvideo.com, Snapchat, LinkedIn, Facebook*). **Les questions relatives à la transparence, à l'utilisation et au fonctionnement des algorithmes ont également été peu traitées, à quelques exceptions près** (notamment *LinkedIn*). Les raisons invoquées sont de trois ordres : le secret des affaires, l'absence de disponibilité des informations et le fait que les moyens dédiés à la lutte contre les fausses informations sont globaux et non répartis spécifiquement sur telle ou telle mesure.

**Le CSA déplore ne pas disposer d'informations plus précises et, de ce fait, de ne pas pouvoir pleinement évaluer l'application et l'effectivité des mesures mises en œuvre en matière de traitement des signalements et de transparence des algorithmes.** À cet égard, il rappelle qu'il a laissé aux plateformes la possibilité de lui communiquer des éléments qui, s'ils relevaient du secret d'affaires, ne seraient pas rendus publics, possibilité dont l'ensemble des opérateurs (à l'exception de Wikipédia) se sont d'ailleurs saisis pour d'autres informations. En outre, si le secret des affaires trouve à s'appliquer aux données qui revêtent une valeur commerciale, effective ou potentielle, on peut s'interroger si relèvent réellement de cette catégorie des éléments tels que des données suffisamment agrégées ou statistiques portant sur la configuration des équipes de modération d'une plateforme ou tels que des informations permettant d'appréhender le fonctionnement général des algorithmes.

**Le CSA invite les opérateurs à préciser leur réponse sur ces points à l'avenir.**

## 2.2. La désignation d'un interlocuteur référent

Aux termes de l'article 13 de loi du 22 décembre 2018, les opérateurs de plateformes en ligne sont tenus de désigner « *un représentant légal exerçant les fonctions d'interlocuteur référent sur le territoire français* » pour l'application notamment du devoir de coopération prévu au titre III de cette loi.

Dans sa recommandation du 15 mai 2019, le CSA a invité les opérateurs concernés à lui faire connaître ce représentant. Pour faciliter ce processus, un formulaire de désignation à l'attention des opérateurs de plateforme a été mis en ligne sur le site [www.csa.fr](http://www.csa.fr). Des procédures internes ont également été mises en œuvres par le CSA pour garantir un traitement des données à caractère personnel ainsi collectées conforme avec le Règlement général sur la protection des données (RGPD).



**À une exception près (*Wikipédia*<sup>17</sup>), l'ensemble des opérateurs ayant fait parvenir leur déclaration annuelle au CSA lui ont fait connaître leur référent, soit au préalable par le biais du formulaire (7 opérateurs), soit dans leur déclaration (3 opérateurs).**

Le CSA rappelle que **la désignation de ce référent est une obligation légale qui participe pleinement du devoir de coopération** prévu par le dispositif de lutte contre la manipulation de l'information.

La première année d'application de ce texte a été l'occasion pour le CSA de relever différentes situations :

- soit l'interlocuteur référent désigné ne figure pas parmi les effectifs de l'organisme opérant la plateforme mais parmi les effectifs d'une filiale française du même groupe ;
- soit l'interlocuteur référent désigné ne figure pas parmi les effectifs de l'organisme opérant la plateforme mais parmi les effectifs d'un organisme tiers (cabinet de conseil ou d'avocats par exemple) ;
- soit l'interlocuteur référent est établi hors du territoire français (cas de *Verizon Media* notamment).

A ce sujet, le CSA estime que, tant la lettre du texte que les travaux parlementaires ayant conduit à son adoption conduisent à ce que ce représentant légal soit une personne physique désignée par l'opérateur de la plateforme, titulaire du pouvoir de la représenter et exerçant son activité sur le territoire français. La désignation d'un interlocuteur référent extérieur à l'opérateur, par exemple employé par un cabinet d'avocat ou un cabinet de relations publiques, est par conséquent exclue. Il en va de même d'une personne qui exercerait cette fonction depuis l'étranger, y compris depuis un autre Etat membre de l'Union européenne. A l'inverse, le CSA considère que les dispositions de l'article 13 de la loi du 22 décembre 2018 sont respectées lorsque le référent désigné figure non pas parmi les effectifs de l'organisme opérant la plateforme mais parmi ceux d'une des filiales du groupe auquel il appartient.

### 3. Moyens mis en œuvre par les plateformes

#### 3.1. Le dispositif de signalement de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un scrutin

Les éléments d'information transmis par les opérateurs de plateforme concernant l'outil de signalement de fausses informations sont particulièrement fournis, par rapport aux autres types de mesures mises en œuvre. Plusieurs raisons peuvent expliquer cette différence : premièrement, l'article 11 oblige les opérateurs concernés à mettre en place un dispositif de signalement, là, où, pour le reste, le législateur impose aux opérateurs de plateformes de

---

<sup>17</sup> L'opérateur explique dans sa déclaration : « *la Wikimedia Foundation Inc. (ci-après la "Fondation Wikimedia") n'emploie aucun salarié en France. Dans ces conditions, la Fondation Wikimedia n'est pas en mesure de désigner un référent légal en France. Cela étant, la Fondation Wikimedia a communiqué dans sa déclaration annuelle les coordonnées d'interlocuteurs au sein de la Fondation qui seront à même de répondre aux questions et coopérer avec le CSA.* ».



mettre en œuvre des « mesures complémentaires », au sujet desquelles il donne une liste illustrative. Ensuite, ces informations peuvent être moins sensibles que celles relatives à d'autres mesures, de sorte que les opérateurs s'estiment plus libre de les partager. Enfin, les mesures prises en faveur d'un dispositif de signalement aisément accessible et visible, telles que recommandées par le CSA, ont une portée large, pouvant recouvrir d'autres moyens complémentaires visés par les textes, qu'il s'agisse de l'identification des faux comptes, des outils de signalement de la publicité politique, du déploiement de moyens automatisés dans la lutte contre la désinformation, de la collaboration avec les entreprises de presse « fiables » et la valorisation de leur production, ou bien encore de la mise à disposition des chercheurs de données recueillies par les opérateurs (ex. : signalements, contenus qualifiés de fausses informations)..

### 3.1.1. Typologie des mesures prises par les opérateurs

DISPOSITIF DE SIGNALEMENT	
Forme, accessibilité, visibilité	<p><b>Outil à proximité immédiate du contenu</b> (<i>YouTube, Facebook, Twitter, LinkedIn</i><sup>18</sup>, <i>Jeuxvideo.com, Dailymotion, Yahoo, Unify, Snapchat</i>) ;</p> <ul style="list-style-type: none"> <li>-&gt; <b>outil accessible à tous</b> (<i>Dailymotion</i><sup>19</sup>, <i>Bing, Google search, Unify, Jeuxvideo.com</i>) <b>ou uniquement aux utilisateurs connectés</b> (<i>Snapchat ; Jeuxvideo.com</i><sup>20</sup>) ;</li> <li>-&gt; <b>signalé par un symbole</b> (<i>Facebook, Instagram, Twitter, YouTube, Snapchat, LinkedIn, Dailymotion, Jeuxvideo.com, Doctissimo</i>) <b>ou un mot</b> comportant un hyperlien (« Signaler » ou « Alerter ») (<i>Dailymotion, Aufeminin</i>).</li> </ul> <p><b>Existence d'un formulaire de signalement</b> (rappelant éventuellement les mentions légales) depuis le bas de la page ou depuis le site institutionnel (<i>YouTube, Facebook, Twitter, LinkedIn, Jeuxvideo.com, Dailymotion, Yahoo, Bing, Google, Wikipédia</i>) ;</p> <ul style="list-style-type: none"> <li>- <b>formulaire accessible à tous</b> (<i>Google, Yahoo, Bing, Jeuxvideo.com, Wikipédia</i>) <b>ou uniquement aux utilisateurs connectés</b> (<i>Facebook, Twitter, LinkedIn, YouTube</i>).</li> </ul> <p><b>Outil de signalement proposé :</b></p> <ul style="list-style-type: none"> <li>- <b>pour tous les contenus</b> (publications, publicités, chaînes, comptes, commentaires) (<i>Microsoft, Dailymotion, Snapchat, Twitter</i>) ;</li> <li>- ou <b>pour certains contenus uniquement</b> (<i>Yahoo Actualités, Facebook, YouTube, Jeuxvideo.com, Doctissimo, Aufeminin</i>) ;</li> </ul> <p><b>Outil accessible en :</b></p> <ul style="list-style-type: none"> <li>- <b>un clic</b> (<i>Snapchat, Twitter, YouTube, Microsoft Advertising, Dailymotion, LinkedIn, Facebook, Jeuxvideo.com</i> pour les utilisateurs connectés,</li> </ul>

<sup>18</sup> *LinkedIn* a entrepris en 2020 une refonte de l'ergonomie de son processus de signalement, prévoyant notamment la réduction du nombre de clics.

<sup>19</sup> *Dailymotion* signale dans sa déclaration qu'au premier semestre 2020, une réforme de l'ergonomie du dispositif de signalement sera mise en place.

<sup>20</sup> Le dispositif de signalement sur *Jeuxvideo.com* diffère selon que l'utilisateur est connecté ou non.



	<p><i>Doctissimo, Aufeminin</i>);</p> <ul style="list-style-type: none"> <li>- <b>en deux clics</b> (<i>Yahoo, Jeuxvideo.com</i> pour les utilisateurs non connectés);</li> <li>- ou dispositif de signalement via le centre d'aide et le formulaire de contact (<i>Google, Bing</i>).</li> </ul> <p>Possibilité de <b>signaler plusieurs contenus</b> en même temps (<i>YouTube</i>).</p>
Motifs proposés	<p><b>Motif :</b></p> <ul style="list-style-type: none"> <li>- « <b>Fausse information</b> » ou formulation équivalente (<i>Facebook, Dailymotion, Unify</i><sup>21</sup>); uniquement pour les contenus publicitaires (<i>Snapchat</i>);</li> <li>- « <b>Fausse information au sujet d'élections</b> » ou formulation équivalente (<i>LinkedIn, Twitter, Doctissimo, Aufeminin</i>);</li> <li>- « <b>Spam ou contenu trompeur</b> » (<i>YouTube</i>);</li> </ul> <p><b>Absence de motif « fausse information » ou s'y rattachant</b> (<i>Snapchat</i> (hors publicité), <i>Microsoft, Jeuxvideo.com, Google</i>).</p>
Formes de l'arborescence	<p><b>Arborescence :</b></p> <ul style="list-style-type: none"> <li>- « à plat » : tous les motifs sont visibles en même temps (<i>signalement d'une publicité sur Microsoft Advertising, Jeuxvideo.com, Doctissimo, Aufeminin</i>);</li> <li>- « par strates » (<i>Facebook, Twitter, YouTube, LinkedIn</i>).</li> </ul> <p><b>Apparition du motif « fausse information » (ou assimilé) :</b></p> <ul style="list-style-type: none"> <li>- proposé <b>dès le niveau 1 de l'arborescence</b> (<i>Facebook, YouTube, Dailymotion, Twitter</i>);</li> <li>- proposé <b>au niveau 2 de l'arborescence</b> (<i>LinkedIn</i>);</li> <li>- <b>absent mais possibilité de choisir « autre »</b> (<i>Microsoft, Snapchat, Yahoo</i>);</li> <li>- <b>absent sans champ « autre »</b> (<i>Jeuxvideo.com</i>).</li> </ul>
Fonctionnalités de l'outil	<p><b>Choix parmi :</b></p> <ul style="list-style-type: none"> <li>- des <b>items</b> proposés uniquement (<i>Facebook, LinkedIn, Microsoft, Aufeminin</i>);</li> <li>- des items proposés + <b>zone de texte</b> libre (<i>Snapchat, Dailymotion, Google, Jeuxvideo.com, Yahoo, Doctissimo</i>);</li> <li>- des items proposés + zone de texte libre + <b>ajout d'informations complémentaires</b> (captures d'écran, contenus supplémentaires) (<i>Twitter</i>).</li> </ul>
Nombre de signalements pour fausses informations déclarés par les	<p><b>Chiffres communiqués :</b></p> <ul style="list-style-type: none"> <li>- 0 (<i>Google, Snapchat</i>);</li> <li>- 40 (<i>Doctissimo</i>);</li> <li>- Au plus 40 000 (<i>Jeuxvideo.com</i><sup>22</sup>);</li> </ul>

<sup>21</sup> Catégorie effective depuis le deuxième trimestre 2018.



opérateurs	<p>- 2512 (<i>LinkedIn</i>) ;</p> <p>- 49 945 signalements via l'outil de signalement concernant l'élection des représentants au Parlement européen et 14 529 signalements via l'outil de signalement de fausses informations spécifique à la France (<i>Twitter</i>).</p> <p><b>Absence de données</b> à communiquer (<i>Facebook, Wikipédia, Auféminin</i>) ou non précisé (<i>Microsoft, Yahoo</i>).</p>
------------	---

SANCTION DES FAUSSES INFORMATIONS	
Politique de traitement des fausses informations	<p><b>Externalisation</b> du traitement des fausses informations par des <b>tiers vérificateurs de confiance</b> (<i>fact-checkers</i>) (<i>Facebook</i>)<sup>23</sup>.</p> <p>Traitement par une <b>équipe de modérateurs pas exclusivement en charge du traitement des fausses informations</b> (<i>Snapchat, Google, Microsoft, Yahoo, Jeuxvideo.com, LinkedIn, Dailymotion, Unify</i>).</p> <p>Approche spécifique :</p> <ul style="list-style-type: none"> <li>- <i>Wikipédia</i> : <b>modération par les utilisateurs</b> ;</li> <li>- <i>Twitter</i> : modération <b>uniquement des fausses informations en matière électorale</b>.</li> </ul>
Doctrine suivie par les plateformes	<p><b>Absence de retrait</b> des fausses informations, sauf en cas d'entrave au processus électoral (<i>Twitter, Facebook</i>).</p> <p><b>Retrait</b> des fausses informations :</p> <ul style="list-style-type: none"> <li>- lorsqu'il y a une <b>violation de la loi</b> (<i>Google, Bing</i>) ;</li> <li>- <b>après déclaration comme telles par l'équipe de modération</b> (<i>LinkedIn, Dailymotion, Snapchat, Jeuxvideo.com</i>).</li> </ul>
Normes de référence du contrôle des plateformes	<p><b>Loi locale</b><sup>24</sup> (<i>Microsoft, Jeuxvideo.com</i>).</p> <p><b>Règles et standards de la communauté</b>, adaptés pour tenir compte de la loi le cas échéant (<i>Snapchat, Facebook, Jeuxvideo.com, Snapchat, Twitter</i>,</p>

<sup>22</sup> Ce chiffre a été obtenu selon une estimation empirique ; il s'agit du nombre maximum de contenus susceptibles d'avoir été signalés en raison de la présence d'une fausse information.

<sup>23</sup> Sur *Facebook* et *Instagram*, le signalement d'un contenu au motif qu'il contient une fausse information n'entraîne pas une revue par l'opérateur lui-même (à moins que le contenu ne contrevienne par ailleurs aux standards de la communauté, qui n'interdisent pas la diffusion de fausses informations pour les contenus non publicitaires). Ce signalement entraîne le « marquage » automatique du contenu. Cela a pour conséquence de le faire figurer et remonter, en fonction notamment du nombre de signalements (ainsi que l'activité antérieure du compte et le niveau de défiance des réactions à la publication), dans la liste des contenus signalés aux partenaires *fact-checkers*.

Cela ne vaut pas pour les publicités, qui lorsqu'elles sont « trompeuses », ainsi que pour 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 : le signalement de ces pratiques, interdites par les Standards de la Communauté, est traité par les modérateurs.

<sup>24</sup> Loi du 22 décembre 2018, en l'espèce.



	<i>LinkedIn</i> ).
Procédure d'instruction des signalements	<p>Concernant l'ouverture et le traitement d'un signalement :</p> <ul style="list-style-type: none"> <li>- <b>traitement de tous les signalements</b> (<i>Snapchat, Yahoo Portal, Jeuxvideo.com, Dailymotion, Unify</i><sup>25</sup>) ;</li> <li>- traitement des signalements <b>en fonction de critères</b> précisés (<i>LinkedIn</i><sup>26</sup>) ou non (<i>Facebook</i>) dans la déclaration ;</li> <li>- traitement <b>uniquement des signalements de fausses informations au sujet d'élections</b> (<i>Twitter</i>) ;</li> <li>- <b>pas d'indication</b> (<i>Google, Microsoft, Yahoo Search</i>).</li> </ul> <p><b>Procédures de priorisation</b> (<i>Dailymotion, Jeuxvideo.com</i>) ;</p> <ul style="list-style-type: none"> <li>- <i>YouTube</i> : programme <i>Trusted Flaggers</i> qui permet un <b>traitement prioritaires des signalements d'acteurs identifiés</b> au regard de leurs compétences.</li> </ul> <p><i>Jeuxvideo.com</i> dispose d'une <b>interface technique propriétaire</b> de traitement des signalements (Gestion et Traitement des Alertes).</p>
Types de mesures adoptées	<p>Mesures sur le contenu :</p> <ul style="list-style-type: none"> <li>- <b>retrait</b> (<i>Google, Microsoft Advertising, Snapchat, Yahoo, Jeuxvideo.com</i>) ;</li> <li>- <b>désindexation</b> de l'URL lorsque la loi l'exige (<i>Bing</i>) ;</li> <li>- <b>réduction de la visibilité</b> (<i>Google, Facebook</i>) ;</li> <li>- <b>restriction aux mineurs</b> : apposition d'un filtre parental (<i>Bing</i>) ou d'une limite d'âge (<i>YouTube</i>).</li> <li>- <b>information de l'utilisateur</b> sur le fait qu'il est en train d'interagir avec un contenu diffusant de fausses informations (<i>YouTube, Facebook</i>).</li> </ul> <p>Mesures sur le compte :</p> <ul style="list-style-type: none"> <li>- <b>suppression du compte</b> (<i>Snapchat, Jeuxvideo.com</i>)</li> <li>- ou mesures disciplinaires à son encontre (<i>LinkedIn, Jeuxvideo.com</i>).</li> </ul>
Nombre de contenus retirés	<b>Peu d'informations sur le nombre de contenus retirés</b> après avoir été signalés comme propageant une fausse information.
Délai d'instruction des signalements	<p>Les opérateurs ayant répondu à ce point déclarent un délai de traitement de <b>24 heures maximum</b> (<i>Snapchat, Dailymotion, Jeuxvidéo.com, Unify</i>).</p> <p>Autres opérateurs : pas d'information sur le délai d'instruction.</p>

#### INFORMATION DES UTILISATEURS

Information de l'ensemble des	Informations générales délivrées sur : - la <b>procédure</b> pour effectuer un signalement (ex : <i>Facebook</i> ) ;
-------------------------------	---

<sup>25</sup> *Unify* déclare formaliser sa procédure d'instruction en 2020.

<sup>26</sup> *LinkedIn* prévoit d'abaisser en 2020 le nombre de signalements conditionnant l'ouverture d'une procédure d'instruction.



utilisateurs	<ul style="list-style-type: none"><li>- les <b>règles de fonctionnement</b> (autorisations/interdictions) ;</li><li>- les <b>sanctions</b> encourues ;</li><li>- les <b>contenus visés par une décision</b> de l'opérateur (contenu masqué, contenu désindexé, contenu retiré ; ex. : <i>YouTube</i>).</li></ul>
Information après un signalement	Informations après un signalement délivrées à : <ul style="list-style-type: none"><li>- <b>l'auteur du signalement</b> : suivi de sa demande, instruction, décision, recours (<i>Microsoft, Jeuxvideo.com, Facebook</i><sup>27</sup>) ;</li><li>- <b>l'auteur du contenu signalé</b> : notification du signalement, décision, recours (<i>Microsoft</i>) ;</li><li>- <b>tous</b> les utilisateurs (<i>Dailymotion</i>).</li></ul> <p><b>Pas de dispositif</b> d'information (<i>Snapchat, LinkedIn</i><sup>28</sup>, <i>Jeuxvideo.com</i><sup>29</sup>, <i>Unify</i><sup>30</sup>).</p>
Existence de voies de recours	Pour : <ul style="list-style-type: none"><li>- <b>l'auteur du contenu</b> (<i>Snapchat, Facebook, Dailymotion</i><sup>31</sup>, <i>Jeuxvideo.com</i>) ;</li><li>- <b>l'auteur du signalement</b> (<i>Facebook</i><sup>32</sup>, <i>Jeuxvideo.com</i>).</li></ul>
<p><b>L'ouverture de l'instruction d'un signalement ne semble pas faire l'objet d'une information particulière auprès des utilisateurs.</b></p>	

### **Mesures prises dans le cadre de la crise de la COVID-19**

- **Retraits de contenus** violant les règles du service (*YouTube* ; plusieurs centaines de milliers en mars 2020 sur *Facebook*) ;
- **blocage ou limitation de hashtags** utilisés pour diffuser de fausses informations liées à la COVID-19 (*Instagram*).

<sup>27</sup> *Facebook* permet aux utilisateurs de suivre le statut des éléments signalés.

<sup>28</sup> Dans sa déclaration, *LinkedIn* indique mettre en place une boucle de rétroaction automatique qui sera prête dans le courant de l'année 2020.

<sup>29</sup> Il est prévu soient développés, courant 2020, un tel dispositif pour l'auteur du signalement ainsi qu'une information de l'ensemble des utilisateurs sur la suppression d'un contenu signalé.

<sup>30</sup> Sauf si l'utilisateur a engagé une démarche proactive de prise de contact avec un modérateur.

<sup>31</sup> Pour les utilisateurs professionnels uniquement.

<sup>32</sup> *Facebook* s'est doté d'un conseil de surveillance indépendant en charge du traitement de ces recours.



### 3.1.2. Analyse des mesures prises par les opérateurs

- **Les opérateurs se sont en grande partie conformés à la recommandation du CSA sur l'existence d'un outil de signalement accessible, mais l'efficacité de ces outils dépend de leur ergonomie.**

#### **Existence d'un outil de signalement**

Le CSA note avec satisfaction que la totalité des opérateurs déclarants, à une exception, **ont développé un outil de signalement**. À cet égard, l'effort des opérateurs pour se conformer à la loi du 22 décembre 2018 ainsi qu'à la recommandation du CSA est notable. Seul *Wikipédia* ne dispose pas d'un tel outil, ce qui s'explique par le modèle coopératif choisi par l'encyclopédie en ligne : chaque personne étant libre de devenir contributeur, seules les règles régissant la publication des articles viennent sanctionner la publication de contenus problématiques, sans que leur respect ne soit confié à une équipe spécifique de modérateurs<sup>33</sup>.

De manière générale, **les outils de signalements situés à proximité du contenu sont visibles, facilement accessibles et aisément compréhensibles**. La procédure d'alerte ne pose pas de difficulté majeure : le chemin d'accès au dispositif de signalement comporte très peu d'étapes (un clic maximum pour accéder au formulaire de signalement), les symboles utilisés sont lisibles, les termes employés dans le formulaire de signalement sont compréhensibles. En moyenne, 4 à 5 clics sont nécessaires pour finaliser l'envoi d'un signalement. Ce chiffre est plutôt satisfaisant. Toutefois, une simplification des dispositifs pourrait permettre aux opérateurs de se conformer à la recommandation du CSA (3 clics maximum<sup>34</sup>).

Les **outils en bas de page renvoyant vers un formulaire** de demande légale semblent moins simples d'utilisation. Moins intuitifs et moins visibles, leur chemin d'accès peut être complexe et relativement long à suivre. Les termes employés dans le formulaire peuvent être plus spécialisés et, par conséquent, moins transparents que ceux employés dans la première catégorie d'outils. En outre, il a été observé que la catégorie « fausse information » n'était pas toujours présente dans le formulaire. Ces caractéristiques s'expliquent notamment par la nature des services (moteurs de recherche) et des réclamations adressées par les utilisateurs à l'opérateur, telles que des demandes de désindexations d'URL. Il demeure que par comparaison, la procédure proposée sur les moteurs de recherche et agrégateurs de contenus est moins favorable au signalement de fausses informations.

Les opérateurs avancent plusieurs raisons pour expliquer **l'absence d'une catégorie « fausse information »** parmi les motifs de signalement proposés : certains soutiennent que leur service est peu propice à la circulation de fausses informations (*LinkedIn*, *Snapchat*) ou mettent en avant les mécanismes mis en place afin d'en limiter la circulation (*Snapchat*<sup>35</sup>) ; un autre fait valoir le risque d'afflux de signalements non pertinents auquel il

<sup>33</sup> *Wikipédia* a cependant développé des outils de signalements spécifiques pour certains motifs d'illégalité présumés (violation du droit d'auteur notamment).

<sup>34</sup> Effectif sur certains services, comme *Dailymotion* : 3 clics pour les utilisateurs non- logués, 2 pour les utilisateurs logués.

<sup>35</sup> Valable pour 2019 seulement (voir note 17).



s'exposerait en proposant un motif dont l'interprétation est difficile (*Jeuxvideo.com*). À cet égard, on peut se demander dans quelle mesure l'ensemble des utilisateurs (et pas seulement les plus aguerris aux usages numériques ou sensibilisés à la problématique) se tournent spontanément vers d'autres champs pour signaler ce qui leur semblerait constituer une fausse information. De fait, ces items de substitution peuvent freiner une démarche de signalement. Par ailleurs, une telle substitution interroge sur la capacité de l'opérateur à identifier, *in fine*, les signalements de « fausse information » et établir des statistiques fiables sur l'activité des utilisateurs en termes de signalements et de propagation de fausses informations<sup>36</sup>. A ce titre, plusieurs opérateurs (*Facebook, Google, Microsoft*) ont souligné qu'ils ne souhaitaient pas être des « arbitres de la vérité », considérant qu'il n'existait pas de définition universelle des fausses informations.

### ***Affichage des motifs de signalement***

Les formulaires proposant une arborescence « à plat » des motifs de signalement permettent à l'utilisateur de prendre connaissance en une seule fois de l'ensemble des motifs, présentant l'avantage notable de limiter le risque qu'il s'égaré dans le chemin de signalement. On peut présumer que, par comparaison et par déduction, il aide à la compréhension par l'utilisateur, qui ne maîtrise pas nécessairement la signification de tous les motifs de signalement proposés.

### ***Items proposés et item « Fausse information »***

Les opérateurs sont libres de choisir **la formulation des motifs de signalement**. Cela leur permet notamment de l'adapter à leur public. Cependant, on pourrait à l'inverse considérer que leur diversité va à l'encontre de leur bonne appréhension par les utilisateurs.

### ***Fonctionnalités de l'outil de signalement***

Les fonctionnalités permettant à l'auteur d'un signalement de **compléter son alerte**, telles que l'ajout d'un champ de texte libre (ex. : *Snapchat, Dailymotion*) ou d'autres contenus que l'auteur du signalement juge problématiques (*Twitter*) ont été saluées par le comité d'experts sur la désinformation. Elles permettent notamment d'améliorer le travail de modération en précisant le contexte de publication du contenu problématique.

- ***Les informations communiquées sur la procédure d'instruction ne permettent pas au CSA d'apprécier pleinement la portée des moyens mis en œuvre pour répondre aux objectifs de la loi***

A l'exception de trois déclarations faisant état de façon exhaustive des outils, des moyens et de l'organisation de la modération (*Dailymotion, Jeuxvideo.com, Unify*), la plupart présentent des lacunes sur les points abordés par le questionnaire du CSA, qu'il s'agisse de l'organisation et des étapes de l'instruction (*Google, Snapchat, LinkedIn*), des moyens humains et automatisés employés (*Twitter, Google, Microsoft*) ou de l'information transmise aux utilisateurs sur le suivi des signalements (*Twitter, Snapchat*). Si des échanges ultérieurs ont permis de recueillir des informations complémentaires, ce défaut d'exhaustivité et de transparence est d'autant plus regrettable que des éléments d'information précis auraient

<sup>36</sup> Sans pouvoir parler d'une causalité évidente, on note à cet égard que deux des trois plateformes ayant déclaré n'avoir reçu aucun signalement pour fausse information en France en 2019 ne proposent pas ce motif dans leur formulaire (*Google* et *Snapchat*).



permis au CSA de développer une connaissance plus fine des procédures des opérateurs et ainsi de mieux identifier les besoins de ceux-ci.

#### *Les textes et critères d'évaluation des contenus signalés*

Les opérateurs se fondent sur la loi (*Jeuxvideo.com*, *Microsoft*), les règles et standards de la communauté et les CGU (*Facebook*, *Snapchat*, *Google*, *Dailymotion*, *Jeuxvideo.com*), des articles de journalistes et le travail des *fact-checkers* (*LinkedIn*), la culture locale (*Jeuxvideo.com*) et le Code de bonne conduite de l'Union européenne sur la désinformation (*Twitter*). Ces instruments peuvent être classés en deux catégories, eu égard à leur fonction pour l'opérateur.

Les **normes de référence** définissent les règles applicables sur le site (interdiction ou non de publier des fausses informations) et conditionnent le traitement d'un signalement et la nature des mesures prises ou non à son issue (retrait, suspension, réduction de la visibilité). Sur ce point, les déclarations sont peu disertes : seuls deux opérateurs ont explicitement évoqué la question de la définition des fausses informations et de l'interprétation de la loi (*Jeuxvideo.com*, *Microsoft*), les autres acteurs rappelant qu'ils appliquent la loi en vigueur dans le territoire sur lequel ils opèrent (« loi locale »).

Les **sources** visent les matériaux (journalistiques, scientifiques, académiques) sur lesquels s'appuient les modérateurs afin de vérifier le contenu et de déterminer la décision à prendre. A l'exception de *Facebook*, qui a mentionné ses partenariats avec les organismes de *fact-checking* et de *LinkedIn*, qui a cité les entreprises de presse auxquelles se réfèrent les équipes de modération, les opérateurs passent ce point sous silence (l'un d'eux (*Dailymotion*) le traite de manière confidentielle). *Snapchat* estime que le travail des entreprises de presse est valorisé par la structure et l'organisation du service, le prémunissant contre le risque de circulation de fausses informations. D'autres s'interrogent sur l'efficacité des outils journalistiques, le principe même de l'élaboration d'une « liste blanche » leur paraissant douteux (*Wikipédia*<sup>37</sup>). Ces interrogations dessinent une ligne de partage entre les opérateurs autour de la question des sources dignes de confiance.

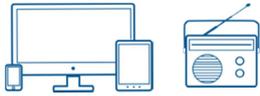
**Le CSA aurait souhaité que tous les opérateurs précisent les sources, notamment externes, sur lesquelles les équipes de modération s'appuient pour traiter les signalements. Il les invite à s'emparer de cette question afin de garantir la transparence de leurs procédés de modération.**

Il est intéressant de noter que Facebook fait état d'un système d'évaluation régulière des modérateurs et des décisions prises.

Certaines déclarations sont très prudentes quant à la communication des **critères d'ouverture d'une instruction**. Il semble donc que cette information recèle un enjeu particulier pour l'opérateur : en fixant la portée des contenus contrôlés, ces critères l'engagent d'une certaine façon quant aux moyens à mettre en œuvre pour répondre aux sollicitations des utilisateurs, pouvant ainsi conduire à augmenter la charge pesant déjà sur les équipes de modération.

---

<sup>37</sup> Accessible depuis le site de Wikipédia, le projet « Manon » relatif aux sources valorisées par les contributeurs de l'encyclopédie pose explicitement ces questions.



Une majorité d'opérateurs déclarent ne pas avoir développé de **procédures d'urgence**, notamment en période électorale. Toutefois, des dispositions particulières peuvent être prises en cas de situation exceptionnelle : un pilotage particulier des indicateurs de volume et de délai de traitement (*Jeuxvideo.com*) ou la révision de l'ordre de priorité de traitement des signalements (*Dailymotion*). De manière générale, il conviendra de voir, à l'issue de l'exercice 2020, si une évolution notable a eu lieu en raison des impératifs liés à la crise de la COVID-19, dans le cadre de laquelle les opérateurs se sont fortement mobilisés.

### **Les moyens humains, financiers et automatisés employés**

L'instruction des signalements de fausses informations est principalement réalisée par des modérateurs humains, en interne pour la plupart (sauf *Facebook*). Les effectifs globaux des équipes de modération oscillent entre 15 000 (*Facebook*) et 15 (*Jeuxvideo.com*), révélant les disparités structurelles et économiques des opérateurs déclarants. Les équipes humaines sont généralement actives 24 heures sur 24.

Les moyens automatisés sont utilisés principalement pour l'identification de certains contenus tels que les spams, les faux comptes, les publicités contraires aux CGU (*Facebook*, *Microsoft*) et pour l'arrêt des actions nuisibles ou malfaisantes telles que le vandalisme<sup>38</sup> (*Wikipédia*).

Les informations fournies sont, sur certains points, évasives (à quelques exceptions près, comme *Facebook*<sup>39</sup>). Pour justifier ce choix, certains font valoir la nécessité de protéger les équipes de modération. Si la taille des effectifs est un élément plutôt renseigné par les opérateurs, des éléments précis n'ont pas toujours été donnés sur la formation des modérateurs et l'organisation du travail (qu'il s'agisse du traitement des signalements ou de celui des éventuels recours). Or, ces informations, loin d'être accessoires, garantissent la transparence et l'équité des procédures.

### **Les actions et mesures des opérateurs de plateformes à l'égard des contenus**

Les mesures qui peuvent être prises à l'égard des contenus signalés sont diverses : retrait, réduction de la visibilité, désindexation de l'URL, apposition d'un filtre sur le contenu, avertissement de l'utilisateur interagissant avec le contenu. Les différences de positionnement des opérateurs quant à la publication de fausses informations renvoient à aux disparités relevées dans les formulaires de signalement.

Il est intéressant de noter qu'un contenu marqué comme contenant une fausse information sur *Facebook* le sera automatiquement sur *Instagram*, et inversement.

<sup>38</sup> Sur *Wikipédia*, le « vandalisme » désigne la détérioration volontaire de l'encyclopédie en tant que comportement visant délibérément à porter atteinte à l'objectif encyclopédique du projet. Il peut s'agir de détérioration délibérée d'un contenu, de suppression d'un contenu compromettant, de promotion d'un point de vue, de suppression de tout ou partie d'un article sans explication, d'insertion au sein d'un article d'éléments erronés mais qui paraissent cohérents et recourant parfois à l'utilisation de fausses sources ou de sources détournées, appelé dans ce cas « vandalisme sournois ». La pratique de vandalisme peut conduire à un blocage en écriture temporaire ou indéfini dans les cas les plus graves.

<sup>39</sup> Informations sur les partenaires, l'implantation géographique, les compétences linguistiques et en matière de culture locale des modérateurs, leur formation et leur encadrement notamment psychologique.



### **Sur le suivi et l'information des utilisateurs et sur la publicité des mesures**

Certains opérateurs ont fait le choix d'informer des mesures de retrait prises à l'égard des contenus soit l'ensemble des utilisateurs (*YouTube, Dailymotion, Wikipédia*) soit ceux directement concernés, à savoir l'auteur du contenu (*Snapchat*) et l'auteur du signalement (*Microsoft, Facebook*). D'autres n'informent pas les utilisateurs en cas de retrait d'un contenu (*LinkedIn, Jeuxvideo.com, Doctissimo, Aufeminin*). *Facebook* affiche un message d'information avertissant l'utilisateur lorsque le contenu est jugé sensible mais n'est pas retiré. Les motifs de la décision prise par l'opérateur sont, semble-t-il, rarement communiqués. Enfin, les opérateurs ont peu développé d'outils permettant d'informer les internautes ayant interagi avec un contenu problématique, à l'exception de *Facebook*.

L'information des utilisateurs apparaît comme un outil crucial dans la lutte contre la circulation des fausses informations, celle-ci résidant précisément dans la viralité du contenu. L'absence d'information des utilisateurs ne favorise pas leur autonomisation et leur responsabilisation en ce qui concerne les contenus qu'ils publient.

### **Sur l'existence de voies de recours**

Les voies de recours peuvent être ouvertes aux seuls auteurs de contenus signalés (*Snapchat, Doctissimo, Aufeminin*), aux auteurs de contenus signalés ainsi qu'aux auteurs de signalement (*Jeuxvideo.com, Facebook*<sup>40</sup>) ou à tout utilisateur s'estimant lésé (*LinkedIn*). Elles peuvent être ouvertes à certaines catégories d'utilisateurs seulement, tels que les partenaires professionnels (*Dailymotion*). Certains opérateurs veillent à ce que les recours dirigés contre une décision soient instruits par une autre équipe ou par un autre modérateur que celle ou celui l'ayant rendue (*LinkedIn, Snapchat, Facebook*).

Les déclarations recèlent peu d'éléments concrets concernant la procédure suivie en interne en cas de recours et l'information des utilisateurs sur l'exercice de ces voies de recours. Il serait souhaitable que les informations fournies au CSA, mais plus encore au public, sur cette question soient claires, exhaustives et transparentes. La responsabilisation des utilisateurs n'est possible qu'à la condition que le cadre dans lequel s'exercent leur liberté d'expression soit équilibré et prévisible.

#### **• Les données chiffrées**

Au regard des chiffres déclarés pour 2019, le nombre de signalements de fausses informations sur les plateformes oscille entre 40 000 (*Jeuxvideo.com*) et 0 (*Google Search*).

Néanmoins, les données chiffrées demandées par le CSA dans son questionnaire ont fait l'objet d'une communication limitée de la part des déclarants. Ce silence s'explique, d'une part, par l'absence d'indicateurs et d'outils de comptage internes (absence de catégorie « Fausse information » et absence d'outil statistique) et, d'autre part, par le refus de certains de communiquer ces chiffres, soit parce qu'ils sont réputés être couverts par le secret des affaires, soit parce qu'ils sont jugés non pertinents (*Facebook*).

---

<sup>40</sup> Cela concerne les contenus examinés à l'aune des Standards de la communauté et non les contenus « fact-checkés ».



Bien que les conditions d'une analyse comparée systématique n'aient pas été réunies en l'espèce, compte tenu de la diversité des acteurs concernés, la communication des données chiffrées aurait permis au CSA de disposer d'éléments concrets lui donnant la mesure de la tâche et des efforts déployés par les opérateurs. L'absence de ces données limite fortement la capacité à appréhender l'ampleur du phénomène de la désinformation sur un service et d'établir des statistiques fiables sur les fausses informations, leur circulation, les signalements et, partant, de développer les dispositifs nécessaires pour l'endiguer.

- **Les discussions entre opérateurs**

La plupart des opérateurs ont indiqué ne pas avoir entamé de discussions avec leurs pairs concernant l'harmonisation des outils de signalement. Certains estiment qu'une telle harmonisation est peu souhaitable et difficilement réalisable, compte tenu de la diversité des modèles (*Snapchat*) et de la surface économique et financière (*Jeuxvideo.com*) des plateformes, comme de leur vocation (*Wikipédia*).

Cet état de fait peut en outre s'expliquer par l'absence d'instance de dialogue comparable à ce qui existe en matière de lutte contre les contenus faisant l'apologie du terrorisme, le GIFCT<sup>41</sup> (*LinkedIn*). Or, la mutualisation des informations et des données, avec le concours éventuel d'organismes de vérification des faits, pourrait constituer un puissant levier de lutte contre la diffusion des fausses informations.

Il convient toutefois de noter les efforts d'harmonisation internes aux groupes opérant plusieurs services (ex : *Facebook* et *Instagram*).

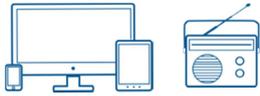
### 3.1.3. Recommandations du CSA

Le CSA salue les dispositifs mis en œuvre par les opérateurs pour permettre à leurs utilisateurs de signaler aisément les contenus susceptibles de propager de fausses informations, ainsi que les efforts réalisés pour en rendre compte avec clarté et précision dans leur déclaration. Il formule toutefois les recommandations suivantes :

1. 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.
2. 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).

---

<sup>41</sup> Pour Global Internet Forum to Counter Terrorism (ou Forum mondial de l'Internet contre le terrorisme), fondé en 2017 à l'initiative de Facebook, Microsoft, Twitter et YouTube.



3. 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.
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 des 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.
5. 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**.
6. **L'information des utilisateurs mériterait d'être améliorée** sur plusieurs points tels que :
  - 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 utilisateurs concernés.
7. Les opérateurs sont **encouragés à développer et à formaliser une procédure d'urgence** pour le signalement de fausses informations en cas de circonstances exceptionnelles.

## 3.2. Transparence des algorithmes

La transparence des algorithmes a fait l'objet de réponses de la part de tous les opérateurs à l'exception de *Jeuxvideo.com* et d'*Unify*, qui déclarent ne pas utiliser d'algorithmes sur leurs services.

**Si le CSA salue la coopération de certains opérateurs sur ce point ainsi que leurs réponses à ses demandes de compléments, il constate néanmoins que les éléments apportés diffèrent en qualité comme en quantité.** Ils n'ont pas toujours permis d'apprécier de manière complète les moyens mis en œuvre en matière de transparence des algorithmes vis-à-vis des utilisateurs.



### 3.2.1. Typologie des mesures prises par les opérateurs

APPROCHES DE LA TRANSPARENCE DES OPÉRATEURS	
L'information des utilisateurs sur le fonctionnement des algorithmes est variable	<p>Emplacement des informations :</p> <ul style="list-style-type: none"> <li>- <b>à partir d'un outil contextuel</b><sup>42</sup> (<i>Snapchat</i>) ;</li> <li>- dans le <b>centre d'aide ou les politiques de confidentialité</b> (<i>Dailymotion, Twitter, Yahoo, LinkedIn, Snapchat</i>).</li> </ul> <p>Contenus des informations : <b>raisons pour lesquelles les utilisateurs sont exposés</b> à certains types de contenus comme les pages suivies, les contenus précédemment consultés, etc., sans précision sur leur caractère exhaustif ou non (<i>Facebook, LinkedIn, Microsoft, Snapchat</i>).</p>
Les opérateurs laissent l'utilisateur plus ou moins libre dans le réglage des paramètres	<p>Plusieurs plateformes revendiquent une volonté de transparence afin de <b>protéger</b> les utilisateurs et, dans cette optique, offrent une <b>faculté de pré-réglages</b> (<i>Google, Microsoft, Yahoo</i>).</p> <p>D'autres mettent l'accent sur <b>la compréhension de leurs algorithmes et la possibilité pour les utilisateurs de changer leurs préférences de manière plus granulaire</b> (<i>Facebook, LinkedIn, Snapchat, Twitter</i>).</p> <p>Deux plateformes soulignent la possibilité d'apposer des filtres qui modifient les critères de recherche (<i>Microsoft, Google</i>).</p>
Certains opérateurs signalent plusieurs motifs de réserve vis-à-vis de la transparence	<p>Motifs de réserve :</p> <ul style="list-style-type: none"> <li>- <b>éviter tout détournement par des acteurs malveillants</b> (<i>Facebook, Google, LinkedIn</i>) ;</li> <li>- risques liés à la <b>concurrence</b> car leurs algorithmes font partie de leur modèle économique (<i>Snapchat</i>).</li> </ul>
Des opérateurs proposent des outils de modification du comportement de leurs algorithmes	<p>Certains algorithmes sont en partie <b>personnalisables</b>, l'utilisateur pouvant décider à quels types de contenus il souhaite être exposé.</p> <p>Les outils peuvent être :</p> <ul style="list-style-type: none"> <li>- <b>contextuels (placés à côté de la publication)</b> et permettent des contrôles (<i>Dailymotion, Facebook, Google, Twitter, LinkedIn, Microsoft, Snapchat</i>).</li> <li>- <b>généraux, disponibles dans les réglages</b> (<i>Dailymotion, Facebook, Google, Twitter, LinkedIn, Microsoft, Snapchat, Yahoo</i>).</li> </ul>
Les opérateurs soulignent	<p><b>Publication d'informations à destination des développeurs</b> (<i>Dailymotion, Google, LinkedIn</i>).</p>

<sup>42</sup> Réglage ou outil contextuel : fonctionnalité accessible à l'utilisateur directement depuis un contenu qui lui est proposé lors de son utilisation du service cité, par exemple en cliquant dessus. Ce type de fonctions s'oppose aux réglages ou outils dits généraux, accessibles au sein de pages d'aide, de sections de réglages ou encore de politiques de confidentialité couvrant des sujets non pas liés à un contenu en particulier mais à l'ensemble de ceux qui peuvent être proposés aux utilisateurs.



l'adaptation de leurs contenus à des publics différents	Volonté de donner du contrôle aux utilisateurs en réalisant un effort pédagogique d' <b>accessibilité de l'information</b> à leur égard ( <i>Facebook, Snapchat, Twitter, Yahoo</i> ).
---	--

### TYPES D'ALGORITHMES UTILISÉS PAR LES OPÉRATEURS

Description des algorithmes d'organisation, de sélection et d'ordonnement utilisés sur les services des opérateurs	<p><b>Organisation :</b></p> <ul style="list-style-type: none"> <li>- <b>des opérateurs décrivent leurs algorithmes d'indexation et comment ceux-ci régissent le flux d'actualités</b> (<i>Dailymotion, LinkedIn<sup>43</sup>, Microsoft, Snapchat, Twitter, Yahoo</i>) ;</li> <li>- autres plateformes : pas d'information sur cet aspect.</li> </ul> <p><b>Sélection : ces algorithmes permettent de recommander certains contenus aux utilisateurs, que ce soit des vidéos, des comptes à suivre ou encore des résultats de requête</b> (<i>Google, LinkedIn, Snapchat, Twitter, Wikipédia, Yahoo</i>).</p> <p><b>Ordonnement des contenus d'actualités et publicitaires</b> (<i>Dailymotion, Google, LinkedIn, Microsoft, Snapchat, Yahoo</i>).</p>
Données utilisées par ces algorithmes	<p>Origine des données :</p> <ul style="list-style-type: none"> <li>- <b>données des partenaires</b> (<i>Dailymotion, Facebook, Google, Microsoft, Snapchat, Yahoo</i>) ;</li> <li>- <b>données des utilisateurs</b> (<i>Dailymotion, Facebook, Google, LinkedIn, Microsoft, Snapchat, Wikipédia<sup>44</sup>, Yahoo</i>).</li> </ul> <p><b>Liste des critères régissant le fonctionnement de leurs algorithmes :</b> <i>Google, LinkedIn, Twitter</i> ont donné plusieurs <b>exemples</b> dont :</p> <ul style="list-style-type: none"> <li>- mots utilisés dans la requête,</li> <li>- genre d'informations que les internautes recherchent,</li> <li>- date de publication des contenus,</li> <li>- connexions des utilisateurs,</li> <li>- nombre de vues sur un contenu,</li> <li>- interactions réalisées avec les publications,</li> <li>- réseau de l'utilisateur,</li> <li>- interprétation des erreurs d'orthographe,</li> <li>- gestion des synonymes,</li> <li>- zone géographique des internautes.</li> </ul> <p><b>Changements conduits</b> en 2019, tels que :</p> <ul style="list-style-type: none"> <li>- exclusion des contenus amateurs des algorithmes de recommandation et d'indexation (<i>Dailymotion</i>) ;</li> </ul>

<sup>43</sup> *LinkedIn* a fourni des informations détaillées sur le modèle de régression logistique de ses algorithmes d'indexation.

<sup>44</sup> Pour *Wikipédia* : type d'appareil utilisé, type et version du navigateur, préférence linguistique du navigateur, type et version du système d'exploitation, nom du fournisseur d'accès internet, site web qui redirige vers *Wikipédia* ainsi que la date et l'heure de chaque demande.



	<ul style="list-style-type: none"> <li>- prise en compte de la vitesse de chargement des pages dans les algorithmes (<i>Google</i>).</li> </ul> <p>Google transmet en amont une information à destination des propriétaires de sites des modifications opérées sur le service afin de leur donner de la visibilité.</p>
--	---

**APPROCHES ALGORITHMIQUES ŒUVRANT À LA LUTTE CONTRE LES FAUSSES INFORMATIONS**

<p>Mesures restrictives en matière de diffusion de contenus, notamment sponsorisés</p>	<p><b>Contenus sponsorisés</b></p> <ul style="list-style-type: none"> <li>- pas de possibilité d'améliorer le référencement d'une information se rattachent à un débat d'intérêt général contre rémunération (<i>Dailymotion</i>);</li> <li>- démonétisation ou masquage du contenu problématique (<i>Facebook, LinkedIn</i>);</li> <li>- examen des annonceurs (<i>Microsoft</i>).</li> </ul> <p><b>Autres</b></p> <ul style="list-style-type: none"> <li>- exclusion systématique des contenus amateurs des algorithmes d'indexation et de recommandation (<i>Dailymotion</i>).</li> </ul>
<p>Modération automatique (utilisation d'algorithmes pour modérer les fausses informations)</p>	<p><b>Modération automatique</b> afin de :</p> <ul style="list-style-type: none"> <li>- <b>prioriser les signalements,</b></li> <li>- <b>repérer les contenus marqués d'une empreinte numérique par l'équipe de modération</b> (<i>Dailymotion</i>);</li> <li>- recourir à de l'apprentissage machine (« <i>machine learning</i> ») (<i>Dailymotion, Facebook, Google, LinkedIn</i>), qui permet notamment de détecter les appâts à clics ou les copies des contenus jugés faux (<i>Facebook</i>) et les faux comptes (<i>Microsoft, Twitter, LinkedIn</i>).</li> </ul> <p><b>Algorithmes modifiant la visibilité des contenus :</b></p> <ul style="list-style-type: none"> <li>- en <b>augmentant la visibilité</b> des contenus de « haute autorité » ou en provenance de sources fiables ;</li> <li>- en <b>diminuant la visibilité</b> des contenus relayant de fausses informations (<i>Dailymotion, Facebook, Google, Microsoft, Twitter</i>).</li> </ul> <p><i>Facebook</i> fournit dans sa déclaration les informations utilisées pour améliorer le fonctionnement de son dispositif basé sur l'apprentissage machine (retours des utilisateurs, commentaires, signaux liés à la source).</p> <p><i>Microsoft</i> fait état des signaux utilisés : réputation d'un site, niveau de discours sur le site, niveau de distorsion vis-à-vis d'autres sources d'information, degré d'origine des informations présentes sur le site.</p>
<p>Outil d'interaction en temps réel</p>	<p><b>Aucun opérateur</b> n'a fait état de l'existence d'un outil d'interaction en temps réel permettant spécifiquement aux utilisateurs d'obtenir des informations personnalisées ou d'exercer des modifications dans leurs préférences algorithmiques et d'en voir les résultats.</p>



### **Mesures prises dans le cadre de la crise de la COVID-19**

Certains opérateurs ont **adapté le fonctionnement de leurs algorithmes** :

- *Microsoft* a accru l'exposition des utilisateurs à des sources d'informations sur la COVID-19 faisant autorité et privilégié les informations d'actualité pour les résultats des requêtes COVID-19 sur Bing ;
- sur *Facebook*, en cas de requête sur le coronavirus, l'utilisateur s'est trouvé redirigé vers le site du Gouvernement.

### 3.2.2. Analyse des mesures prises par les opérateurs

- **Qualité des informations transmises au CSA**

Le niveau de détail des informations transmises au CSA dans certaines déclarations est le même que celui des informations disponibles pour l'utilisateur. Cela n'a pas permis au CSA d'évaluer pleinement les moyens mis en œuvre en matière de transparence des algorithmes. **Une connaissance plus complète du contexte dans lequel ils opèrent sur les plateformes – quels algorithmes, agissant sur quels contenus, avec quelles données et selon quels critères – aurait sans doute permis une meilleure appréciation des efforts mis en place par ces opérateurs en matière de construction d'un esprit éclairé chez les utilisateurs.** Cela aurait permis d'analyser les principes mis en œuvre en la matière et d'étudier les informations afférentes, si ce n'est de manière exhaustive, du moins plus amont que ce qui est actuellement délivré en aval aux utilisateurs de ces services.

Certains **opérateurs** (*Facebook, Microsoft*) **ont ciblé une portion congrue de la thématique** en transmettant des déclarations centrées sur les algorithmes utilisés pour la lutte contre la désinformation en général ou celle liée à la crise sanitaire de la COVID-19. Ces éléments, précis, n'ont cependant pas toujours permis au CSA de mener son travail d'instruction dès lors qu'il ne disposait pas d'une information complète ni de bases comparables entre opérateurs de plateformes. En outre, **le CSA rappelle que la nocivité des contenus propageant de fausses informations tient notamment à leur degré d'exposition aux utilisateurs et à leur viralité**, qui peuvent en rendre l'impact important avant même qu'ils ne soient identifiés par l'opérateur ou ses partenaires : aussi, il est essentiel d'inclure dans la réflexion le fonctionnement des algorithmes utilisés pour mettre en avant les contenus en général, et pas seulement ceux permettant de les déceler et de les modérer.

- **Types d'algorithmes utilisés par les plateformes**

Si les réglages proposés aux utilisateurs en matière d'algorithmes d'organisation, de sélection et d'ordonnement de contenus ont été bien documentés (*Dailymotion, Google, LinkedIn, Twitter, Wikipédia*), **aucune liste complète d'algorithmes n'a été fournie par les**



**opérateurs**, même limitée à une description très synthétique de leurs principes de fonctionnement et objectifs poursuivis.

De manière générale, **il a été délicat de déterminer si les informations données étaient exhaustives ou non**. Seul *LinkedIn* a fait preuve d'efforts plus poussés dans l'explication de son modèle en fournissant des éléments techniques ; *Google* a quant à lui fourni des éléments d'explication intéressants sur les étapes de fonctionnement des algorithmes de *YouTube*.

Ce constat tient au fait que le fonctionnement des algorithmes est éminemment complexe à retranscrire, notamment dès lors qu'ils fonctionnent avec une intervention humaine plus ou moins limitée. Par ailleurs, le CSA comprend également la nécessaire confidentialité sur ce sujet, au cœur des modèles d'affaires des opérateurs. Possibilité leur avait donc été laissée, sur le fondement du secret des affaires, de fournir au CSA des versions confidentielles de leurs déclarations. Seul *Dailymotion* a utilisé cette possibilité. En outre, les informations attendues par le CSA sur les principes de fonctionnement et objectifs de certains algorithmes pouvaient être fournies sans dévoiler les modèles d'affaires.

Il convient de noter par ailleurs que *Yahoo* a indiqué dans sa déclaration que les algorithmes utilisés sont ceux de *Bing* et que les modalités de leur gestion relèvent de *Microsoft*.

- **Transparence des opérateurs de plateformes vis-à-vis de leurs utilisateurs**

Certains services de plateformes offrent des **fonctionnalités intéressantes permettant aux utilisateurs d'intervenir**, dans une certaine mesure, sur le choix algorithmique des contenus qui leur sont proposés.

Un certain nombre d'entre elles **fournissent également certaines informations sur le fonctionnement des algorithmes de recommandation, sélection et organisation**, accessibles de manière soit contextuelle (depuis le contenu), soit générale.

**Le CSA regrette néanmoins que davantage d'informations sur les critères ou les données utilisées par les opérateurs dans le cadre de ces algorithmes n'aient été communiquées aux utilisateurs en 2019<sup>45</sup>**. Les déclarations reçues n'ont pas permis non plus d'appréhender l'ensemble des possibilités qui leur sont offertes en matière de réglages, des exemples donnés par les opérateurs s'étant parfois substitués à des explications plus générales sur leurs approches concrètes en matière de transparence des algorithmes.

---

<sup>45</sup> *Facebook* a cependant mis en œuvre, en 2020, des moyens dont il fait état dans sa déclaration et qui augurent d'un constat plus positif en la matière. Ainsi en est-il de la mise en place de sa section « Activité en dehors de Facebook » depuis le 28 janvier 2020 qui permet d'accéder à un résumé des informations envoyées à la plateforme par d'autres applications et sites internet via ses outils commerciaux, de déconnecter ces informations de son compte ou bien de consulter toute future activité enregistrée en dehors de Facebook.



### 3.2.3. Recommandations du CSA

**Le CSA considère, en l'état des informations reçues, que l'ensemble des opérateurs ayant répondu n'offrent pas à leurs utilisateurs les moyens d'exercer de manière éclairée leur esprit critique sur les contenus qui leur sont proposés par voie algorithmique.**

Aussi, il formule les recommandations suivantes :

1. Dans le cadre du prochain bilan, il encourage les opérateurs **à lui communiquer davantage d'éléments sur l'intelligibilité<sup>46</sup> et la redevabilité<sup>47</sup>, de leurs algorithmes d'organisation, de sélection et d'ordonnement des contenus.**

L'ensemble de ces éléments lui permettra de mener à bien son travail d'instruction et de formuler des recommandations concrètes à destination des opérateurs, tout en respectant les impératifs économiques qui ont été soulevés par ces derniers dans le cadre des échanges avec le régulateur. Le CSA est conscient que le modèle d'affaires de certaines plateformes repose en grande partie sur la performance de leurs algorithmes. Pour autant, il estime qu'il leur est possible de fournir des éléments lui permettant de comprendre les principes de fonctionnement et les objectifs de certains d'entre eux sans dévoiler leur modèle d'affaires.

2. **Le CSA recommande une transparence accrue en matière de fonctionnement des algorithmes vis-à-vis des utilisateurs.** Ces informations sont à ce stade fournies au sein de rubriques denses tels que les centres d'aide comme les politiques de confidentialité, ce qui n'en facilite pas toujours l'accessibilité et la compréhension. D'une manière générale, il considère que **les opérateurs devraient être proactifs pour**

---

<sup>46</sup> L'intelligibilité est entendue ici comme « explicabilité » des algorithmes, notion qui renvoie à la publicité d'éléments relatifs à la finalité des traitements pour lesquels les algorithmes ont été programmés, afin d'informer les utilisateurs auxquels ces traitements s'appliquent. Les règles définissant ces traitements, les principales caractéristiques de leur mise en œuvre, les données traitées et leurs sources, les paramètres de traitement et leur pondération, les opérations effectuées par le traitement et, le cas échéant, le degré et mode de contribution à la décision de, par exemple de proposer un contenu plutôt qu'un autre, font partie des éléments communicables en vertu de ce principe. Celui-ci a notamment été cité dans le cadre des relations administrations – usagers par le décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique, pris en application de l'article L. 311-3-1 du code des relations entre le public et l'administration, tel que modifié par l'article 4 de la loi du 7 octobre 2016 pour une République numérique.

<sup>47</sup> La redevabilité des algorithmes se définit comme « l'attribution explicite des responsabilités impliquées par le fonctionnement des algorithmes » selon la Commission nationale de l'informatique et des libertés (CNIL), définition citée page 51 de son rapport *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, paru en septembre 2017. Il s'agit de connaître, pour des traitements algorithmiques donnés, les personnes responsables de leur définition, de leur mise en œuvre et de la prise de décision qui en découle, ainsi que leurs étapes et degrés d'intervention par rapport à la machine. Ceci est d'autant plus important que les algorithmes développés récemment reposent sur une part d'apprentissage (« machine learning ») dont le détail des résultats par exemple en matière de classement des signaux échappent aux concepteurs mêmes de ces modèles..



**proposer ces informations** nécessaires à la compréhension et à la bonne utilisation de leurs services.

3. **Le CSA recommande également une transparence accrue en matière de capacité de réglage des algorithmes par les utilisateurs de ces plateformes.** Les réglages contextuels, qu'offrent déjà plusieurs opérateurs, permettent notamment d'améliorer ce niveau de transparence. Une attention doit également être portée à l'ergonomie afin de s'assurer que ce type de réglages est facilement accessible en un minimum d'actions de la part de l'utilisateur.

Le Conseil envisage de faire **évoluer ces deux points dans le questionnaire** qu'il adressera aux plateformes dans la perspective de l'élaboration du bilan d'application des mesures mises en place en 2020.

4. Enfin, ainsi que l'ont souligné les membres du comité d'experts sur la désinformation en ligne, 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, qui ont conduit à un usage accru des systèmes automatisés. Le CSA invite les opérateurs à lui **fournir toutes les informations utiles sur les rôles et degrés respectifs de la part de curation humaine d'une part, et de celle des algorithmes d'autre part, dans la détection et le traitement des fausses informations** dans leur prochaine déclaration annuelle.

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

#### 3.3.1. Typologie des mesures prises par les opérateurs

La lutte contre la propagation des fausses informations sur Internet ne peut reposer uniquement sur la suppression des comptes et contenus véhiculant ce type d'informations. Elle réside également dans la capacité à détecter puis analyser les contenus trompeurs et à assurer la diffusion d'une information vérifiée provenant de sources auxquelles tout utilisateur peut se fier. C'est le sens des recommandations formulées par le CSA dans sa recommandation du 15 mai 2019 en ce qui concerne la promotion de contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle (ci-après EAP-SCA), articulées autour de trois types de mesures : **l'identification des sources des contenus issus de ces organismes** au moyen d'indicateurs clairement visibles ; **la prise en compte des démarches de labellisation** et notamment de celles réalisées par ces mêmes organismes ; la mise en avant des informations provenant de ces sources, en particulier de « **fact-checking** ».

Deux plateformes (*Jeuxvideo.com* et *Wikipédia*) indiquent **ne pas prendre de mesure en la matière** en raison de la nature même de leur service : *Jeuxvideo.com* estime que dès lors que seuls des individus ayant accès au forum, les contenus issus d'EAP-SCA ne peuvent y être



publiés ; Wikipédia, dont les articles sont uniquement issus et revus par des contributeurs, place le principe de sources « fiables » au cœur de son fonctionnement.

*Aufeminin* ne dispose pas de journalistes et n'a pas conclu de contrats avec des agences de presse ou mis en place d'actions de *fact-checking*.

Deux plateformes déclarent sélectionner en amont les partenaires médias dont les contenus sont disponibles au sein de tout ou partie de leur service (*Snapchat* (dans la partie Discover) et *Yahoo Portal*) et estiment moins nécessaire, de ce fait, le recours à la vérification des informations publiées.

IDENTIFICATION DES SOURCES DES CONTENUS	
Mesures relatives à l'identification des sources	<p><b>Identification réalisée en interne :</b></p> <ul style="list-style-type: none"> <li>- après un travail humain et à l'aune de critères préétablis, la plateforme décide de <b>certifier le compte et/ou la chaîne vidéo d'un utilisateur (non réservé aux EAP-SCA) ;</b></li> <li>- l'éditeur de contenu est sélectionné par la plateforme et contractualise directement avec celle-ci (<i>Doctissimo, Snapchat et Yahoo Portal</i>).</li> </ul> <p><b>Identification réalisée en externe par les EAP-SCA ou un tiers :</b></p> <ul style="list-style-type: none"> <li>- la source (EAP-SCA ou un autre) <b>renseigne sur son compte les informations sur son identité</b> (<i>Twitter, LinkedIn</i>) ;</li> <li>- les EAP-SCA peuvent « <b>marquer</b> » leurs contenus : cas de la <b>fonctionnalité ClaimReview</b> (<i>Google, Microsoft</i>) où l'éditeur peut apposer le libellé « <i>fact-checking</i> » au contenu de <i>fact-checking</i> qu'il édite ;</li> <li>- l'identification de la source est <b>réalisée par un tiers</b> : module « NewsGuard » (<i>Microsoft</i>) qui évalue la fiabilité des sites d'information à l'aune de 9 critères et émet une note de fiabilité.</li> </ul>
Mesures relatives aux indicateurs visuels	<ul style="list-style-type: none"> <li>- <b>Apposition d'un visuel</b> (badge, logo, icône, bouton d'informations) : <ul style="list-style-type: none"> <li>- visuel certifiant la « <b>fiabilité</b> » d'un éditeur (<i>Dailymotion</i>) ;</li> <li>- bouton d'<b>information sur les éditeurs</b> relayant du contenu externe (<i>Facebook, Google Search, YouTube</i>).</li> </ul> </li> <li>- <b>visuel commun à tout type d'utilisateurs (EPA-SCA ou non)</b> : badge certifiant le compte de l'utilisateur comme étant authentique (<i>Twitter</i>) et/ou digne de confiance (<i>Dailymotion</i>).</li> <li>- <b>visuel spécifique aux EAP-SCA :</b> <ul style="list-style-type: none"> <li>- fonctionnalités permettant d'<b>identifier visuellement</b>, de façon plus ou moins évidente, les contenus provenant d'éditeurs de presse (les boutons de contexte permettant d'avoir de plus amples informations sur l'éditeur de <i>Facebook</i><sup>48</sup>, <i>Instagram, YouTube, Google et Microsoft</i>) ;</li> <li>- identification spécifique des contenus de <b>fact-checking</b> (<i>Google, Bing</i>) ;</li> <li>- identification se résumant à <b>mentionner le nom de l'utilisateur</b> (EAP-SCA ou non) ayant publié un contenu (<i>Yahoo</i>).</li> </ul> </li> </ul>

<sup>48</sup> Facebook envisage de labelliser les médias qui sont sous le contrôle d'un État.



	<b>Visuel indépendant de la plateforme</b> : dans les faits, les partenaires média rendent visible leur logo sur les vignettes des contenus qu'ils publient ( <i>Snapchat</i> ).
<b>PRISE EN COMPTE DES DÉMARCHES DE LABELLISATION ET GARANTIE DE LA FIABILITÉ DES INFORMATIONS</b>	
Sélection de l'EAP-SCA sur le critère de sa participation à une démarche de labellisation	<p>La plateforme recherche :</p> <ul style="list-style-type: none"> <li>- <b>l'adhésion à une charte éthique ou déontologique</b> pour identifier une source d'EAP-SCA ;</li> <li>- <b>la participation à des initiatives de « fact-checking »</b> et la mise en place d'outils et/ou dispositifs de « fact-checking » chez le tiers afin de labéliser son compte ou son contenu (<i>Google, Microsoft via Claim Review</i>).</li> </ul> <p>Possibilité <b>d'intervenir en tant que tiers vérificateur conditionnée</b> à l'appartenance à des organisations reconnues (IFCN<sup>49</sup>) (<i>Facebook</i>).</p> <p>Mise en avant des seules sources de confiance dans les recommandations (<i>Dailymotion</i>).</p>
Règles posées par la plateforme	<p>La plateforme communique préalablement ses propres <b>lignes directrices</b>, que les EAP-SCA partenaires doivent suivre (<i>Snapchat</i>).</p> <p>Activité <b>d'édition de contenus en propre par la plateforme</b>, selon des règles incluant l'absence de fausses informations (<i>Doctissimo, LinkedIn</i>).</p>
Partenariats et démarches	Participation, notamment financière, à des démarches, initiatives et programmes de recherches pour <b>développer la qualité du travail journalistique numérique</b> : <i>Facebook</i> ( <i>Code de bonnes pratiques contre la désinformation en ligne de l'UE</i> ), <i>Google</i> (IFCN, GNI, Trust Project...), <i>Twitter</i> ( <i>Disinfo Lab</i> ).
<b>MISE EN AVANT DES CONTENUS ISSUS DES EAP-SCA ET NOTAMMENT DE FACT-CHECKING</b>	
<p>Si plusieurs orientations et types de mesures peuvent se dessiner en matière de mise en avant des contenus issus des EPA-SCA, un élément commun se dégage : l'intervention algorithmique (<i>Google, Microsoft, Facebook</i>). Plusieurs démarches sont à noter à cet égard :</p> <ul style="list-style-type: none"> <li>- mise en avant, de façon active, des contenus issus des EPA-SCA, en assurant un <b>certain référencement</b>, en dédiant à ces contenus des canaux spécifiques ;</li> <li>- <b>dégradation de la visibilité des sources réputées peu « fiables »</b> (ayant pour effet de faire remonter les informations de sources fiables) ;</li> <li>- apport d'<b>éléments de contexte</b> lorsqu'une information a été vérifiée et jugée fausse.</li> </ul>	
Mise en avant des contenus	<b>Canaux de diffusion dédiés</b> : page d'accueil de la plateforme ou pages spécialement dédiées aux contenus provenant de sources dites « fiables » dont certains EPA-SCA ( <i>Dailymotion, Snapchat, LinkedIn, YouTube</i> ).

<sup>49</sup> International Fact-Checking Network.



	<p><b>Sélection algorithmique de contenus provenant de sources « fiables »</b> (<i>Google, LinkedIn, Bing, Snapchat, Dailymotion</i>) afin de les faire apparaître de manière préférentielle dans les listes de résultats.</p> <p>→ Ces deux types de mesures vont généralement ensemble, la sélection algorithmique (reposant sur le nombre d'internautes ayant souligné la fiabilité d'un contenu et les signaux d'engagement des utilisateurs) des contenus provenant des sources fiables permettant de les faire apparaître de façon préférentielle ou au sein de canaux de diffusions dédiés.</p>
<p>Mise en avant des contenus de « <i>fact-checking</i> »</p>	<p><b>Fenêtre pop-up invitant à accéder aux contenus des tiers vérificateurs</b> lorsqu'un utilisateur s'apprête à partager un contenu jugé faux par ces derniers et <b>filtres</b> sur ce contenu (<i>Facebook</i>).</p> <p>Possibilité de <b>rechercher uniquement des contenus fact checkés</b> (<i>Bing, Google</i>).</p> <p>Affichage, sous le résultat de recherche, d'<b>éléments de contexte</b> et de liens renvoyant vers du contenu de <i>fact-checking</i>.</p> <p>Absence de dispositif de vérification des faits (<i>Twitter</i><sup>50</sup> ; <i>Dailymotion</i>).</p>
<p>« <i>Fact-checking</i> » confié à des tiers vérificateurs</p>	<p><b>De manière directe : accès privilégié des EAP-SCA directement à la plateforme via des outils dédiés :</b> <i>Facebook</i> dispose d'un système faisant remonter aux tiers vérificateurs partenaires les fausses informations signalées par les utilisateurs de la plateforme ; les tiers vérificateurs peuvent également se saisir d'une vérification. Ils peuvent publier leurs contenus de vérification en regard du contenu vérifié et jugé comme faux.</p> <p><b>Avec un intermédiaire entre les EAP-SCA et la plateforme</b> (<i>Bing avec NewsGuard</i>) : utilisés par certains moteurs de recherche, les tiers vérificateurs participent à un travail global de certification de sites dont la fonctionnalité <i>NewsGuard</i> fait ensuite état en signalant grâce à un symbole (bouclier coloré) la fiabilité d'une source.</p> <p><b>Marquage par les EAP-SCA de leurs contenus</b> pour assurer le référencement mais sans accès au « back-office » de la plateforme : <i>ClaimReview</i> (<i>Google, Bing, Facebook, YouTube</i>).</p>
<p>« <i>Fact-checking</i> » réalisé en interne</p>	<p>Vérification des informations par une équipe en interne, <b>en s'appuyant sur un guide pratique rédigé par la plateforme ou sur les informations produites par les EAP-SCA</b> faisant autorité sur leurs propres environnements (<i>LinkedIn</i>)<sup>51</sup>.</p>
<p>Actions sur les</p>	<p>Actions s'effectuant contenu par contenu, au cas par cas :</p>

<sup>50</sup> *Twitter* fait néanmoins état d'un partenariat avec l'association DisinfoLab UE.

<sup>51</sup> *Doctissimo* et *Aufeminin* ont prévu en 2020 de développer des activités de *fact-checking*.



<p>contenus de sources réputées peu fiables</p>	<ul style="list-style-type: none"><li>- <b>dégradation de la visibilité</b> des contenus jugés faux par des tiers vérificateurs (<i>Google, Bing, Facebook</i>) ;</li><li>- <b>informations apposées sur le contenu</b> vérifié à l'égard des utilisateurs (<i>Facebook, Instagram</i>).</li></ul> <p>Dans le cadre d'une vérification par des tiers, peu de plateformes (<i>Facebook</i><sup>52</sup>) peuvent procéder à une <b>suppression des contenus</b> dont les <i>fact-checkers</i> ont pu déterminer qu'ils contenaient de fausses informations.</p> <p>Aucune plateforme n'a déclaré avoir mis en œuvre des restrictions techniques des possibilités de partage des contenus ayant été jugés comme contenant une fausse information.</p>
---	---

### **Mesures spécifiques prises dans le cadre de la crise de la COVID-19**

#### **Promotion des contenus fiables :**

- *LinkedIn* a mis en place une promotion des contenus de sources fiables (sources officielles nationales et internationales) via son Récap Actu ;
- *Doctissimo* a lancé une newsletter quotidienne et un article dédié afin de combattre les fausses nouvelles et a développé un programme de vidéos expliquant les fausses informations avec le pictogramme « *fake news* ». Il s'est associé à *Facebook* afin de mettre en place un dispositif live en partenariat avec TF1 et LCI qui permettait à un médecin urgentiste de répondre aux questions des utilisateurs ;
- la plupart des opérateurs ont redirigé les utilisateurs vers les informations officielles, notamment dans les propositions et résultats de recherche (*Facebook, Google, YouTube, LinkedIn, Twitter, Microsoft*) et ont mis en place des chat-bots dédiés (*Facebook*) ;
- *Google* a mis en place, avec le Gouvernement, une campagne sur les gestes barrières ;
- *Facebook* a affiché des renvois vers des articles de l'OMS dans le fil d'actualité des personnes ayant interagi avec des informations nuisibles liées à la COVID-19.

**Aide financière au *fact-checking* :** subvention à des organismes comme l'IFCN (*Facebook*), le Fonds mondial d'aide d'urgence au journalisme, l'International Center for Journalists, le DART center for journalism and Trauma (*Google*).

<sup>52</sup> Dans des cas particuliers : suppression des fausses informations contribuant à causer un dommage physique ou à la pression sur les électeurs, ou lorsque les standards ou règles de la communauté sont violés par ailleurs.



### 3.3.2. Analyse des mesures prises par les opérateurs

- **Identification des sources de contenus issus d'EAP-SCA : différents degrés de précisions dans les informations portées à la connaissance de l'utilisateur**

Si la majorité des plateformes déclarent prendre des mesures, celles-ci sont mises en œuvre de différentes manières et avec différents degrés de précision. Cette hétérogénéité tient essentiellement au modèle de la plateforme et à la nature du service.

Le **dénominateur commun consiste à faire apparaître le nom et/ou le logo de la marque de presse publiant un contenu, ces éléments étant renseignés par l'organe de presse lui-même** depuis son compte utilisateur. À cet égard, si l'affichage simple d'informations renseignées par l'EAP-SCA (nom et logo) permet son identification, il ne semble pas suffire à répondre parfaitement à un objectif de promotion de ces contenus.

Sur certains moteurs de recherche (*Google* et *Bing*), l'identification des sources de contenus issus d'EAP-SCA via la fonctionnalité *ClaimReview* résulte également d'une action externe à la plateforme, c'est-à-dire exécutée par l'organe de presse lui-même, tout en délivrant des informations plus précises comme par exemple la participation à la vérification de l'information du contenu libellé.

Au-delà de la simple identification, certaines plateformes **informent l'utilisateur quant à la fiabilité d'une source par l'apposition d'un visuel** qui peut recouvrir plusieurs fonctions comme authentifier le compte utilisateur (*Twitter*) ou, de façon spécifique, certifier qu'un EAP-SCA est digne de confiance (ex : *Dailymotion*). Il s'agit ici d'une action de la plateforme elle-même, à l'aune de critères qu'elle préétablit.

Sur *Facebook*, le bouton de contexte à côté du contenu permet à l'utilisateur d'**obtenir des informations synthétiques sur l'organe de presse** (dont la ligne éditoriale : informations, satire...), un rapport des derniers articles publiés par cet organe ainsi qu'un renvoi vers sa page Wikipédia. Cette fonctionnalité est spécifiquement dédiée aux contenus d'éditeurs de sites internet, dont les éditeurs de presse. Sans certitude, il semble qu'elle résulte à la fois d'une action externe (les informations délivrées étant susceptibles de provenir des organes de presse eux-mêmes) et d'une action algorithmique. Les mesures de transparence prises sur *Instagram* méritent d'être relevées, qui consistent à attacher ces informations aux comptes ayant de nombreux abonnés (seuil non spécifié) : date de création du compte, pays du titulaire, publicités diffusées, anciens noms d'utilisateurs sur un an.

Peu d'éléments permettent de déterminer si les plateformes recherchent, préalablement à l'identification des contenus ou la certification de des comptes, si les EAP-SCA participent à des **démarches de labellisation**.

Les opérateurs ne mentionnent pas d'éventuelles mesures leur permettant de contrôler la véracité des informations liées à l'identité des organes de presse lorsqu'ils n'en sont pas eux-mêmes à l'origine. Seuls deux moteurs de recherche (*Google* et *Microsoft*) indiquent procéder à de telles vérifications, sans préciser avec quelle régularité et selon quels moyens.



- **Mise en avant des informations, notamment de vérification, provenant des EAP-SCA**

Si l'ensemble des opérateurs déclarent prendre des mesures visant à assurer à l'utilisateur l'accès à des sources réputées dignes de confiance, cela se traduit de manière différente d'une plateforme à l'autre. La nature de la plateforme ne semble pas expliquer à elle seule cette situation. On note néanmoins que l'intervention algorithmique joue ici un rôle prépondérant. Par ailleurs, on constate que les actions d'identification et de mise en avant ne sont pas isolées l'une de l'autre, la première permettant la mise en œuvre de la seconde, notamment grâce aux outils algorithmiques de référencement.

**Une mise en avant algorithmique des contenus provenant des EAP-SCA**

Une fois les contenus identifiés comme provenant d'organes de presse (ou de sources « fiables »), ils sont, **de manière algorithmique, sélectionnés et affichés au sein de canaux ou d'espaces dédiés** (*Dailymotion, Snapchat, LinkedIn, YouTube*) **ou remontés parmi les premiers résultats de recherche** (*Google, LinkedIn, Bing*). Il est à noter que, dans le cadre d'échanges avec le comité d'experts, certaines plateformes ont invoqué l'application de droits voisins comme contrainte financière à prendre en compte dans la mise en avant des contenus issus des EAP-SCA.

La **dégradation de la visibilité des contenus jugés comme véhiculant de fausses informations** est une autre manière algorithmique de mettre en avant les contenus jugés fiables.

Certaines plateformes **mettent spécifiquement en avant les actions de *fact-checking***, soit en réservant aux contenus de vérification une place privilégiée aux côtés d'une recherche associée à l'information vérifiée, soit en marquant d'un visuel le contenu ayant fait l'objet de la vérification et jugé comme pouvant véhiculer de fausses informations. L'utilisateur souhaitant partager ce contenu peut ainsi prendre connaissance du travail de vérification.

**La vérification de l'information : un recours encore peu fréquent aux tiers vérificateurs**

Certaines plateformes déclarent ne prendre aucune mesure en matière de vérification de l'information ou sont silencieuses sur ce point (*JeuxVideo.com, Twitter, Wikipédia, Yahoo, AuFeminin*). Pour certaines d'entre elles, la nature de leurs services tend à l'expliquer (*Jeuxvideo.com, AuFeminin* et *Wikipédia*). Néanmoins, pour d'autres, cette situation soulève plus d'interrogations et on peut regretter la modestie des informations déclarées au Conseil, notamment de la part de *Twitter*.

**Deux plateformes déclarent sélectionner « soigneusement » et selon leur « fiabilité », dans le cadre de partenariats, les EAP-SCA** dont les contenus sont disponibles au sein de tout ou partie de leur service (*Snapchat* (dans la partie Discover) et *Yahoo Portal*). Dans cette configuration, elles estiment moins nécessaire le recours à la vérification des informations publiées sur leurs services car émanant « nécessairement » de sources fiables ou tenues contractuellement de s'assurer elles-mêmes de la véracité des informations publiées. S'agissant de la partie Discover de *Snapchat*, on rappellera néanmoins qu'elle accueille aussi des contenus d'influenceurs (certes, également sélectionnés préalablement par *Snapchat*) et donne accès aux *Stories* publiées par les contacts de l'utilisateur. En outre, certains membres du comité d'experts ont fait valoir que cette situation posait la question de l'éventuelle



responsabilité éditoriale de la plateforme dans ce type de cas, dès lors qu'elle est susceptible de choisir d'exposer ses utilisateurs à tel ou tel type d'informations.

Peu de plateformes font **directement appel au travail de fact-checkers**. *Facebook* a mis en place un outil spécifique permettant à ses partenaires d'accéder aux contenus signalés comme véhiculant de fausses informations et d'injecter leurs propres contenus vérifiant les informations. Avec la fonctionnalité « *ClaimReview* », *Google* et *Microsoft* permettent aux organismes de presse de marquer leurs contenus participant à la vérification de l'information afin que ces derniers puissent jouir d'une remontée algorithmique dans les résultats de recherche et de figurer aux côtés des résultats liés à l'information recherchée.

Dans ce cadre, la prise en compte des différentes démarches de labellisation revêt une importance particulière. En effet, le recours au travail des tiers vérificateurs pose nécessairement la question de l'identité et de la fiabilité de ces tiers. A cet égard, le choix de sélectionner des organismes préalablement membres de structures telle que l'IFCN est à saluer.

Lorsque **la vérification de l'information est réalisée en interne** (*Dailymotion*, *LinkedIn* et *Doctissimo*), elle s'apparente d'avantage à un travail de modération traitée sous l'angle du respect des CGU propres à chaque plateforme. Les sources journalistiques externes ne sont toutefois pas écartées puisqu'elles peuvent servir de support aux équipes de modérateurs pour vérifier une information. Néanmoins, une vérification de l'information par la plateforme elle-même n'est pas sans soulever certaines questions quant à la position qu'elle occupe pour « dire le vrai » et au degré de son intervention éditoriale sur les contenus.

Le module *NewsGuard* apparaît comme une mesure hybride par laquelle la plateforme fait appel indirectement aux tiers vérificateurs. Ceux-ci participent à un travail global de certification de sites dont la fonctionnalité *NewsGuard* fait ensuite état en signalant grâce à un symbole (bouclier coloré) la fiabilité d'une source. Sans ouvrir directement leurs services aux tiers vérificateurs ou sans déployer d'espaces qui leur soient spécifiques, les plateformes proposant ce module permettent néanmoins de faire connaître leurs travaux aux utilisateurs.

### ***L'activité d'édition de contenus rédactionnels et de contractualisation avec des organes de presse comme source de contenus fiables***

*Doctissimo*, *LinkedIn* et *Yahoo Portal* publient au sein de leur service leurs **propres contenus rédactionnels, signalés comme tels**. Au vu des efforts décrits dans leur déclaration pour garantir la fiabilité, la traçabilité et la rigueur de l'information délivrée, cette pratique permet d'offrir aux utilisateurs des contenus de référence. Une mesure du même ordre consiste pour *Doctissimo*, *Snapchat* et *LinkedIn* à contractualiser avec des organes de presse.



### 3.3.3. Recommandations du CSA

**Au vu de ces éléments d'analyse, le CSA formule les recommandations suivantes.**

- ***En matière d'identification des contenus issus des EAP-SCA***

1. Le CSA relève l'existence de fonctionnalités, spécifiques ou non, permettant d'identifier les contenus issus des EAP-SCA. Elles se trouvent essentiellement sur les plateformes ayant une activité de réseau social. Celles-ci apparaissent de fait les plus concernées par l'objectif poursuivi par cette mesure, eu égard à la diversité des sources de contenus qu'elles proposent à leurs utilisateurs.

Il salue et encourage les pratiques consistant à non seulement **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.

2. Lorsque l'identification des ESP-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.

3. 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.

4. Le CSA **encourage les démarches de labellisation des EAP-SCA et estime utile que les utilisateurs des plateformes puissent en avoir connaissance.**

- ***En matière de fact-checking***

5. De manière générale, **le CSA encourage fortement le développement de partenariats entre les plateformes et des fact-checkers (tiers vérificateurs de l'information).**

6. 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).



7. 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.
8. 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.

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

De manière générale, **l'appréhension de la notion de « comptes propageant massivement de fausses informations » dans les déclarations est protéiforme** et englobe tant des pratiques abusives ou trompeuses permettant la viralisation des fausses informations que des pratiques non assimilables à des comportements participant de la propagation massive de fausses informations. Cette appréhension conditionne les réponses des opérateurs s'agissant des mesures mises en œuvre en la matière.

Les opérateurs adoptent une approche globale de la lutte contre les pratiques de propagation massive de fausses informations sur leur plateforme, qui peuvent inclure des mesures et mécanismes de détection et de traitement des comptes propageurs. Ainsi, peuvent être observés des procédés de lutte « collatérale » contre ces comptes. A ce titre, les comptes propageant massivement de fausses informations ne bénéficient ni d'une modération ni d'une catégorisation spécifique et sont **appréhendés sous l'angle des pratiques interdites par les règles internes des plateformes**. Il pourrait sembler intéressant à l'avenir d'accompagner l'ensemble des opérateurs dans la mise en place de mesures spécifiques contre les comptes propageant massivement de fausses informations.

Par ailleurs, certains opérateurs se considèrent comme peu concernés par la lutte contre les comptes propageant massivement de fausses informations, du fait de l'absence de « fonctionnalités sociales et de promotion de contenus » (*Dailymotion*) ou d'une structure de leur service qui empêche la viralisation des contenus (*Snapchat*<sup>53</sup>, *Jeuxvideo.com*, *Bing*).

#### 3.4.1. Typologie des mesures prises par les opérateurs

##### DÉFINITION DES COMPTES PROPAGEANT MASSIVEMENT DE FAUSSES INFORMATIONS

**Aucune plateforme n'utilise, dans son travail de modération, de catégorie « comptes propageant massivement de fausses informations » ; les opérateurs utilisent les catégories suivantes, auxquelles ils se sont référés dans leur déclaration.**

Comptes	Catégorie des « <b>faux comptes</b> » ( <i>Facebook</i> et <i>LinkedIn</i> ) : comptes <b>dont le but</b>
---------	---

<sup>53</sup> *Snapchat* invoque l'absence de profils publics d'utilisateurs et d'outils de « vanité » (Likes, partages).



désignés comme « faux comptes »	<b>est de créer et/ou partager des informations trompeuses</b> (notamment dans le cadre de « <i>campagnes d'informations coordonnées, comme celles qui sont souvent utilisées pour influencer des élections</i> », comme rappelé par <i>Facebook</i> ).
Comptes contrevenant aux règles de la plateforme s'agissant de pratiques trompeuses	<b>Comptes dont les comportements contreviennent aux règles internes de la plateforme</b> et participent de la propagation de fausses informations, sous l'angle d'une <b>pratique particulière et définie par l'opérateur</b> : fausses déclarations, usurpations d'identité, spams, pratiques trompeuses, vidéos truquées et trompeuses, vandalisme <sup>54</sup> ( <i>Google, Facebook, Snapchat, Twitter, Yahoo, Dailymotion, Wikipédia, Unify</i> ).
Comptes pratiquant le <i>flood</i>	Comptes pratiquant le <b>flood</b> (c'est-à-dire la multipublication de contenus) ( <i>Jeuxvideo.com</i> ).
Manipulation par des bots	Pratique de <b>manipulation par des bots</b> <sup>55</sup> qui cherchent à influencer le système de classement d'un moteur de recherche par exemple, notamment en se livrant à une « fraude au clic » ou en exploitant des « fermes de liens » pour faire croire qu'une site est plus populaire qu'il ne l'est en réalité ( <i>Microsoft</i> ).

#### DÉTECTION DES COMPTES PROPAGEANT MASSIVEMENT DE FAUSSES INFORMATIONS

Moyens d'identification des comptes propageant massivement de fausses informations	<p><b>Moyens :</b></p> <ul style="list-style-type: none"> <li>- <b>exclusivement non automatiques</b> se fondant uniquement sur les signalements (<i>Dailymotion ; Jeuxvideo.com</i> pour la pratique de <i>flood</i>) ;</li> <li>- <b>en partie automatiques et en partie humains</b> (<i>Facebook, Google, LinkedIn, Microsoft, Yahoo Twitter</i><sup>56</sup>, <i>Wikipédia, Unify</i>) : certains opérateurs expliquent que la détection des faux comptes et des comportements de spams à grande échelle nécessite de déployer des moyens technologiques complétés par des moyens humains.</li> </ul>
--	---

#### TRAITEMENT DES COMPTES PROPAGEANT MASSIVEMENT DE FAUSSES INFORMATIONS

Mise en œuvre de la décision : humaine ou automatisée	<p>La décision de clore un compte revient à une <b>équipe humaine</b> de modération (<i>Jeuxvideo.com, Snapchat, LinkedIn, Dailymotion</i>) ;</p> <p>La plateforme prévoit un <b>procédé de lutte hybride utilisant l'intelligence humaine et l'automatisation</b> (<i>Google, Wikipédia, Facebook, Unify</i>) :</p> <ul style="list-style-type: none"> <li>- <i>Facebook</i> : détection et suppression d'un « faux compte » fondées sur <b>le signalement d'utilisateurs et les systèmes automatiques</b> de détection (parfois en amont pour bloquer toute tentative) ;</li> <li>- <i>Wikipédia</i> : possibilité : <ul style="list-style-type: none"> <li>- pour les administrateurs, de <b>bloquer en écriture</b> les comptes</li> </ul> </li> </ul>
---	--

<sup>54</sup> Cf. note 38.

<sup>55</sup> Désignant ici un logiciel informatique agissant automatiquement pour reproduire, parfois massivement, un comportement humain.

<sup>56</sup> Pour détecter les comptes automatisés, *Twitter* utilise un dispositif de reCAPTCHA (demande à l'utilisateur de reconnaître un mot ou une image affichée à l'écran) et des demandes de réinitialisation du mot de passe.



	<p>d'utilisateurs ne respectant pas les principes fondateurs et valeurs du service (notamment en diffusant de fausses informations) ;</p> <ul style="list-style-type: none"> <li>- pour les administrateurs et contributeurs, de <b>supprimer des contenus</b> diffusant de fausses informations ;</li> <li>- pour les programmes informatiques développés par les contributeurs (tels que « ClueBotNG »), d'<b>annuler rapidement les contributions</b> qu'il classe dans la catégorie vandalisme<sup>57</sup> ;</li> <li>- <i>Unify</i> : outil de modération automatique pour supprimer des messages ou bannir des utilisateurs <b>en cas d'utilisation de mots précis</b>, en parallèle du traitement humain des contenus signalés.</li> </ul>
<p>Critères de détection et traitement des comptes propageant massivement des fausses informations</p>	<p><b>Violation par le compte des CGU et règles internes</b> (ex. : <i>Facebook, Yahoo</i>), notamment détecté sur <b>signalement</b> (ex. : <i>Facebook</i>).</p> <p>Comportement de l'utilisateur :</p> <ul style="list-style-type: none"> <li>- <b>création abusive de comptes par le même utilisateur</b> (<i>Google, LinkedIn</i>) ;</li> <li>- <b>republification fréquente d'un message</b> par un utilisateur (<i>Unify</i>) ;</li> <li>- <b>activité anormalement élevée avec le même hashtag</b> (<i>Twitter</i>) ;</li> <li>- <b>comportements anormaux</b> (<i>Snapchat</i> : publication massive de <i>Stories</i> publiques) ;</li> <li>- <b>publication de plusieurs contenus signalés puis retirés par la plateforme</b> (<i>Jeuxvideo.com, Snapchat</i>).</li> </ul> <p><i>LinkedIn</i> :</p> <ul style="list-style-type: none"> <li>- <b>système de filtrage au moment de la création du contenu</b> pour s'assurer du respect des politiques du service : <i>Universal Content Filtering</i> qui effectue une série de contrôles sur tous les contenus notamment via des modèles de <i>machine learning</i><sup>58</sup>, des « classificateurs » et du <i>hash matching</i><sup>59</sup> ;</li> <li>- <b>analyse automatisée successive au regard de plusieurs critères</b> : score d'inscription, partitionnement<sup>60</sup>, score basé sur l'activité, score de signalements comme spams puis revue manuelle (« entonnoir de défense »).</li> </ul> <p>Pas de réponse (<i>Unify</i> et <i>Wikipédia</i>).</p>
<p>Mesures pouvant être</p>	<p>Mesures sur le contenu :</p> <ul style="list-style-type: none"> <li>- <b>modification et protection en écriture du contenu</b> (<i>Wikipédia</i><sup>61</sup>) ;</li> </ul>

<sup>57</sup> Cf. note 38.

<sup>58</sup> Apprentissage automatique.

<sup>59</sup> Recherche de correspondance d'empreintes.

<sup>60</sup> D'après la déclaration de *LinkedIn*, technique de *machine learning* utilisée « en aval [de l'inscription] pour capturer de plus petits groupes de faux comptes » consistant à « [créer] des groupes de comptes en les regroupant en fonction de caractéristiques partagées » puis à « [trouver] des groupes de comptes qui montrent une distribution statistiquement anormale des données, ce qui indique qu'ils sont créés ou contrôlés par un seul acteur nuisible (...). Les modèles notent les groupes puis propagent la note du groupe aux comptes individuels ». Cette approche par groupe et non par membre permet à l'opérateur « d'attraper plus de faux comptes [plus rapidement] » que s'il choisissait d'attendre que ceux-ci « commencent à entreprendre des actions abusives sur le site ».



<p>prises en cas de propagation massives de fausses informations</p>	<p>- <b>suppression</b> du contenu (<i>Snapchat, Dailymotion, Facebook, Wikipédia, Unify</i>).</p> <p>Mesures sur le compte :</p> <ul style="list-style-type: none"> <li>- <b>blocage</b> du compte (<i>Google, Twitter, Facebook, Yahoo, Wikipédia</i>) ;</li> <li>- <b>suppression</b> du compte (<i>Google, Snapchat, Twitter, LinkedIn, Dailymotion, Facebook</i>) ;</li> <li>- <b>exclusion du service</b> (<i>Jeuxvideo.com, Unify</i>).</li> </ul>
<p>Mesures économiques (restrictions ou incitations)</p>	<p><i>Facebook</i> : <b>suspension de la possibilité de monétiser</b> les contenus pour les personnes à l'origine des comptes.</p> <p><i>Google</i> : <b>monétisation non ouverte à tous</b> ; règles et procédures visant à perturber les tentatives de monétisation émanant d'acteurs malveillants et trompeurs (spammeurs, usurpateurs d'identité, etc.)<sup>62</sup>.</p>
<p>Mesures particulières en direction des comptes relayant des opérations d'information soutenues par des entités étatiques</p>	<p><i>Google</i> : mesures de <b>lutte contre les tentatives politiques pour influencer sur les élections</b>.</p> <p>Ex. : identification des campagnes d'influence politique liées aux gouvernements russe et iranien puis suppression de dizaines de chaînes YouTube et blogs liés aux organismes à l'origine de ces opérations.</p> <p><i>Twitter</i> : publication des <b>archives complètes des opérations d'information soutenues par des entités étatiques</b>.</p> <p>Ces archives ont été enrichies en 2019 par de nouveaux corpus de données, afin de permettre l'analyse de ces comportements par des tiers<sup>63</sup>.</p> <p><i>Facebook</i> : contrôle accru des comptes de médias contrôlés par un État.</p>

#### INFORMATION DES UTILISATEURS

<p>Information aux utilisateurs sur les mesures de détection et de traitement de tels comptes</p>	<p>Information présente :</p> <ul style="list-style-type: none"> <li>- dans les <b>centres d'aide, charte d'utilisation ou Standards de la communauté</b> (<i>Dailymotion, Yahoo, Facebook, Google, Twitter, Wikipédia, Jeuxvideo.com</i>) ;</li> <li>- dans les <b>rapports de transparence et les archives</b> (<i>Twitter, LinkedIn</i>).</li> </ul> <p>Pas de réponse (<i>Snapchat, Unify</i>).</p>
---	---

<sup>61</sup> Wikipédia permet une protection totale ou une semi-protection de pages, signalée par l'apparition d'une icône « cadenas » en haut à gauche de la page.

<sup>62</sup> A ce titre, voir la déclaration de l'opérateur renvoyant vers :

- Les **règles Google Ads et AdSense** relatives au contenu trompeur, au contenu original insuffisant, à l'inventaire de valeur et au contenu inapproprié concernant *Google Search* ;
- Les **critères d'éligibilité au Programme Partenaire YouTube** dont les membres peuvent monétiser leurs contenus : pour candidater au programme, une chaîne *YouTube* doit avoir généré 4000 heures de visionnage au cours des 12 derniers mois et compter plus de 1000 abonnés.

<sup>63</sup> *Twitter* fournit, dans sa déclaration, une liste des opérations d'interférence étrangère potentielles observées sur son service en 2019 qui comprend les États et les contenus concernés et les mesures prises.



<p>Information aux utilisateurs sur les risques liés à la pratique de création de tels comptes</p>	<p><b>Accessibilité et niveau de détail :</b></p> <ul style="list-style-type: none"> <li>- dans les <b>CGU et règles internes</b> de la plateforme (<i>Dailymotion, Facebook, Google, LinkedIn, Snapchat, Twitter, Wikipédia, Jeuxvideo.com, Verizon</i>) ;</li> <li>- dans un langage simple et clair avec des illustrations des comportements interdits (<i>Twitter</i>) ;</li> <li>- délibérément décrites <b>à un niveau de détail moindre</b> pour les utilisateurs, afin de ne pas donner les clés de contournement aux acteurs nuisibles (<i>LinkedIn</i>) ;</li> <li>- annoncées comme <b>prochainement prévues</b> dans les CGU et règles internes (<i>Unify</i>).</li> </ul> <p><b>Les informations délivrées aux utilisateurs concernent :</b></p> <ul style="list-style-type: none"> <li>- les <b>interdictions</b> (ex. : <i>Wikipédia</i> : création de plus d'un compte, vandalisme<sup>64</sup>) ;</li> <li>- les <b>risques encourus</b> en cas de création de comptes propageant massivement de fausses informations (ou catégorie similaire) (<i>Dailymotion, Facebook, Google, Snapchat, Jeuxvideo.com, Yahoo, LinkedIn</i>) ;</li> <li>- les <b>modalités de clôture</b> d'un compte ayant une activité illicite (<i>Dailymotion</i>).</li> </ul>
--	--

DONNÉES CHIFFRÉES	
<p>Moyens financiers et humains alloués à la lutte contre les comptes propageant massivement de fausses informations</p>	<p><b>Moyens humains :</b></p> <ul style="list-style-type: none"> <li>- <b>informations globales</b> sur les moyens humains alloués au traitement de l'ensemble des dispositifs de modération développés pour le service (<i>Facebook, Google, Yahoo, LinkedIn, Snapchat, Jeuxvideo.com</i>) (voir partie 3.1) ;</li> <li>- <i>Wikipédia</i> : la version française de son encyclopédie compte près de 5 000 contributeurs « actifs »<sup>65</sup> et plus de 20 000 contributeurs ayant réalisé au moins une contribution dans les trente derniers jours ;</li> <li>- <b>type de services sollicités</b> pour le traitement des signalements (<i>Microsoft</i>) : premier niveau d'analyse par une équipe de 50 personnes ; second niveau (si nécessaire) par ses services techniques et juridiques spécialisés.</li> </ul> <p><b>Moyens financiers</b> alloués (<i>Snapchat</i>).</p> <p>Pas d'information (<i>Twitter, Unify</i>, ce dernier précisant n'avoir détecté aucune activité sur ses forums provenant de tels comptes).</p>
<p>Nombre de faux comptes détectés et traités au niveau mondial (sauf précision)</p>	<p><i>Facebook</i> : les faux comptes (hors bloqués en amont) représentent environ 5 % des utilisateurs actifs mensuels en 2019 ; 5,39 milliards désactivés entre le 1<sup>er</sup> et le 3<sup>e</sup> trimestres 2019.</p> <p><i>Google</i> : suspension d'1,5 million<sup>66</sup> de comptes d'annonceurs pour violation des règles publicitaires et retrait de 2,7 milliards de publicités contrevenant à ces</p>

<sup>64</sup> Cf. note 38.

<sup>65</sup> C'est-à-dire ayant réalisé au moins cinq contributions dans les trente derniers jours.

<sup>66</sup> 500 000 de plus qu'en 2018.



	<p>règles publicitaires ; au niveau national, 19 942 comptes identifiés comme en violation de la règle Google Ads relative aux déclarations trompeuses.</p> <p><i>YouTube</i> : au 4<sup>e</sup> trimestre 2019, suppression de 5,8 millions de vidéos (dont 31 031 en France), 2 millions de chaînes et 540 millions de commentaires.</p> <p><i>LinkedIn</i> : suppression de 104 600 contenus « spam et arnaque » et blocage ou suppression de 20 millions de faux comptes, au 1<sup>er</sup> semestre 2019 ;          - en France : 1 427 membres français signalés pour publication de 1 805 contenus faux ou trompeurs. Sur la base d'un examen manuel, ces comptes n'ont pas été sanctionnés au seul titre de leur activité de publication de fausses informations, et certains ont été restreints pour d'autres raisons (ex. : publication de discours haineux).</p> <p><i>Twitter</i> : au 1<sup>er</sup> semestre 2019, 16 millions de vérification anti-spam par mois ; du 1<sup>er</sup> au 3<sup>e</sup> trimestre 2019, suspension de 144 000 applications dans le cadre de sa lutte contre l'automatisation malveillante et les abus provenant de son API ; au 1<sup>er</sup> semestre 2019, 4 544 096 signalements de spam.</p> <p><i>Wikipédia</i> : sa partie « caractérisation automatisée » (programme ClueBot NG) a classé correctement plus de 90 % des contributions analysés soit dans la catégorie vandalisme, soit dans la catégorie non-vandalisme<sup>67</sup>.</p> <p><i>Unify</i> : aucune activité détectée en 2019.</p>
<p>Publicité de ces chiffres</p>	<p><b>Rapports de transparence biannuels</b> (<i>Twitter</i>, <i>LinkedIn</i><sup>68</sup>, <i>Facebook</i><sup>69</sup>, <i>Google</i>).</p> <p>Pas de réponse (<i>Snapchat</i>, <i>Verizon</i>, <i>Jeuxvideo.com</i>).</p>
<p>Recettes générées pour la plateforme par les comptes propageant massivement de fausses informations et celles qu'elle leur aurait reversées</p>	<p>Aucune recette : <i>Facebook</i> (suppression très rapide de ce type de comptes ; pas de rémunération versées aux comptes de manière générale) ; <i>Google</i> (du fait de ses règles de monétisation, cf. <i>supra</i>) ; <i>LinkedIn</i> (les influenceurs et autres utilisateurs n'étant pas rémunérés ; mais il ne peut être tout à fait exclu qu'un « faux » compte ait acheté un abonnement premium ou un autre produit du service).</p> <p>Absence de données :</p> <ul style="list-style-type: none"> <li>- pas de réponse (<i>Twitter</i>, <i>YouTube</i>, <i>Jeuxvideo.com</i>).</li> <li>- pas de compte propageant massivement de fausses informations détecté sur la plateforme (<i>Snapchat</i>, <i>Unify</i>) ;</li> <li>- pas concerné (<i>Wikipédia</i>).</li> </ul>

<sup>67</sup> Cf. note 38.

<sup>68</sup> « Community Report ».

<sup>69</sup> « Rapport d'Application de Standards de la communauté ».



### 3.4.2. Analyse des mesures prises par les opérateurs

- ***Sur la notion de « compte propageant massivement de fausses informations »***

Les opérateurs de plateformes n'utilisent pas à proprement parler la notion de « compte propageant massivement de fausses informations ».

Leur action en direction de ces comptes s'inscrit en effet dans une politique plus large visant les **comptes relevant de pratiques considérées comme abusives**. Ainsi, c'est au titre de la lutte contre ces pratiques que les opérateurs mettent en œuvre **de manière « collatérale »** des mesures permettant de lutter contre les comptes propageant massivement de fausses informations.

De manière générale, les mécanismes de traitement des comptes propageant massivement de fausses informations prévus par les plateformes ne sont donc pas propres à ces derniers, mais touchent également d'autres types de contenus (spams, incitation à la haine, contenus violents ou pornographiques). Le fait que les opérateurs prévoient des **dispositifs communs à plusieurs catégories ou motifs de signalement** peut se comprendre. Elle conduit cependant à des politiques de modération indifférenciées, ce qui peut poser question.

Il convient en outre de relever que l'objectif poursuivi par la loi de lutter contre les comptes propageant massivement de fausses informations ne concerne pas uniquement les comptes considérés comme « faux », mais également les **« vrais » comptes adoptant des comportements trompeurs**. Ce peut être le cas, pour prendre quelques exemples parmi de nombreux cas possibles, d'une vidéo truquée ou biaisée visant à polariser un débat ou à créer un climat général de polarisation, de l'utilisation de métadonnées trompeuses ne correspondant pas au contenu pour biaiser les algorithmes de recommandation et augmenter la visibilité de son contenu, de la création de comptes (exemple : chaînes YouTube, blogs, comptes Twitter) à l'origine d'opérations d'interférence étrangère soutenues par des entités étatiques.

Le **CSA salue l'esprit de coopération dont ont fait preuve les opérateurs en lui fournissant des éléments relatifs aux catégories et problématiques qui leur sont spécifiques** mais dont ils ont bien identifié qu'elles relevaient de cet axe général de lutte contre la propagation de fausses informations.

- ***Sur la fourniture de données statistiques liées à la lutte contre les comptes propageant massivement de fausses informations***

**De manière générale**, les opérateurs ne sont pas en mesure de fournir des données spécifiques à la lutte contre les comptes propageant massivement de fausses informations, de la même façon qu'ils éprouvent des difficultés à identifier les mesures spécifiquement destinées à lutter contre ce type de comptes. Cette situation découle probablement du fait que les opérateurs assimilent les comptes propageant massivement de fausses informations à des **pratiques trompeuses et abusives** (généralement désignés – tel que vu



précédemment – par les termes « faux comptes », « spams », « comptes abusifs », « usurpations d'identité », « comportements malveillants », etc.).

Les chiffres fournis sont nombreux, intéressants et démontrent les efforts réalisés par les opérateurs. Toutefois, **le suivi statistique d'ensemble du phénomène et de sa modération en est rendu très complexe**, *a fortiori* avec des données ne portant pas sur les mêmes périodes (expression en année, en semestre, en trimestre...) ou zones géographiques (France ou Monde).

**Très peu d'informations ont été communiquées au Conseil sur les éventuelles recettes financières générées** pour la plateforme par ces comptes. *Facebook* indique que ces recettes sont inexistantes dans la mesure où ils sont supprimés rapidement<sup>70</sup> voire immédiatement après leur création. En outre, la plateforme ne verse pas de rémunération aux comptes de manière générale. Elle indique néanmoins que ce type de comptes est en grande majorité utilisé dans des campagnes de spams et créé pour des raisons financières, notamment la petite minorité de ceux n'ayant pas été détectés à la création.

- ***Sur l'absence de modération spécifique des comptes propageant massivement de fausses informations***

Certains opérateurs indiquent le nombre de personnes composant leurs équipes de modération mais n'en précisent pas la **composition**. La majorité d'entre eux adoptent une **approche globale de la modération des contenus** sur leur plateforme qui rend difficile la caractérisation des comptes propageant massivement de fausses informations et des mesures prises pour assurer leur modération.

- ***Sur les garanties contre la viralisation des contenus relayant massivement de fausses informations liées au modèle et aux fonctionnalités de la plateforme***

*Snapchat* fait valoir que son modèle empêche la viralité des contenus et, partant, œuvre en soi à réduire les risques de propagation de fausses informations. En effet, sur le service, les contenus des utilisateurs privés sont éphémères (notamment, les *Stories* sont disponibles 24 heures) ; par défaut, il n'est pas possible de partager un contenu d'un contact avec un utilisateur non connecté à ce contact et très peu de contenus sont ouverts à un cercle plus large que les contacts de l'utilisateur les publiant. En outre, aucun contenu disponible sur le service n'est assorti de fonctionnalités d'engagement de type « *like* » ni ne peut être commenté. Enfin, les messages privés ne peuvent être envoyés qu'à un nombre restreint d'utilisateurs (les « groupes » ne peuvent dépasser plus d'une trentaine de membres).

Toutefois, les *Stories* d'un utilisateur sont, par défaut, visibles (de façon éphémère) par tous ses contacts dont le nombre est certes limité mais peut être « très élevé »<sup>71</sup>. Elles sont également partageables avec d'autres utilisateurs si l'utilisateur en fait le choix. Si la viralité d'un contenu contenant une fausse information est indéniablement bien moindre que sur d'autres plateformes, il n'est pas exclu que ce contenu puisse être vu par un nombre important d'utilisateurs.

<sup>70</sup> 99,8 % des faux comptes sont détectés et supprimés avant tout signalement.

<sup>71</sup> <https://support.snapchat.com/fr-FR/a/add-friends>.



- **Sur l'application des mesures de lutte contre les comptes propageant massivement de fausses informations**

Certaines réponses sont moins étayées que d'autres, ce qui peut s'expliquer par la différence de taille et de moyens des services considérés, et par le fait que certains se considèrent peu concernés par la problématique.

Il peut être regretté que certains opérateurs, dans leur déclaration, renvoient systématiquement vers leurs CGU et règles internes au lieu d'en expliciter les mesures pertinentes en répondant précisément aux questions les concernant.

Le **relatif silence des opérateurs sur la question de l'information délivrée aux utilisateurs** peut s'expliquer par le caractère hétéroclite de cette catégorie et par le fait qu'ils privilégient l'information liée à un contenu jugé trompeur à celle relative à un compte jugé trompeur.

### 3.4.3. Recommandations du CSA

En matière de lutte contre les comptes propageant massivement de fausses informations, le CSA formule les recommandations suivantes.

1. 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.
2. 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.
3. 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.



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

La notion de contenus sponsorisés est ici entendue comme la publicité au sens large, englobant les publicités « classiques » et la promotion des contenus d'information se rattachant à un débat d'intérêt général<sup>72</sup>.

A l'exception de *Wikipédia*, **le modèle économique des plateformes repose sur la publicité**. Pour autant, les mesures mises en place par ces acteurs en ce domaine sont très hétérogènes.

Les contenus sponsorisés mettent en relation les plateformes, les utilisateurs et les annonceurs. Afin de lutter efficacement contre les fausses informations en ligne dans le respect de la liberté d'expression, les opérateurs doivent assurer un haut niveau de transparence à l'égard des citoyens, mettre en œuvre des mesures favorisant la sécurité des marques et restreindre la monétisation des pages diffusant massivement des fausses informations. Ces exigences valent pour tous les contenus sponsorisés, mais revêtent une importance particulière dans le cadre de la promotion des contenus d'information se rattachant à un débat d'intérêt général, ces derniers pouvant avoir un impact direct sur des enjeux démocratiques.

#### 3.5.1. Typologie des mesures prises par les opérateurs

RESTRICTIONS DE LA PUBLICATION DE CONTENUS SPONSORISÉS SUR LA PLATEFORME	
Absence de publicités	<i>Wikipédia, forum « blabla » de Jeuxvideo.com</i> <sup>73</sup> (non traités ci-après).
Refus des contenus sponsorisés d'information se rattachant à un débat d'intérêt général	<b>De manière permanente</b> ( <i>Dailymotion, Jeuxvideo.com, LinkedIn, Unify, Twitter</i> ) ; - <i>Microsoft</i> interdit les publicités pour les contenus liés aux élections, aux « <i>candidats politiques</i> », aux partis, aux mesures de vote et aux collectes de fonds politiques à l'échelle mondiale, ainsi que les publicités traitant de certaines « thématiques » (« <i>contenu politique ou religieux extrême</i> », terrorisme...).  <b>De manière temporaire</b> , principalement pendant les périodes électorales ( <i>Google</i> <sup>74</sup> , <i>Snapchat, Yahoo, Facebook</i> ).

<sup>72</sup> Le Conseil d'État, dans son avis du 19 avril 2018, s'est référé à la jurisprudence de la Cour européenne des droits de l'homme qui a estimé que relevait de l'intérêt général l'ensemble des questions qui touchent le public dans une mesure telle qu'il peut légitimement s'y intéresser, qui éveillent son attention ou le préoccupent sensiblement, notamment parce qu'elles concernent le bien-être des citoyens ou la vie de la collectivité.

<sup>73</sup> En outre, les utilisateurs ne sont pas autorisés à publier des publicités sur les forums.

<sup>74</sup> Sauf pour les « *annonces d'information sur les élections diffusées par les organes officiels communication du Gouvernement* ».



**TRANSPARENCE ET INFORMATION GÉNÉRALE DES UTILISATEURS**

<p>Identification des contenus sponsorisés</p>	<p><b>Labellisation :</b></p> <ul style="list-style-type: none"> <li>- <b>des contenus sponsorisés en général (toutes les plateformes).</b> L'étiquette apparaissant alors sur le contenu varie : « sponsorisé » (<i>Facebook</i>) ; « promu », « publicité » ou « sponsorisé » (<i>LinkedIn, Snapchat, Twitter, Microsoft</i>) ; « annonce » ou « annonce sponsorisée » (<i>Bing, Google</i>) ;</li> <li>- <b>particulière pour les contenus sponsorisés d'information se rattachant à un débat d'intérêt général :</b> mention « payé par » (<i>Facebook</i><sup>75</sup>) ou « financé par » (<i>Snapchat</i>) ; informations sur l'annonceur telles que le nom de l'organisation, le candidat au scrutin, le nom de la personne qui a financé la publicité... (<i>Facebook, Snapchat</i>).</li> </ul>
<p>Information des utilisateurs sur la raison pour laquelle ils voient une publicité</p>	<p><b>Informations aux utilisateurs sur le ciblage des publicités</b> grâce à un outil « pourquoi je vois cette publicité » (ou équivalent), accessibles en quelques clics depuis le contenu (<i>Facebook, Google, Snapchat, Twitter, Microsoft</i> dans certains cas<sup>76</sup>). Le niveau de détail des informations semble varier d'une plateforme à l'autre.</p> <p><b>Pas d'information</b> sur le ciblage (<i>LinkedIn</i>).</p>
<p>Politiques publicitaires</p>	<p>L'ensemble des plateformes disposent de <b>politiques publicitaires publiques, adaptées aux lois en vigueur dans le pays où la plateforme est proposée et accessibles aux utilisateurs.</b> Elles doivent être respectées par les annonceurs.</p> <ul style="list-style-type: none"> <li>- politiques <b>en français, sauf pour Yahoo et LinkedIn. LinkedIn est un cas particulier</b> car il y a deux documents : les politiques publicitaires de la plateforme (en anglais) et les « règles relatives à la publicité » (en français).</li> </ul> <p>Ces <b>politiques publicitaires peuvent interdire les contenus trompeurs ou inappropriés :</b></p> <ul style="list-style-type: none"> <li>- <i>Google</i> interdit les publicités qui « ont pour but de tromper les utilisateurs en omettant ou en masquant des informations pertinentes ou en fournissant des informations trompeuses » (règles Google Ads) et les contenus trompeurs : « les annonces Google ne peuvent pas être placées sur des pages qui dénaturent, déforment ou dissimulent des informations vous concernant ou concernant votre contenu ou l'objectif principal de votre site Web » (<b>règles AdSense</b>) ;</li> <li>- <i>Twitter</i> interdit les publicités ayant un contenu trompeur ou manipulé et les sujets ou événements sensibles ;</li> <li>- <i>LinkedIn</i> interdit (en anglais) les publicités frauduleuses et indique (en français) que les publicités ne peuvent « tromper ou mentir » aux membres ;</li> <li>- <i>Snapchat</i> interdit les publicités qui sont « fausses ou trompeuses » ;</li> </ul>

<sup>75</sup> Sur *Facebook*, les utilisateurs peuvent signaler les publicités pour lesquelles la mention n'apparaît pas.

<sup>76</sup> Une icône « information » ou une flèche permet d'afficher les paramètres de la publicité qui « peut » être associée à un message spécifique avec des informations de ciblage.



	- <i>Facebook</i> interdit les publicités mensongères.
Mesures liées à la responsabilisation des utilisateurs	<b>Outils permettant à l'utilisateur :</b> - <b>de comprendre le fonctionnement des publicités et de paramétrer ses préférences publicitaires</b> ( <i>Facebook, Google, LinkedIn, Snapchat, Twitter</i> ) ; - <b>de « masquer »</b> une publicité ( <i>Facebook</i> ).

### SIGNALEMENT DES CONTENUS

Possibilité de signaler les contenus sponsorisés	Processus de <b>signalement spécifique pour « fausse information »</b> ( <i>Facebook, Microsoft</i> ).  Pas de processus spécifique ( <i>autres opérateurs</i> ).
Suivi de l'état du signalement par l'utilisateur	<b>Peu d'information sur ce qu'il advient des contenus signalés et non encore traités</b> sauf : - <i>LinkedIn</i> : une publicité est mise en quarantaine après plus d'un signalement afin que les équipe de modération l'étudient ; - suivi du signalement : possible sur <i>Facebook</i> (pas sur <i>LinkedIn</i> ).  Autres opérateurs : pas d'information sur ce point.
Traitement (automatisé ou humain)	Seule une partie des répondants ont précisé ce point : - <i>LinkedIn</i> : les publicités ayant reçu plus d'un signalement sont suspendues puis <b>examinées manuellement</b> ; - <i>Google</i> et <i>Twitter</i> vérifient le respect des règles à <b>la fois par des algorithmes et des moyens humains</b> (répartition des tâches entre l'humain et la machine non décrite) ; - <i>Microsoft</i> : neutralisation ou retrait des publicités qui violent les politiques par des méthodes <b>d'exécution automatisées et manuelles</b> .

### MESURES LIÉES A LA SÉCURITÉ DES MARQUES

Mesures de sécurité des marques	<b>Mise en place de mesures de <i>brand safety</i></b> , notamment pour qu'une marque ne soit pas associée à un compte propageant des fausses informations, tels que : - l'examen des signaux contextuels des produits publicitaires, - l'interdiction des publicités superposées ou en pop-up» ( <i>Google</i> <sup>77</sup> , <i>Snapchat, Twitter, Dailymotion</i> ).
Mesures permettant aux contenus	<b>Peu de plateformes ont fait état de ces mesures</b> et certaines se sont étonnées du champ très large de la question par rapport à la problématique de lutte contre les fausses informations.

<sup>77</sup> Mise en place d'« outils de contrôle supplémentaires » à destination des annonceurs leur permettant d'« exclure des contenus ayant notamment trait à la politique, à l'actualité, au sport, à la beauté, à la mode ».



<p>sponsorisés de gagner en visibilité</p>	<p>Campagnes de publicités spécialisées ; outils de prospection commerciale (<i>Snapchat, LinkedIn</i>).</p>
<p>Mesures liées à la démonétisation des contenus propageant des fausses informations</p>	<p><b>Facebook suspend les capacités de monétisation</b> des « pages et domaines qui partagent » des fausses informations. Cette suspension peut être levée une fois que les pages « ne sont plus récidivistes » (durée de « probation » non précisée).</p> <p>Sur <i>YouTube</i>, les utilisateurs voulant monétiser leur chaîne doivent <b>déposer une candidature et répondre à certains critères</b> et, même après avoir été acceptés, il reste possible que leurs vidéos soient démonétisées.</p>

### SÉLECTION DES ANNONCEURS

<p>Approbation des annonceurs et des publicités</p>	<p>Certaines plateformes précisent les conditions pour pouvoir sponsoriser un contenu (<i>on parle ici des publicités en général et non de la promotion de contenus d'information se rattachant à un débat d'intérêt général</i>). Les annonceurs doivent toujours <b>respecter les politiques et règles de la plateforme</b>.</p> <p>Mesures d'approbation :</p> <ul style="list-style-type: none"> <li>- validation des annonceurs qui souhaitent diffuser des annonces électorales dans l'Union européenne (<i>Google</i>) ;</li> <li>- blocage de certains annonceurs (<i>Dailymotion, Twitter</i>) ;</li> <li>- <b>restriction de l'accès à l'outil pour mener une campagne</b> aux membres ayant accepté les CGU dont celles relative aux publicités (<i>LinkedIn</i>).</li> </ul>
---	--

### PROMOTION DES CONTENUS D'INFORMATION SE RATTACHANT À UN DÉBAT D'INTÉRÊT GÉNÉRAL

<p>Définition</p>	<p><b>Pas de définition des « contenus d'information se rattachant à un débat d'intérêt général » <i>stricto sensu</i>, souvent rapprochés d'autres notions</b> (définies dans les politiques publicitaires de la plateforme) :</p> <ul style="list-style-type: none"> <li>- « publicités à caractère politique et militant » (<i>Snapchat</i>) ;</li> <li>- « publicités défendant une question législative d'importance nationale, publicités faisant référence à une élection, un parti politique ou un candidat clairement identifié » (<i>Twitter</i>).</li> </ul>
<p>Identification des annonceurs politiques</p>	<p>Afin notamment d'éviter l'influence de puissances étrangères, des plateformes ont mis en place <b>des processus d'identification particuliers pour les annonceurs politiques</b> (autorisation dans un seul pays, obligation de fournir un document d'identité pour le pays ciblé...) (<i>Facebook, Google</i><sup>78</sup>, <i>YouTube, Snapchat</i><sup>79</sup>, <i>Twitter</i><sup>80</sup>).</p>

<sup>78</sup> A noter dans la déclaration : « nous avons annoncé le 23 avril 2020 des mesures allant dans le sens de plus de transparence pour nos utilisateurs et notamment l'extension du processus de validation des annonceurs à tous nos annonces (et plus uniquement aux annonceurs politiques) ainsi qu'une fonctionnalité permettant d'avoir plus



Restrictions et examen particulier	<p><b>Restrictions de ciblage</b> (<i>Twitter</i><sup>81</sup>);</p> <p><b>Examens préalables avant diffusion</b> (<i>Yahoo, Snapchat, Microsoft</i>).</p>
------------------------------------	--

MISE EN PLACE DE BIBLIOTHEQUES PUBLICITAIRES	
Mise en place de bibliothèques publicitaires publiques	<p><b>Bases de données publicitaires publiques</b> (<i>Facebook, Google, Twitter, Snapchat</i>).</p> <p>- Ces bases peuvent prendre <b>différentes formes</b> : bases de données avec moteurs de recherche, API, centre de transparence, rapports, pages...</p> <p><i>LinkedIn</i> : l'onglet « publicités » sur les pages <b>Entreprise</b> (les pages vers lesquelles le contenu sponsorisé renvoie) permet de voir les contenus sponsorisés actuels ou passés pour lesquels <i>LinkedIn</i> a été payé par le propriétaire de la page ou une personne affiliée.</p> <p>Pas d'information fournie (<i>Yahoo, Jeuxvidéo.com, Microsoft</i>).</p>
Contenus	<p>Les <b>contenus peuvent varier</b> :</p> <p>- <i>Facebook</i> : bibliothèque publicitaire publique et consultable <b>même sans compte</b>, incluant toutes les publicités « actives et en cours » (sauf les publicités politiques qui sont conservées 7 ans);</p> <p>- <i>Snapchat</i> : bibliothèque des publicités politiques et militantes (uniquement).</p>

#### Mesures prises dans le cadre de la crise de la COVID-19

Selon les déclarations, **blocage de publicités « tirant parti » de la COVID-19** (*Google* : retrait de plusieurs millions d'annonces en mars-avril ; *Facebook*).

### 3.5.2. Analyse des mesures prises par les opérateurs

Dans l'ensemble, les plateformes ont répondu positivement aux questions du CSA relatives aux contenus sponsorisés, même si le niveau de détail des informations communiquées est variable. Certaines réponses sont effet assez laconiques (*Microsoft, Yahoo*). *Unify* s'estime

*d'informations sur l'annonceur, pour n'importe quelle publicité diffusée sur Google Ads. Cette fonctionnalité devrait être active dès l'été 2020 ».*

<sup>79</sup> Les publicités politiques ne peuvent pas être payées par des personnes ou entités qui ne résident pas dans le pays où sera diffusée la publicité.

<sup>80</sup> *Twitter* limite la publicité et exige la **certification des annonceurs** (selon une procédure non définie) pour les publicités « qui éduquent, sensibilisent ou appellent les gens à agir en rapport avec l'engagement civique, la croissance économique, la gestion de l'environnement ou les causes d'équité sociale ».

<sup>81</sup> Pour les publicités dédiées à une cause, les possibilités de ciblage se limitent au ciblage géographique, par mot-clé et par intérêt. Aucun autre type de ciblage n'est autorisé, y compris les audiences personnalisées.



non concerné par les questions relatives aux contenus sponsorisés, dès lors que la promotion de contenus d'information se rattachant à un débat d'intérêt général en contrepartie d'une rémunération n'est pas permise sur *Doctissimo* et *Aufeminin* ; on y trouve néanmoins des contenus sponsorisés au sens large. Le CSA appelle l'ensemble des opérateurs à faire preuve de davantage de coopération à l'avenir.

- **Une transparence nécessaire pour l'utilisateur**

L'ensemble des plateformes disposent de **politiques publicitaires publiques**, notamment en matière de publicités trompeuses ou inappropriées, et accessibles aux utilisateurs. **La forme et le fond peuvent varier**. A titre d'exemple, certaines plateformes font apparaître ces règles dans un document unique quand d'autres ont pris le parti d'adopter plusieurs documents. Dans un souci de clarté et de transparence pour l'utilisateur, il est essentiel que ces règles soient facilement accessibles et compréhensibles.

Dans une logique de transparence, **l'ensemble des plateformes distinguent les publicités par leur labellisation**. Cette dernière peut varier dans sa forme. Elle est toujours assez claire et permet ainsi à l'utilisateur de comprendre la nature du contenu vu.

Certaines plateformes permettent à l'utilisateur de comprendre en quelques clics, depuis le contenu promu, **pourquoi il a été ciblé par une publicité**.

Enfin, des plateformes permettent aux utilisateurs de **paramétrer leurs préférences publicitaires, voire de masquer des publicités**. Les réponses au questionnaire sur ce point sont malheureusement lacunaires, mais il s'agit d'une initiative positive reposant sur la responsabilisation des utilisateurs.

- **Un besoin de transparence accru dans le traitement des signalements**

Seules deux plateformes offrent la faculté de signaler une publicité pour « fausse information ». D'autres permettent des signalements pour les contenus « inappropriés » ou « trompeurs ». D'autres encore donne la possibilité de rédiger des commentaires pour préciser l'objet du signalement.

En matière de communication avec l'auteur du signalement, seul *Facebook* précise que l'utilisateur peut suivre l'état de son signalement par le biais d'une page dédiée. Le CSA incite les plateformes à intensifier leurs efforts sur ce point en proposant, par exemple, un suivi des signalements de l'utilisateur par message ou sur une page dédiée.

La recommandation formulée par le CSA dans sa délibération du 15 mai 2019 invite les plateformes à « *appeler la vigilance des utilisateurs sur les contenus qui ont fait l'objet de signalements* ». Si *LinkedIn* précise mettre les publicités signalées en quarantaine afin de les étudier, les autres acteurs n'ont pas précisé si des mesures étaient prises en ce sens. Afin d'éviter ou de ralentir la dissémination de fausses informations par le biais de publicités, plusieurs actions sont envisageables, telles que, par exemple, la mise en quarantaine du contenu sponsorisé jusqu'à vérification par les modérateurs ou la publication d'un message d'avertissement sur un contenu promu qui aurait fait l'objet d'un certain nombre de signalements pour « fausse information » ou équivalent.



Enfin, il est intéressant de noter que la majorité des répondants affirment que le traitement des signalements se fait à la fois par des moyens automatisés et humains.

- **Sécurité des marques et identification des annonceurs**

Certaines plateformes ont affirmé avoir mis en place des **outils visant à garantir la sécurité des marques**. Ces mesures peuvent permettre d'éviter qu'une marque ne soit associée à un compte, une page ou tout autre contenu disséminant des fausses informations. Elles assèchent par ailleurs les ressources financières tirées par le créateur du contenu de la propagation de fausses informations. Dans une logique similaire, des plateformes ont précisé avoir mis en place des mesures de démonétisation (*Facebook*), supprimant l'association de publicités à des contenus problématiques.

La grande majorité des plateformes n'ont **pas précisé les mesures permettant aux contenus sponsorisés de gagner en visibilité**, qui sont au cœur de leur modèle économique. La compréhension de ces mécanismes est pourtant essentielle pour lutter contre les fausses informations.

Enfin, certaines plateformes ont précisé les conditions pour pouvoir faire la promotion d'un contenu sur leur service. Des **mesures préalables d'identification des annonceurs** peuvent permettre d'empêcher les annonceurs malveillants, souhaitant propager des fausses informations, de le faire.

- **La promotion de contenus d'information se rattachant à un débat d'intérêt général**

La promotion de contenus d'information se rattachant à un débat d'intérêt général revêt un enjeu démocratique significatif. Ces derniers peuvent en effet contribuer à former l'opinion des citoyens. Lutter contre les fausses informations dans ce contexte est essentiel pour préserver le pluralisme des opinions.

L'ensemble des plateformes semblent reconnaître la particularité et l'importance de ces contenus. Elles en ont toutefois des approches différentes. Il apparaît donc important d'œuvrer à l'harmonisation<sup>82</sup> de la définition de cette notion.

Les mesures prises en ce domaine sont particulièrement variées. Certaines plateformes interdisent de manière permanente la promotion de ces contenus ; d'autres appliquent des restrictions de ciblage. D'autres encore distinguent ces promotions des autres publicités par le biais d'une labellisation particulière comportant parfois davantage d'informations sur l'annonceur. Ces initiatives répondent à une logique de transparence et d'information des utilisateurs et ne peuvent être qu'encouragées.

---

<sup>82</sup> Cette problématique avait déjà été relevée par le groupe des régulateurs européens des services de médias audiovisuels (ERGA) dans les rapports suivants : *Report of the activities carried out to assist the European Commission in the intermediate monitoring of the Code of practice on disinformation*, juin 2019, et *ERGA report on disinformation: Assessment of the implementation of the code of practice*, mai 2020.



De nombreux acteurs ont mis en place un **processus d'identification particulier des annonceurs voulant faire la promotion de contenus d'information se rattachant à un débat d'intérêt général**. La pertinence de ces mesures est réelle, notamment dans l'objectif de parer de possibles influences étrangères dans le débat démocratique interne. Néanmoins, elles peuvent varier d'une plateforme à l'autre<sup>83</sup>.

- **La mise en place de bibliothèques publicitaires**

La mise en place, par certaines plateformes, de bases de données publiques relatives aux contenus promus est une **avancée importante**. Bien que la forme, le contenu et les filtres de recherches varient beaucoup, ces outils sont des éléments clés pour améliorer la transparence mais également pour comprendre l'impact des publicités.

Cependant, **les informations contenues dans ces bases de données ne semblent pas toujours suffisantes**. Dans deux rapports relatifs à la lutte contre la désinformation<sup>84</sup> publiés en 2019 et en 2020, le groupe des régulateurs européens des services de médias audiovisuels (ERGA) relève que ces informations sont agrégées, parfois difficilement trouvables ou filtrées, et qu'il est parfois impossible d'en garantir la fiabilité.

### 3.5.3. Recommandations du CSA

En conclusion, le CSA note que les plateformes ont mis en place des mesures, particulièrement variées, allant dans le bon sens. Cependant, des efforts soutenus doivent être faits en matière de contenus sponsorisés. Ainsi, le CSA formule les recommandations suivantes.

1. Le CSA encourage les plateformes à continuer de mettre en place une **labellisation claire et directement accessible** des contenus promus, aisément visible par l'utilisateur et précisant distinctement l'identité de l'annonceur.
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).
3. **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.
4. 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**

<sup>83</sup> Cf. rapports ERGA précités.

<sup>84</sup> *Ibid.*



**gagner en visibilité.** La compréhension de ces mécanismes est essentielle pour lutter contre les fausses informations.

5. 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.
6. 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.**
7. 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.

### 3.6. Education aux médias et à l'information

Les citoyens doivent pouvoir rester maîtres de leurs choix en matière de sources d'information. Il est donc impératif que l'utilisateur d'une plateforme soit aidé et accompagné dans ses usages afin de comprendre les spécificités de son offre d'information, les logiques économiques sous-jacentes et les risques d'influence ou de manipulation. L'éducation aux médias et à l'information (EMI), pour tous, est ainsi devenue une clé essentielle pour comprendre les enjeux et maîtriser l'utilisation des outils numériques.

#### 3.6.1. Typologie des mesures prises par les opérateurs

Mesures permettant d'identifier des contenus pertinents et fiables	<b>Lettres hebdomadaires récapitulant l'information</b> ( <i>Dailymotion, LinkedIn</i> ).  <b>Mise en avant d'informations officielles ou fact-checkées</b> ( <i>Facebook, Microsoft, Twitter</i> ).  <i>Bing</i> et <i>Newsguard</i> (en partenariat) ont développé un système d' <b>évaluation des contenus proposés</b> : des notes (couleurs) en fonction de la fiabilité des informations relayées.
Mesures permettant de former à l'usage	<b>Initiatives permettant de former les publics à l'utilisation des plateformes :</b>



<p>des plateformes et à l'identification des contenus</p>	<ul style="list-style-type: none"> <li>- <b>initiatives d'éducation aux médias sur le service</b> : <i>Twitter</i> : hashtags pendant la semaine de l'EMI ; <i>Facebook</i> : portail pour les jeunes, Ressources sur le numérique ; <i>Google</i> : campagne Be Internet Awesome, Google News Initiative) ;</li> <li>- <b>soutien financier</b> à des initiatives tierces : <i>Twitter</i> : organisation de la semaine de l'EMI à l'UNESCO ; <i>Google</i> : soutien au projet Les Super Héros du Net porté par l'association e-enfance ; <i>Snapchat</i> : finance des actions d'e-enfance ; <i>Facebook</i> : participation au Yo !Fest, Freeformers ;</li> <li>- <b>partenariats avec des institutions publiques</b> : <i>LinkedIn</i> avec Pôle Emploi ; <i>Snapchat</i> avec l'OMS.</li> </ul> <p><b>Initiatives permettant aux publics de développer leur esprit critique :</b></p> <ul style="list-style-type: none"> <li>- <b>campagnes d'information et de sensibilisation pendant les élections</b> (<i>Facebook</i> pendant les élections européennes et municipales) ;</li> <li>- <b>initiatives centrées sur le public jeune</b> : <i>Facebook</i> : ressources sur le numérique pour les 11 à 18 ans, Fonds pour le civisme en ligne ; <i>Google</i> : Programme Be Internet Citizens, Les Complots rigolos, « Il paraît que » ; <i>Twitter</i> : Semaine de la presse ; <i>Snapchat</i> : jeu en réalité augmentée pour lutter contre les fausses informations (dans le cadre de la COVID-19).</li> </ul>
<p>Aides à la recherche</p>	<p><b>Financement de chaires universitaire et de colloques et groupes de recherche</b> (<i>Microsoft</i> : financement de la recherche contre les deepfakes, Oxford Internet Institute, Woodrow Wilson Disinformation project, Truth and Trust Online).</p> <p><b>Financement d'initiatives en faveur du fact-checking</b> : <i>Google</i> : First Draft Coalition, Trust Project ; <i>Microsoft</i> : Trust Project).</p> <p><b>Accès facilité aux algorithmes et aux API</b>, avec notamment des API publiques (<i>Twitter</i> ; <i>Facebook</i> pour sa bibliothèque publicitaire).</p>

### 3.6.2. Analyse des mesures prises par les opérateurs

Si certaines plateformes ont développé des actions en matière d'éducation aux médias et à l'information, elles sont **encore trop peu nombreuses à s'être investies** ; sur les onze ayant répondu au questionnaire, seules cinq ont rendu compte d'actions concrètes d'ampleur : *Facebook*, *Google*, *Microsoft*, *Twitter* et *Snapchat*.

- **Des approches multidimensionnelles en matière d'éducation aux médias et à l'information**

Certains des opérateurs ayant mené des actions en matière d'EMI ont pris soin de développer des projets embrassant l'ensemble des problématiques, dans une logique de complémentarité d'approches : **apprendre par l'intermédiaire des plateformes mais**



**également donner les lignes directrices permettant un bon usage de ces outils.** *Facebook* et *Twitter* se démarquent particulièrement à cet égard.

S'agissant de *Facebook*, on relève notamment le lancement en mai 2019 du « Portail pour les jeunes utilisateurs de Facebook », et à l'automne, des « Ressources sur le numérique » : une collection de cours disponibles en 45 langues qui permet aux jeunes de 11 à 18 ans de développer leur esprit critique et de mieux se protéger sur les réseaux sociaux.

*Twitter* a privilégié le développement d'un partenariat avec l'UNESCO qui lui a permis, en s'appuyant sur l'expertise reconnue de cette institution, de développer plusieurs guides visant à accompagner ses nouveaux utilisateurs ou à décrypter ce qui leur est donné à voir (ex : distinguer les faits de la fiction, etc.).

Par ailleurs, plus récemment, lors de l'épidémie de la COVID-19, ces deux plateformes se sont illustrées en mettant en avant sur leurs interfaces des contenus pédagogiques et des informations vérifiées concernant l'épidémie.

- **Des initiatives encore trop souvent destinées aux seules cibles jeunes**

De manière générale, les actions déclarées par les opérateurs en matière d'EMI **visent presque exclusivement les publics jeunes** : de 6 ans à 20 ans.

Si les adultes sont parfois visés, c'est uniquement en tant que parents de jeunes utilisateurs et jamais comme potentiels utilisateurs de la plateforme. Or, d'après une étude publiée le 9 janvier 2019 sur « Science Advances », les utilisateurs âgés de 65 ans et plus partagent deux fois plus d'articles faux sur *Facebook* que les 45-65 ans et même sept fois plus que les personnes âgées de 18 à 29 ans<sup>85</sup>.

- **Des données sur les audiences touchées par les initiatives mais aucune concernant leur impact sur les comportements des utilisateurs**

Si certains opérateurs ont communiqué au CSA des données concernant le nombre de personnes touchées par leurs actions (*Google*), **aucun ne lui a transmis des informations rendant compte de l'impact des initiatives menées sur le comportement des utilisateurs.**

L'ERGA, dans son rapport de mai 2020, avait déjà souligné l'impossibilité de se prononcer sur l'effectivité de ces actions en l'absence de telles données : « *as highlighted by the French CSA and several other NRAs, in the absence of any data on the uptake and impact of these initiatives provided by the platforms it is not possible for the monitors to evaluate their efficiency.* »<sup>86</sup>

---

<sup>85</sup> « *Less than you think : Prevalence and predictors of fake news dissemination on Facebook* », Andrew Guess, Jonathan Nagler, Joshua Tucker, *Science Advances*, le 9 janvier 2019 : <https://advances.sciencemag.org/content/5/1/eaau4586>.

<sup>86</sup> Cf. note 82.



- **Des rapports insuffisants avec le secteur de la recherche s'agissant de l'exploitation des données**

En 2019, **seuls trois opérateurs sur onze** ont déclaré avoir mené des actions avec le monde de la recherche : *Microsoft*, *Twitter* et *Google*.

Ces actions consistent principalement en un **financement de projets de recherche** (*Google* qui soutient une étude sur la désinformation et la confiance dans le journalisme, *Microsoft* avec « Oxford Internet Institute » et le projet « WoodrowWilson », etc.), de conférences, de forums (*Microsoft* avec la conférence « Truth and Trust Online » et le forum « Technologie/University/Politic (TUP) », etc.) ainsi que d'initiatives en faveur du *fact-checking* (*Google* avec le « First Draft Coalition » et le « Trust Project » et *Microsoft* avec le « Trust Project »).

A noter que *Twitter*, en plus de financer des projets d'études, publie les **archives complètes de ses tweets et rend accessible ses API**.

Toutefois, bien que ces initiatives soient nombreuses et paraissent volontaristes, **le CSA s'interroge**, comme a pu le faire l'ERGA dans son rapport précité<sup>87</sup>, **sur l'exploitabilité des données transmises aux chercheurs** par les plateformes. Par ailleurs, l'ERGA a souligné que si des partenariats avaient bien été noués, ils nécessitaient des échanges plus fréquents : « *any NRAs confirm that specific events/discussions and partnerships with research and academic institutions remain episodic and largely inadequate to support any rigorous analysis and monitoring of online disinformation trends.* »<sup>88</sup>

### 3.6.3. Recommandations du CSA

Aussi, le CSA formule les recommandations suivantes.

1. Il invite les opérateurs à **développer des approches multidimensionnelles en matière d'EMI**, 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.).
2. 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.).
3. Il les incite à **évaluer l'impact de leurs actions**.
4. 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**.

<sup>87</sup> *Ibid*, p. 29.

<sup>88</sup> *Ibid*, p. 30.



### 3.7. Moyens complémentaires mis en œuvre

#### Des opérateurs ont fait état de mesures complémentaires dans le cadre de la lutte contre les fausses informations.

- **Actions de lutte contre les deepfakes**

*Facebook* rappelle que si ce type de contenus est encore rare, leur usage augmente et constitue un défi. Il coopère avec des experts afin d'améliorer les solutions de détection des vidéos manipulées. Est supprimé du service tout contenu vidéo « *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* » ou « *produit d'une intelligence artificielle ou d'un apprentissage automatique qui fusionne, remplace ou superpose des contenus dans une vidéo qui semble vraie* »<sup>89</sup>.

En 2019, *Facebook* a lancé et subventionné le *Deepfake detection challenge*, projet de recherche international, afin de faire avancer la recherche dans le champ de la détection des *deepfakes*. Il est également associé avec Reuters afin d'aider les sociétés d'information, par des formations, à identifier les *deepfakes*.

*Google* et *Microsoft* financent la recherche en matière de lutte contre les *deepfakes*.

*LinkedIn* interdit les *deepfakes* et ses équipes peuvent « hacher » les cas connus de *deepfakes* afin d'éviter leur propagation sur la plateforme.

- **Actions de lutte contre les attaques informatiques pendant les élections**

*Google* a développé les projets « Protection Elections » et « Project Shield » qui consistent en un ensemble de mesures de sécurité afin de lutter contre les applications malveillantes et le hameçonnage. L'opérateur mène aussi des actions de formation des équipes de campagne et des partis politiques, afin de les aider à mettre en œuvre des dispositifs de protection contre ce type d'attaques.

*Microsoft* a développé un « Programme de défense de la Démocratie » rassemblant des outils de lutte contre les attaques informatiques et s'est doté d'une équipe centralisée de sécurité numérique dans le cadre de sa réponse à l'Appel de Christchurch.

Fin 2019, *Facebook* a renforcé sa politique de lutte contre la pression sur les électeurs en supprimant les contenus visant à dissuader ou empêcher les utilisateurs de voter.

- **Affichage d'une information exhaustive dans le cadre de leurs recherches**

*Google* et *Bing* disposent de fonctions qui permettent aux utilisateurs d'avoir l'ensemble des articles et sources d'informations lors d'une requête données (fonctions « Voir la couverture complète » et « Recherche intelligente »).

---

<sup>89</sup> Sauf contenus parodiques, satiriques ou modifiés dans le but de supprimer ou modifier l'ordre des mots.



- **Restriction d'accès au « Live »**

*Facebook* limite l'utilisation de Facebook *Live* en empêchant l'utilisation par les personnes ayant violé certaines règles, dont celles relatives aux « Individus et Organismes Dangereux », pour une période déterminée.



## Conclusion

Le bilan des déclarations des opérateurs atteste des **efforts importants et croissants** fournis par les opérateurs pour œuvrer à la lutte contre la propagation des fausses informations sur leur(s) service(s). À cet égard, leurs efforts lors de la crise de la COVID-19 témoignent de leur prise de conscience de l'importance de ces enjeux. Le CSA tient en outre à saluer **l'esprit de coopération et la disponibilité dont ont fait preuve la grande majorité d'entre eux**, ainsi que la richesse du dialogue noué avec eux.

On ne saurait sous-estimer la complexité inhérente à cette entreprise qui nécessite, comme le rappellent certains opérateurs dans leurs déclarations, de trouver un **équilibre entre la lutte contre les fausses informations et la liberté d'expression des utilisateurs**, équilibre complexe au vu de la masse de contenus à traiter et de la difficulté de juger de la véracité de certains contenus.

Toutefois, comme il l'a noté au long de ce bilan, **le CSA demande que les mesures mises en œuvre soient encore améliorées.**

Le dispositif de signalement, présent sur toutes les plateformes, ne permet pas toujours à l'utilisateur **d'identifier et de signaler aisément une fausse information**. En outre, toutes les plateformes devraient pouvoir mettre en place un **outil de suivi des signalements** pour leurs auteurs, ainsi qu'une **procédure de recours** contre les décisions prise à l'issue de l'examen de ces signalements. Le CSA les invite à développer des outils à cette fin.

**Le CSA souhaite en outre une plus grande transparence à son égard et à celui des utilisateurs.** En particulier, des **données chiffrées fiables et exhaustives** portant sur la présence des contenus susceptibles de propager de fausses informations, leur signalement par les utilisateurs et leur traitement par les opérateurs seraient très précieux pour mieux comprendre le phénomène et ses enjeux.

Il engage également les opérateurs à **communiquer davantage d'éléments sur l'intelligibilité et la redevabilité de leurs algorithmes** d'organisation, de sélection et d'ordonnancement des contenus vis-à-vis de leurs utilisateurs et à lui transmettre les principes de fonctionnement et les objectifs de ces algorithmes. Dans ce domaine, **l'ouverture et la collaboration avec le monde de la recherche, si elles existent, devraient être grandement développées pour permettre des travaux experts et indépendants sur le fonctionnement et les effets de ces algorithmes.**

Par ailleurs, le CSA **encourage fortement le développement de partenariats entre les plateformes et des tiers vérificateurs de confiance**, comme la mise en avant des contenus de ces derniers, qui ont à nouveau démontré leur nécessité lors de la crise sanitaire en 2020.

À la lumière du présent bilan, le CSA entend, **dans les prochains mois, poursuivre la réflexion** sur l'amélioration de la lutte contre la diffusion de fausses informations **avec son Comité d'experts sur la désinformation en ligne ainsi qu'avec les opérateurs.**



## ANNEXES

### Annexe 1 : Dispositions légales et réglementaires

#### 1. La définition légale de l'opérateur de plateforme en ligne

**Code de la consommation, Livre 1<sup>er</sup>, Titre 1<sup>er</sup>, Chapitre 1<sup>er</sup>, Article L111-7, modifié par la loi n°2016-1321 du 7 octobre 2016**

« I.- Est qualifiée d'opérateur de plateforme en ligne toute personne physique ou morale proposant, à titre professionnel, de manière rémunérée ou non, un service de communication au public en ligne reposant sur :

- 1° Le classement ou le référencement, au moyen d'algorithmes informatiques, de contenus, de biens ou de services proposés ou mis en ligne par des tiers ;
- 2° Ou la mise en relation de plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service. (...) »

#### 2. Le devoir de coopération des opérateurs de plateforme en ligne en matière de lutte contre la diffusion de fausses informations

**Loi n°2018-1202 du 22 décembre 2019 relative à la lutte contre la manipulation de l'information.**

**Titre III : Devoir de coopération des opérateurs de plateforme en ligne en matière de lutte contre la diffusion de fausses informations**

##### Article 11

*I. - Les opérateurs de plateforme en ligne mentionnés au premier alinéa de l'article L. 163-1 du code électoral mettent en œuvre des mesures en vue de lutter contre la diffusion de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un des scrutins mentionnés au premier alinéa de l'article 33-1-1 de la loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication.*

**Ils mettent en place un dispositif facilement accessible et visible permettant à leurs utilisateurs de signaler de telles informations, notamment lorsque celles-ci sont issues de contenus promus pour le compte d'un tiers.**

*Ils mettent également en œuvre des mesures complémentaires pouvant notamment porter sur :*

- 1° La transparence de leurs algorithmes ;
- 2° La promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle ;
- 3° La lutte contre les comptes propageant massivement de fausses informations ;
- 4° L'information des utilisateurs sur l'identité de la personne physique ou la raison sociale, le siège social et l'objet social des personnes morales leur versant des rémunérations en contrepartie de la promotion de contenus d'information se rattachant à un débat d'intérêt général ;
- 5° L'information des utilisateurs sur la nature, l'origine et les modalités de diffusion des contenus ;
- 6° L'éducation aux médias et à l'information.



Ces mesures, ainsi que les moyens qu'ils y consacrent, sont rendus publics. **Chaque opérateur adresse chaque année au Conseil supérieur de l'audiovisuel une déclaration dans laquelle sont précisées les modalités de mise en œuvre desdites mesures.**

II. - Le présent article est applicable en Polynésie française, dans les îles Wallis et Futuna, en Nouvelle-Calédonie et dans les Terres australes et antarctiques françaises.

#### **Article 12 :**

« Art. 17-2.-Le Conseil supérieur de l'audiovisuel contribue à la lutte contre la diffusion de fausses informations susceptibles de troubler l'ordre public ou de porter atteinte à la sincérité d'un des scrutins mentionnés au premier alinéa de l'article 33-1-1 de la présente loi.

En cas de nécessité, il adresse, à ce titre, aux opérateurs de plateforme en ligne mentionnés au premier alinéa de l'article L. 163-1 du code électoral des recommandations visant à améliorer la lutte contre la diffusion de telles informations.

**Il s'assure du suivi de l'obligation pour les opérateurs de plateforme en ligne de prendre les mesures prévues à l'article 11 de la loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information.**

**Il publie un bilan périodique de leur application et de leur effectivité.** A cette fin, il recueille auprès de ces opérateurs, dans les conditions fixées à l'article 19 de la présente loi, toutes les informations nécessaires à l'élaboration de ce bilan. »

#### **Article 13 :**

Les opérateurs de plateforme en ligne mentionnés au premier alinéa de l'article L. 163-1 du code électoral désignent un représentant légal exerçant les fonctions **d'interlocuteur référent sur le territoire français** pour l'application des dispositions prévues au présent titre et au troisième alinéa du 7 du I de l'article 6 de la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique. (...)

**Décret n° 2019-297 du 10 avril 2019 relatif aux obligations d'information des opérateurs de plateforme en ligne assurant la promotion de contenus d'information se rattachant à un débat d'intérêt général**

#### **Article 1**

«Le nombre de connexions au-delà duquel les opérateurs de plateforme en ligne sont soumis aux obligations de l'article L. 163-1 est fixé à cinq millions de visiteurs uniques par mois, par plateforme, calculé sur la base de la dernière année civile. (...) ».



## Annexe 2 : Questionnaire du CSA aux opérateurs

### 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

#### Rappel du cadre

*Le titre III de la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information instaure un devoir de coopération en matière de lutte contre la diffusion de fausses informations à la charge des opérateurs de plateforme en ligne<sup>1</sup>.*

*L'article 11 de cette même loi impose à ces opérateurs de prendre des mesures en vue de lutter contre la diffusion de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un des scrutins mentionnés au premier alinéa de l'article 33-1-1 de la loi du 30 septembre 1986<sup>2</sup>. A cet égard, l'article 11 précité dispose que les mesures ainsi prises, de même que les moyens que les opérateurs de plateforme en ligne y consacrent, sont rendus publics et font l'objet d'une déclaration annuelle au Conseil supérieur de l'audiovisuel.*

*Le 15 mai 2019, sur le fondement de l'article 12 de la loi du 22 décembre 2018, le Conseil a adressé aux opérateurs une recommandation visant à améliorer la lutte contre la diffusion de fausses informations susceptibles de troubler l'ordre public ou de porter atteinte à la sincérité du scrutin. Il les invite à mettre en œuvre un certain nombre de mesures et à déployer les moyens humains et techniques nécessaires en vue d'atteindre les objectifs fixés par la loi, et leur demande de lui communiquer leur déclaration au plus tard le 31 mars de l'année suivant l'année d'exercice sur laquelle elle porte.*

*Ainsi, l'année d'exercice 2019 doit faire l'objet d'une déclaration adressée au Conseil au plus tard le 31 mars 2020.*

#### Objectifs du questionnaire

*Le questionnaire, rédigé à partir de la recommandation du 15 mai 2019, a été nourri par la veille et les observations que le Conseil a menées et enrichies à la lumière des réflexions des membres du Comité d'experts sur la désinformation en ligne auprès du Conseil.*

*Il a pour objectifs d'accompagner les opérateurs dans la préparation de leur déclaration annuelle et d'alimenter le bilan comparé que le Conseil dressera de l'application et de l'effectivité des mesures prises par ces derniers. A cet égard, le Conseil rappelle qu'il prendra en compte la pluralité des modèles des plateformes et l'adéquation des moyens mis en œuvre sur chacune d'entre elles à l'ampleur et à l'impact du phénomène de manipulation de l'information.*

---

<sup>1</sup> Sont concernés les opérateurs mentionnés au premier alinéa de l'article L 163-1 du code électoral, à savoir les opérateurs de plateforme en ligne au sens de l'article L. 111-7 du code de la consommation dont l'activité dépasse cinq millions de visiteurs uniques par mois, par plateforme, calculé sur la base de la dernière année civile.

<sup>2</sup> Sont visées par ces dispositions l'élection du Président de la République, les élections générales des députés, l'élection des sénateurs, l'élection des représentants au Parlement européen et les consultations référendaires.



Les déclarations des opérateurs pourront également, grâce à l'éclairage qu'elles apporteront au regard de leurs éventuelles lacunes, amener le Conseil à ajuster le questionnaire relatif à l'exercice 2020.

### **Modalités et précisions méthodologiques :**

Les opérateurs sont invités à communiquer une déclaration respectant les modalités suivantes :

- une réponse par service de plateforme en ligne doit être communiquée ; dans l'hypothèse où l'application des mesures différerait au sein d'un même service, d'un espace à un autre, l'opérateur est invité à le préciser dans ses réponses ;
- les informations et données fournies doivent concerner l'année d'exercice 2019 (sauf indications spécifiques sur des dispositions passées ou des projets à venir) ; à cet égard, l'opérateur peut fournir des données ou des informations sur des dispositifs existants en 2019 et qui auraient été arrêtés depuis (en mentionnant cet arrêt) ;
- les informations et données fournies doivent, dans la mesure du possible, concerner l'activité du service de plateforme en France ; à défaut, l'opérateur peut fournir des données agrégées sous réserve qu'elles incluent l'activité en France et que leur base géographique soit explicitement mentionnée ;
- les déclarations seront considérées comme publiques et publiées sur le site internet du Conseil, à l'exception des éléments couverts par le secret des affaires dont la confidentialité sera explicitement demandée et qui resteront à la seule disposition du Conseil. Le Conseil invite les opérateurs à justifier strictement le recours au secret des affaires ;
- l'opérateur est invité à fournir tous les éléments quantitatifs ou qualitatifs permettant d'attester les informations déclarées (notamment des données chiffrées, des captures d'écran, de la documentation interne ou publique ou des hyperliens) ;
- de manière générale, l'opérateur s'attachera à expliquer ses choix et à indiquer au Conseil pourquoi un dispositif n'a pas ou a partiellement été mis en œuvre ;
- il précisera, à chaque fois que cela sera pertinent :
  - si l'utilisateur doit ou non disposer **d'un compte pour accéder au dispositif et informations, et ce dans leur intégralité,**
  - la **facilité d'accès aux dispositifs** mis en œuvre : mise à disposition de l'ensemble des utilisateurs, visibilité, nombre d'actions et/ou d'hyperliens pour y parvenir depuis la page d'accueil du service et depuis un contenu en particulier,
  - **l'accessibilité des** dispositifs mis en œuvre (ex. : signalement, informations) aux personnes en situation de handicap visuel ou auditif,
  - le déploiement des dispositifs sur **l'ensemble des versions du service**, quel que soit le support utilisé (ex. : smartphone, ordinateur, tablette, smart tv, enceintes connectées), la version de sa plateforme (ex. : applications, site internet, assistants vocaux), le navigateur, le moteur de recherche, le système d'exploitation ; ainsi que la nature des différences qui pourraient exister entre les versions,



- le déploiement du dispositif sur **tous les types de contenus** (texte, vidéo, image, commentaires, contenus sponsorisés, autres) ; ainsi que, le cas échéant, la nature des différences en fonction des contenus,
  - la **langue** dans laquelle les dispositifs et informations sont proposés à l'utilisateur par défaut et comment elle est déterminée ainsi que, le cas échéant, la possibilité de modifier ou configurer aisément en français,
  - comment l'opérateur fait en sorte de garantir que l'application des mesures se fait **dans le respect de la liberté d'expression et de communication** ;
- pour chaque ensemble de mesures mis en place, le Conseil souhaite connaître :
    - la **date** de mise en œuvre,
    - s'il a été fait **appel à un tiers** pour les mettre en place et/ou les appliquer,
    - les **changements** d'ordres technologique, ergonomique, en matière de ressources humaines ou autres qui ont été nécessaires afin de se conformer à la loi, ainsi que les coûts de cette mise en conformité,
    - le **bilan** qu'en tire l'opérateur ; en particulier, ce dernier est invité à communiquer les résultats de mesures quantitatives ou qualitatives qu'il aurait effectuées, fait effectuer ou dont il aurait connaissance sur la visibilité, l'utilisation, les performances et les effets des dispositifs mis en œuvre,
    - les **freins ou difficultés** qu'il a rencontrés dans leur mise en œuvre et plus généralement dans la mise en conformité du service aux textes en vigueur en matière de lutte contre la manipulation de l'information,
    - les éventuelles **initiatives complémentaires** qu'il aurait prises afin de lutter contre la désinformation sur son service.

De manière générale, les opérateurs sont encouragés à partager avec le Conseil des données lui permettant d'exercer sa mission de contribution à la lutte contre la diffusion de fausses informations prévue à l'article 17-2 de la loi du 30 septembre 1986 relative à la liberté de communication (dans le respect du cadre juridique en matière de protection des données personnelles).

### Précisions terminologiques

Dans le questionnaire, sont désignés par :

- « Service » : le service de plateforme en ligne faisant l'objet de la déclaration ;
- « Opérateur » : l'opérateur du service de plateforme en ligne faisant l'objet de la déclaration ;
- « Loi » : la loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information ;
- « Conseil » : le Conseil supérieur de l'audiovisuel.

### Questions liminaires : présentation du service

- **Nom du service :**
- **Nom et raison sociale de la société opératrice :**
- **Identité et fonction du référent légal de l'opérateur de la plateforme en ligne en France :**
- **Nombre de visiteurs uniques par mois de la plateforme en ligne, calculé sur la base de l'année d'exercice (2019) :**



- **Modalités de mise à disposition du service auprès du public français :**
- **Disponibilité sur internet (URL du site internet) :**
- **Disponibilité via des applications (systèmes d'exploitation concernés) :**
- **Disponibilité via des assistants vocaux :**
- **Langue du service sur le territoire français (par défaut) :**
- **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) ? :**
  
- **Description du service :**
- **Objet :**
- **Population(s) ciblée(s) (tranches d'âges, communautés d'intérêt...) :**
- **Types de contenus qui peuvent être publiés et consommés sur le service (vidéos, textes, sons, images, jeux vidéo...) :**
- **Thématique(s) :**
- **Modèle économique du service (types d'activités rémunératrices pour l'opérateur tirées de l'exploitation du service en France) :**
- **Effectifs employés aux activités du service dans le monde et en France :**
- **Comptes de résultats pour l'activité du service en France en 2019 :**
  - **Chiffres d'affaires :**
  - **Produits d'exploitation :**
  - **Charges d'exploitation :**
  - **Résultat net :**
- **Autres informations ou spécificités :**

## 1. Dispositif de signalement de fausses informations

1.1. L'opérateur indiquera si son service dispose d'un dispositif de signalement des contenus publiés, et si ce dispositif permet à l'utilisateur de signaler de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité du scrutin (et le cas échéant, dans quels termes). Plus précisément, le Conseil souhaite connaître le nombre de motifs de signalement proposés à l'utilisateur au total ainsi que leur arborescence et s'il existe une possibilité de saisir un motif autre que ceux proposés par la plateforme.

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

- sa présentation visuelle,
- son affichage par défaut ou non,
- son emplacement à proximité immédiate du contenu ou non,
- le nombre d'actions nécessaires pour accéder au dispositif de signalement général, et à celui de signalement d'une fausse information.

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

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



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

- avant le déclenchement de cette procédure : les **outils** (techniques, scientifiques, journalistiques, lignes directrices internes, etc.) permettant d'instruire les signalements ;
- les critères d'ouverture de l'instruction d'un signalement ;
- dès le déclenchement de la procédure :
  - l'organisation et les étapes de la procédure d'instruction, notamment s'il existe une procédure d'instruction spécifique en période électorale ou en cas d'urgence (ex. : sanitaire),
  - les **moyens humains** mis en œuvre dans le cadre de procédure, en précisant le nombre de personnes en charge de cette activité, leur formation, leur localisation ainsi que les modalités d'organisation de leur travail (notamment si elles impliquent une astreinte) et leur statut interne ou externe à la société,
  - les **moyens automatisés** mis en œuvre dans le cadre de procédure, en précisant leur nature, leur nombre et leur fonctionnement,
  - à quelles étapes et par quelle voie **l'auteur du signalement et/ou celui du contenu signalé sont tenus informés** par l'opérateur des actions que celui-ci a entreprises ou décidées ;
    - à la fin de la procédure :
  - la liste des **actions** que peut mettre en œuvre l'opérateur vis-à-vis des contenus signalés dont il a considéré à l'issue de la procédure qu'ils contenaient de fausses informations, ainsi que des comptes sur lesquels ils ont été publiés,
  - si les résultats de la procédure sont **portés à la connaissance des autres utilisateurs**, notamment par l'affichage d'un message sur le contenu signalé,
  - si la mise en œuvre de la décision de l'opérateur s'effectue par des **moyens humains et/ou automatisés**,
  - l'existence de **voies de recours** pour les auteurs du contenu signalé et du signalement.

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

- le **nombre de contenus signalés** comme fausses informations,
- le **nombre de procédures de traitement** mises en œuvre,
- le **nombre de contenus considérés comme étant de fausses** informations à l'issue de la procédure de traitement,
- le **déai 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.

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



## 2. Transparence des algorithmes

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

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

- son ou ses objectif(s),
- ses grands principes de fonctionnement, en détaillant notamment leur(s) critère(s) (définis succinctement), le poids de ces critères dans l'algorithme considéré ainsi que les données qu'ils utilisent, la façon dont elles sont collectées, leur durée de conservation, ainsi que la longueur de l'historique utilisée par l'algorithme<sup>3</sup>,
- les effets de l'algorithme, soit la manière dont les résultats sont utilisés par la plateforme en lien avec les objectifs précités,
- si cet algorithme est personnalisable par l'utilisateur et/ou s'il est ajusté aux données propres à chaque utilisateur<sup>4</sup>,
- enfin, le nombre de changements opérés dans ces algorithmes au cours de l'année d'exercice, notamment lorsqu'ils concernent la lutte contre la manipulation de l'information.

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

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

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

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

2.5. Par ailleurs, l'opérateur indiquera s'il existe un **outil d'interaction en temps réel** entre l'utilisateur et l'opérateur aux fins notamment d'obtenir des informations personnalisées et



précises sur le fonctionnement des algorithmes de recommandation et de hiérarchisation des contenus. Le Conseil invite la plateforme à lui préciser, dans l'hypothèse où un tel outil existerait :

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

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

### 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.

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

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

- **qui la réalise** (l'opérateur, les entreprises et agences de presse et SCA, ou un tiers) et, dans l'hypothèse où l'opérateur ne la réalise pas directement, **s'il la vérifie** (avant ou après publication),
- **par quels moyens et sur la base de quelles sources** elle est réalisée,
- **comment elle est portée à la connaissance de l'utilisateur** sur le service,
- les **moyens humains et technologiques déployés par l'opérateur afin d'assurer la visibilité et la mise en avant des informations provenant de ces sources** dans les résultats des moteurs de recherche, des fils d'actualité et autres espaces de diffusion de contenus opérant par classement algorithmique.

3.2. Les moyens mis en œuvre par l'opérateur pour **tenir compte des démarches de labellisation et de bonnes pratiques**, notamment celles réalisées en tout ou partie par les entreprises et agences de presse et les SCA (ex. : Trust Project, Journalism Trust Initiative), seront également portés à la connaissance du Conseil. Outre la description de ces **moyens mis en œuvre** (ex. : réalisation directe, financement, partenariats, mise en avant sur le service (et comment), informations aux utilisateurs, etc.), l'opérateur précisera, le cas échéant, **sur quels critères sont sélectionnées les démarches de labellisation** concernées.

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



3.3. Le Conseil s'interroge sur les dispositifs mis en œuvre par l'opérateur pour œuvrer au « fact-checking <sup>3</sup> ». Il souhaite connaître les **partenariats et/ou actions (en interne et externe) existants** en matière de « fact-checking » (notamment aux fins de lutter contre les hypertrucages ou « deepfakes ») et en particulier :

- les critères retenus par l'opérateur pour **sélectionner les « fact-checkers<sup>4</sup> »** (notamment : nationalité et langue de travail pour un partenariat sur une zone géographique donnée) ainsi que les **contreparties** qui leur sont offertes et dans quelles conditions,
- si **les contenus sur la plateforme pouvant être soumis aux « fact-checkers »** sont laissés à leur libre appréciation ou limités (le cas échéant, selon quels critères),
- l'ensemble des **moyens technologiques** mis à la disposition des « fact-checkers » dans leur travail de vérification de l'information (notamment : outil de suggestion des contenus à vérifier et, le cas échéant, selon quels critères ; mise à disposition d'une API).

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

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

3.5. L'opérateur est invité à décrire les **moyens technologiques et humains permettant de traiter les contenus dont le « fact-checking »** a montré qu'ils contenaient ou étaient susceptibles de contenir une fausse information, notamment **les mesures prises** le cas échéant (ex. : déréférencement) et leur gradation, ainsi que l'existence de **voies de recours** pour les auteurs des contenus concernés.

3.6. Le Conseil souhaite également connaître la communication de l'opérateur ou les moyens qu'il met en œuvre pour permettre aux « fact-checkers » de **communiquer auprès des utilisateurs ayant publié un tel contenu, l'ayant partagé ou ayant interagi avec** (ex. : commentaire, like).

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

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

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

- **détecter** de tels comptes (ainsi que ceux ayant recours aux « deepfakes »),
- **faire obstacle** à ces comptes (notamment la procédure, les mesures prises et leur gradation : avertissement, restriction des droits, mise en quarantaine, déréférencement,

<sup>3</sup> Vérification de l'information.

<sup>4</sup> Tiers vérificateurs de l'information.



suppression du contenu, du compte, etc.) et à l'accélération et à la viralité des contenus contenant de fausses informations qu'ils véhiculent.

4.2. S'agissant de **l'information aux utilisateurs sur les mesures de détection et de traitement** des comptes propageant massivement de fausses informations, l'opérateur décrira les **dispositifs de suivi et de statistiques de ces mesures**, et notamment :

- si ces dispositifs sont à usage interne ou s'ils sont **accessibles aux utilisateurs, et selon quels forme et niveau de détail**,

- les **informations** auxquelles ils donnent accès : nombre de comptes détectés, signalés, traités (par des moyens humains d'une part, et des outils technologiques d'autre part) ; caractéristiques des comptes et des contenus concernés ; actions prises ; résultats,

- l'accès et le **traitement possible de ces données par l'utilisateur**, via une fonctionnalité sur le service ou permettant de les télécharger.

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

- l'information des utilisateurs sur les **risques encourus** notamment dans les cas de création de comptes dans des volumes anormaux, de partage de contenus à des fréquences anormales et d'utilisation de renseignements faux, volés ou trompeurs,

- **la manière dont cette information est fournie** à l'utilisateur : quelle forme prend-elle, où se situe-t-elle, est-elle claire et accessible ? est-elle poussée à l'utilisateur (ou nécessite-t-elle une recherche de sa part) et, le cas échéant, dans quelle(s) situation(s) ?

4.4. Le Conseil souhaite disposer de **données chiffrées** sur les mesures de lutte prises par l'opérateur contre les comptes propageant massivement de fausses informations, notamment :

- **le nombre de comptes propageant massivement de fausses informations détectés** sur le service, et la proportion qu'ils représentent par rapport au nombre total de comptes sur le service,
- les **origines de cette détection** (moyens technologiques, signalements par des tiers, signalements en interne) et leur répartition,
- parmi les comptes détectés, le **nombre de ceux ayant fait l'objet d'un traitement** par l'opérateur, et de ceux ayant fait l'objet de mesures (en déclinant par types de mesures),
- les **moyens financiers et humains** mis en œuvre par l'opérateur pour lutter contre les comptes propageant massivement de fausses informations,
- le cas échéant, les recettes générées pour la plateforme par de tels comptes et celles qu'elle leur a reversées en 2019.

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

Les indications suivantes s'appliquent aux opérateurs qui auraient mis en œuvre des mesures relatives à l'information des utilisateurs sur la nature, l'origine et les modalités de diffusion des contenus et à l'information des utilisateurs sur l'identité de la personne physique ou la raison sociale, le siège social et l'objet social des personnes morales leur versant des rémunérations en



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

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

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

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

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

5.4. Dans la mesure où le dispositif différerait du dispositif de signalement général, le Conseil souhaite que l'opérateur lui indique si l'utilisateur **peut signaler un contenu sponsorisé pour « fausse information » (ou équivalent), et en particulier un contenu sponsorisé d'information se rattachant à un débat d'intérêt général**. Le cas échéant, il souhaite également connaître les modalités de signalement (faut-il avoir un compte pour signaler le contenu, quelle est la démarche pour signaler ces contenus...) et dans quelles conditions et comment l'opérateur interagit avec les utilisateurs concernés par ce signalement (auteur du contenu et auteur du signalement ; information concernant l'avancée du traitement du signalement). L'opérateur pourra ici préciser le nombre de signalements reçus et traités relatifs à des contenus sponsorisés et, plus particulièrement, aux contenus sponsorisés d'information se rattachant à un débat d'intérêt général, en indiquant, quand cela est possible, le nombre de signalements pour des raisons de « fausses informations ». Enfin, il pourra détailler les moyens (technologiques, humains et financiers) mis en œuvre pour traiter ces signalements.



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

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

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

- si cette base est publique,
- les types de données qu'elle contient,
- si celles-ci peuvent être filtrées spécifiquement pour la France,
- si l'utilisateur, via une fonctionnalité sur le service ou permettant de les télécharger, peut y avoir accès.

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

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

6.1. L'opérateur est invité à déclarer au Conseil s'il a développé ou mis en place ;

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

6.2. L'opérateur **décrira en quelques lignes chaque outil, action ou soutien**, en s'attachant notamment à indiquer : sa nature, son objectif, en quoi il/elle consiste, sa date de lancement, son accessibilité et sa mise en valeur auprès du public et des utilisateurs, le ou les public(s) visé(s) et l'impact de son action sur ces derniers (notamment en termes quantitatifs, ex. : nombre de personnes effectivement touchées), les partenaires éventuels, la nature de sa propre implication et le montant des dépenses qu'il a engagées et/ou des recettes qu'il a tirées pour chacune de ces opérations. Il pourra fournir toute présentation illustrée de ces différentes mesures.



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

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

En particulier, le Conseil souhaite savoir si l'opérateur a conclu des partenariats avec le monde de la recherche pour mieux cerner certains comportements et phénomènes liés à la diffusion de fausses informations en ligne et **les moyens** qu'il met à leur disposition (ex. : concernant la propagation de fausses informations, l'opérateur donne-t-il accès à des données - notamment via une API -, et si oui, sous quelles conditions cet accès est-il possible ?).

## 7. Remarques générales

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

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

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

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



## Annexe 3 : Composition du comité d'experts sur la désinformation en ligne

- Christine Balagué, professeure à l'Institut Mines-Télécom Business School, titulaire de la Chaire Good in Tech ;
- Valérie-Laure Benabou, professeure de droit privé à Aix-Marseille Université ;
- Julia Cagé, Assistant Professor en économie à Sciences Po ;
- Yves Caseau, directeur des systèmes d'information du Groupe Michelin, membre de l'Académie des technologies ;
- Lucien Castex, secrétaire général d'Internet Society France ;
- Raphaël de Andréis, président-directeur général d'Havas Village France, président des agences créatives Europe du sud chez Havas ;
- Christophe Deloire, secrétaire général de Reporters sans frontières ;
- Catherine Emprin, directrice générale de BETC, membre du comité de Paris de Human Rights Watch ;
- Frédéric Filloux, président-directeur général de Deepnews.ai ;
- Divina Frau-Meigs, professeure en sciences de l'information et de la communication à l'Université Sorbonne nouvelle - Paris III, experte auprès de l'Unesco, du Conseil de l'Europe et de la Commission européenne ;
- Marie-Anne Frison-Roche, professeure de droit économique à Sciences Po ;
- Boris Jamet-Fournier, responsable du département recherche et innovation de l'Institut national de l'audiovisuel ;
- Lucas Menget, directeur-adjoint de la rédaction de France Info ;
- Albert Moukheiber, docteur en neurosciences cognitives, psychologue clinicien, chargé de cours à l'université Paris 8 Vincennes - Saint-Denis ;
- Bruno Patino, directeur éditorial d'Arte France, directeur de l'Ecole de journalisme de Sciences Po (*membre du Comité jusqu'à juin 2020*) ;
- Cécilia Ragueneau, directrice des marques et du développement de Radio France ;
- Pauline Talagrand, adjointe au chef des réseaux sociaux et du fact-checking de l'AFP ;
- Sophie Viger, développeuse, directrice générale de l'Ecole 42.