

Lutte contre la manipulation de l'information sur les plateformes en ligne

Bilan annuel des moyens
et mesures mis en oeuvre
par les opérateurs de plateforme
en ligne en 2021

Sommaire

Synthèse.....	3
Introduction	8
I. Les plateformes déclarantes et leur approche en matière de lutte contre la manipulation de l'information	10
1. Présentation des plateformes au vu des déclarations reçues	10
2. Approches générales en matière de lutte contre la manipulation.....	11
3. Une transparence qui demeure insuffisante	14
II. Analyse des moyens déployés par les plateformes pour lutter contre la manipulation de l'information en 2021	17
1. Dispositif de signalement de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un scrutin	17
2. Transparence des algorithmes.....	27
3. Promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle.....	35
4. Lutte contre les comptes propageant massivement de fausses informations.....	41
5. Mesures de lutte contre les fausses informations en matière de communications commerciales et de promotion des contenus d'information se rattachant à un débat d'intérêt général.....	49
6. Éducation aux médias et à l'information et relations avec le monde de la recherche	56
7. Mesures complémentaires prises face à des situations spécifiques	61
Synthèse des préconisations de l'Arcom.....	67
Annexe – Points-clés par plateformes	70

Synthèse

La loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information impose aux opérateurs de plateforme en ligne un devoir de coopération avec l'Arcom, des obligations de moyens et une transparence à l'égard de leurs utilisateurs.

Ces obligations de moyens incluent notamment : la mise en place d'un dispositif de signalement, une meilleure transparence des algorithmes, la promotion de contenus de issus de sources vérifiées, la lutte contre les comptes participant activement à la manipulation de l'information, une transparence renforcée sur certaines publicités et des actions d'éducation aux médias et à l'information.

Les opérateurs doivent rendre publics les moyens et mesures qu'ils y consacrent dans une déclaration annuelle communiquée à l'Arcom. Le régulateur est chargé de publier un bilan périodique de l'application et de l'effectivité des mesures prises par les opérateurs.

Ce rapport sur les moyens et mesures mis en œuvre en 2021 vient compléter le bilan, très satisfaisant, de la coopération avec les plateformes durant les deux périodes électorales dont le rapport a été publié par l'Arcom le 2 novembre 2022¹.

1. Remarques générales sur le respect des obligations de moyens

Douze opérateurs de plateforme en ligne soumis au devoir de coopération prévu par le titre III de la loi du 22 décembre 2018 se sont livrés à l'exercice de déclaration à l'Arcom pour l'exercice 2021 : *Dailymotion*, *Google*, *LinkedIn*, *Meta*, *Microsoft*, *Pinterest*, *Snap*, *TikTok*, *Twitter*, *Webedia*, *la Fondation Wikimédia* et *Yahoo*.

De manière générale, l'Arcom observe une amélioration de la quantité d'informations déclarées, conséquence en particulier de l'intensification de la coopération entre les plateformes et le régulateur par la mise en place d'un dialogue riche et constant. Néanmoins, ce constat doit être nuancé par un niveau de transparence très hétérogène dans l'exercice de déclaration. *TikTok* (soumis pour la première année à ces obligations), *Yahoo* (*a priori* soumis pour la dernière année à ces obligations) et, dans une moindre mesure, *Google*, se démarquent particulièrement par l'absence d'informations tangibles permettant à l'Arcom d'analyser la pertinence et l'efficacité des moyens et mesures mis en œuvre pour lutter contre la manipulation de l'information.

Les opérateurs se soumettent désormais de façon globalement satisfaisante à l'obligation de mise en place d'un outil de signalement des fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un des principaux scrutins. L'accessibilité et la visibilité de cet outil demeurent néanmoins inégales d'une plateforme à l'autre, en particulier sur les moteurs de recherche. De plus, l'Arcom déplore un manque flagrant d'indicateurs et de données chiffrées relatives aux signalements et à l'efficacité de leur traitement, ce qui empêche une évaluation de la pertinence et l'efficacité des mesures au regard des risques associés à chacun des services.

Une certaine amélioration est également relevée s'agissant des informations sur les critères utilisés pour la recommandation de contenus, notamment ceux entraînant une réduction de la visibilité des contenus considérés comme étant de faible qualité. Mais là aussi, l'absence de données chiffrées ne permet pas au régulateur d'évaluer la pertinence des outils, notamment les systèmes de modération automatisés. Une attention particulière a

¹ <https://www.arcom.fr/nos-ressources/etudes-et-donnees/mediatheque/rapport-sur-les-campagnes-electorales-2022-election-la-presidence-de-la-republique-et-elections-legislatives>

par ailleurs été portée cette année aux processus d'évaluation des systèmes algorithmiques, internes ou externes, mis en œuvre par les opérateurs.

Concernant l'identification et la promotion de sources fiables, les mesures déclarées sont globalement identiques à celles déclarées lors du précédent exercice chez la majorité des répondants. Dans un contexte marqué par une instabilité géopolitique forte, l'analyse des moyens et mesures mis en œuvre pour l'identification des contenus issus de sources vérifiées s'est focalisée sur l'identification des médias ne bénéficiant pas des garanties d'indépendance éditoriale nécessaires vis-à-vis de l'État qui les contrôle.

L'Arcom remarque plus particulièrement un défaut de transparence quant aux pratiques identifiées par les opérateurs pouvant porter atteinte à l'intégrité de services, ainsi qu'aux moyens et outils qu'ils mettent en œuvre pour y répondre. Elle rappelle à cet égard que les opérateurs moins vertueux sont ceux qui restent les plus opaques, se gardant ainsi de toute critique autre que celle, fondamentale, de leur manque de coopération et de transparence.

Sur les mesures de lutte contre la manipulation en matière de communications commerciales, un manque d'informations est particulièrement remarqué concernant les modèles de vente et d'enchères et le ciblage publicitaire.

Enfin, en matière d'éducation aux médias et à l'information, l'Arcom note des disparités dans la quantité d'actions engagées par les opérateurs en direction des utilisateurs, notamment concernant les risques spécifiques à leurs propres services. Si des améliorations sont à noter concernant les coopérations avec le monde de la recherche, l'accès de ce dernier aux données des plateformes demeure toujours très limité. Comme l'a rappelé l'Arcom lors du lancement de la consultation publique sur l'accès aux données des plateformes, il est essentiel que soit mis en œuvre le cadre prévu par le DSA permettant l'étude des comportements en ligne et les risques systémiques induits. Cette évolution contribuera de façon décisive à la protection et au renforcement de l'indépendance, de l'autonomie et de la capacité d'analyse propres des chercheurs.

L'Arcom déplore les lacunes répétées dans la communication par certains opérateurs d'informations et de données chiffrées. Les justifications fournies interrogent sur leur volonté et leur capacité à évaluer leurs propres politiques en matière de lutte contre la manipulation de l'information.

La mise en œuvre de cette régulation montre ainsi les limites d'un cadre essentiellement souple qui ne permet pas d'exiger des plateformes qu'elles rendent des comptes, d'une part, de la réalité des phénomènes de manipulation de l'information sur leur service et, d'autre part, de l'ensemble des mesures prises pour les atténuer. En effet, si la loi du 22 décembre 2018 a instauré un cadre précurseur fondé sur des obligations de moyens et de transparence qui incombent aux opérateurs de plateforme en ligne, les efforts de transparence apparaissent toujours très insuffisants et ce d'autant plus que les opérateurs restent muets sur des enjeux essentiels.

Toutefois, l'évolution majeure du cadre de régulation avec l'adoption du règlement européen sur les services numériques (« *Digital Services Act* » ou DSA) va changer la donne. Avec ses obligations « en dur » et leur corollaire, un pouvoir de sanction, ce texte ambitieux et novateur est destiné à mieux répondre encore aux risques soulevés par les plateformes en ligne. Ce sera le cas en particulier en matière de protection des publics et de stabilité de nos sociétés démocratiques, dans le strict respect de la liberté d'expression et des droits fondamentaux garantis en Union européenne. Les dispositions du Code européen renforcé de bonnes pratiques contre la désinformation du 16 juin 2022 ont ainsi vocation à compléter la palette d'outils d'atténuation des risques dont la régulation va pouvoir se saisir dans le cadre du DSA. Les différents engagements pris par ses signataires, parmi lesquels figurent la majorité des opérateurs concernés en France par le titre III de

la loi du 22 décembre 2018, et les obligations de publier des rapports de transparence qui en découlent, permettront de nourrir les réflexions de l'Autorité dans une optique de coordonner et d'harmoniser le suivi et l'analyse des moyens de lutte contre la manipulation de l'information aux différentes échelles, européenne et nationale.

Le DSA instaure ainsi un cadre de régulation rénové fondé sur une imbrication entre échelons national et européen, qui appelle une forte coordination entre les régulateurs nationaux et la Commission européenne. À cet égard, l'accès des régulateurs aux données et aux informations des plateformes est essentiel, tout comme celui, désormais consacré par le DSA, du monde académique aux données nécessaires à l'évaluation des risques et des mesures prises pour les réduire.

2. Synthèse des préconisations de l'Arcom aux opérateurs

Sur la transparence en général

- ❖ **Préconisation n° 1 :** faire preuve de plus de transparence vis-à-vis du public en fournissant davantage d'indicateurs chiffrés aussi bien sur les contenus, pratiques et phénomènes de manipulation de l'information identifiés que sur les moyens et mesures mises en œuvre pour lutter contre ceux-ci.
- ❖ **Préconisation n° 2 :** mettre en place davantage d'initiatives pour améliorer la transparence vis-à-vis de la société civile et de la recherche, notamment en matière d'accessibilité des données, afin de permettre une évaluation indépendante et plurielle des risques et des mesures déployées pour les contrer.
- ❖ **Préconisation n° 3 :** communiquer au régulateur davantage d'informations, de manière confidentielle si nécessaire, pour permettre une meilleure compréhension des moyens et mesures mis en œuvre pour lutter contre la manipulation de l'information.

Sur le dispositif de signalement

- ❖ **Préconisation n° 4 :** améliorer la visibilité et la facilité d'utilisation du dispositif de signalement en le plaçant à proximité immédiate du contenu (notamment pour les moteurs de recherche [Google Search](#) et [Bing](#)), et ce, sur toutes les versions du service (particulièrement pour les versions mobiles de [TikTok](#) et [YouTube](#)).
- ❖ **Préconisation n° 5 :** proposer, au sein des dispositifs de signalement, des motifs ou sous-motifs relatifs à la manipulation de l'information (ou notion équivalente) clairs et suffisamment précis (notamment pour [Twitter](#), [Jeuxvideo.com](#), [LinkedIn](#) et [YouTube](#)). Ces motifs doivent également couvrir la manipulation de l'information susceptible de troubler l'ordre public ou de nuire au bon déroulement des processus électoraux.
- ❖ **Préconisation n° 6 :** mieux informer les auteurs de signalement et utilisateurs ayant publié un contenu signalé de l'avancée des procédures de traitement des signalements en cours et leur en communiquer l'issue dans un délai raisonnable et de façon motivée.
- ❖ **Préconisation n° 7 :** permettre aux auteurs de signalement et aux utilisateurs affectés par une décision de modération de former un recours et leur fournir une information claire sur leurs possibilités de recours au moment de la notification de la décision prise.

Sur la transparence des algorithmes

- ❖ **Préconisation n° 8 :** améliorer la transparence des processus et moyens dédiés à l'évaluation des outils automatiques en publiant régulièrement, dans un espace dédié, des éléments sur les modifications qui ont été réalisées à l'aide d'analyses internes ou externes.
- ❖ **Préconisation n° 9 :** communiquer aux utilisateurs les critères utilisés pour la recommandation de contenus, et ce, de manière personnalisée et contextuelle à l'aide d'outils et fonctionnalités accessibles directement sur le service.
- ❖ **Préconisation n° 10 :** instaurer une transparence accrue des politiques en matière de réduction de la visibilité des contenus.
- ❖ **Préconisation n° 11 :** faire preuve d'une transparence plus importante sur le fonctionnement des outils de détection de contenus, comptes et comportements utilisés à des fins de modération vis-à-vis du public et communiquer au régulateur davantage d'éléments permettant d'en évaluer la pertinence.

Sur la promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle

- ❖ **Préconisation n° 12 :** mettre en place des dispositifs spécifiques d'identification des médias qui ne bénéficient pas des garanties d'indépendance éditoriale nécessaires vis-à-vis de l'État qui les contrôle et les appliquer de façon homogène, transparente et selon des critères objectifs.
- ❖ **Préconisation n° 13 :** mettre en place des dispositifs permettant d'ajouter des « frictions » dans la diffusion de contenus potentiellement problématiques.
- ❖ **Préconisation n° 14 :** développer le recours aux dispositifs (i) de *fact-checking* et (ii) de mise en avant des contenus en résultant, y compris hors période électorale.

Sur la lutte contre les comptes propageant massivement de fausses informations

- ❖ **Préconisation n° 15 :** établir une liste claire de l'ensemble des pratiques portant atteinte à l'intégrité de leur service afin de détailler plus clairement, avec des indicateurs chiffrés, les moyens et mesures mis en œuvre pour y répondre. Pour ce faire, les opérateurs sont encouragés à échanger entre eux sur les types de pratiques identifiées et à se nourrir de la typologie commune qui sera élaborée par les signataires du code européen de bonnes pratiques renforcé contre la désinformation.
- ❖ **Préconisation n° 16 :** créer un espace unique et facilement accessible sur le service listant les interdictions en matière de pratiques portant atteinte à l'intégrité des services, celles identifiées et les décisions prises en conséquence.
- ❖ **Préconisation n° 17 :** travailler à la mise en œuvre de fonctionnalités d'information des utilisateurs victimes de certaines pratiques de manipulation, de manière individualisée et contextuelle.

Sur les mesures de lutte contre les fausses informations en matière de communications commerciales et de promotion de contenus d'information se rattachant à un débat d'intérêt général

- ❖ **Préconisation n° 18 :** l'ensemble des modèles d'enchères devraient faire l'objet de mesures de transparence renforcées, tel qu'un audit par des tiers indépendants, en particulier lorsqu'ils présentent des métriques propriétaires basées sur la performance qui pourraient être exploitées dans un but de manipulation de l'information.
- ❖ **Préconisation n° 19 :** dans la perspective d'une meilleure transparence des algorithmes de recommandation publicitaire, les dispositifs d'information contextuels devraient comprendre des informations sur le ciblage (de manière individualisée) et la gestion des paramètres et servir de lien vers d'autres outils d'information plus complets comme les bibliothèques publicitaires.
- ❖ **Préconisation n° 20 :** intensifier les efforts de vérification des liens de redirection encapsulés dans les communications commerciales, qui peuvent rediriger vers des sites tiers diffusant de la désinformation.
- ❖ **Préconisation n° 21 :** l'Arcom rappelle en outre que la présence de désinformation dans les communications commerciales doit pouvoir être signalée par l'utilisateur via un dispositif facile d'accès et visible.

Sur l'éducation aux médias et à l'information ainsi que sur les relations avec le monde de la recherche

- ❖ **Préconisation n° 22 :** mettre en place des actions d'éducation aux médias et à l'information sur les risques en matière de manipulation de l'information liés à l'utilisation de leur service et proposer des fonctionnalités permettant aux utilisateurs de prendre en compte ces risques lors de cette utilisation.
- ❖ **Préconisation n° 23 :** développer davantage la mesure de l'impact réel des actions d'EMI des opérateurs sur les utilisateurs et en rendre compte.
- ❖ **Préconisation n° 24 :** coopérer davantage avec la société civile à des initiatives communes permettant une meilleure compréhension des phénomènes en ligne, et plus particulièrement avec les organismes de recherche sur une meilleure accessibilité des données.

Introduction

L'Arcom publie son troisième bilan annuel des moyens et mesures mis en œuvre par les opérateurs de plateforme en ligne en France pour lutter contre la manipulation de l'information, portant sur l'exercice 2021.

Il intervient dans un contexte marqué par une évolution majeure du cadre de régulation avec l'adoption du règlement européen sur les services numériques (« Digital Services Act », ci-après DSA)². Ce texte ambitieux et novateur est destiné à mieux répondre encore aux risques soulevés par les plateformes en ligne en matière de protection des publics et de stabilité de nos sociétés démocratiques, dans le strict respect de la liberté d'expression. Il est fondé sur une plus grande responsabilisation des opérateurs et une exigence renforcée de transparence. Les dispositions du Code européen renforcé de bonnes pratiques contre la désinformation du 16 juin 2022 (ci-après code européen de bonnes pratiques renforcé contre la désinformation), dont sont signataires la majorité des opérateurs concernés par le présent bilan, ont vocation à s'insérer dans les outils d'atténuation des risques systémiques dans le cadre de ce règlement.

L'actualité récente est également marquée par l'annonce du rachat de *Twitter* fin octobre 2022, suivie d'une très forte réduction (près de la moitié) du nombre d'employés. S'il est encore trop tôt pour mesurer les conséquences de cette décision sur l'engagement de l'opérateur en matière de lutte contre la manipulation de l'information et sur l'efficacité des moyens mis en œuvre à cette fin, l'Arcom en appelle à sa responsabilité pour poursuivre les efforts et mettre en œuvre une modération protectrice des droits des utilisateurs et respectueuse de leur liberté d'expression.

Début novembre, *Meta* annonçait également une réduction de près de 13 % du nombre d'employés et un gel des recrutements jusqu'en fin mars 2023. L'Autorité souhaite que cette annonce ne soit pas de nature à remettre en cause la mobilisation de l'opérateur sur les questions de lutte contre la manipulation de l'information.

En mai 2022, pour la troisième année consécutive, l'Autorité a adressé aux opérateurs de plateforme en ligne concernés par le titre III de la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information un questionnaire dans la perspective de son bilan. Elle les y interrogeait sur les enseignements tirés de la mise en œuvre de la loi et les éventuelles difficultés rencontrées en la matière. Des questions spécifiques ont par ailleurs été posées à chaque opérateur pour mieux prendre en compte les spécificités de leur service et les interroger sur des actualités particulières survenues l'année passée. Enfin, le questionnaire tenait compte des observations réalisées pendant la période électorale et des échanges que l'Arcom a eus avec les opérateurs au cours du premier semestre 2022.

Douze opérateurs de plateforme en ligne soumis au devoir de coopération prévu par la loi précitée se sont livrés à l'exercice de déclaration à l'Arcom : *Dailymotion*, *Google*, *LinkedIn*, *Meta*, *Microsoft*, *Pinterest*, *Snap*, *TikTok*, *Twitter*, *Webmedia*, la *Fondation Wikimédia* et *Yahoo*.

Dans un objectif de transparence et d'accès du public à l'information, les déclarations des opérateurs ont été publiées par l'Arcom sur son site internet³ dès leur réception en juillet 2022, en amont de leur instruction et de la publication du présent bilan. L'Arcom invite les lecteurs à se référer à ces déclarations ainsi qu'à ses deux précédents bilans.

² Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques)

³ <https://www.arcom.fr/vos-services-par-media/internet-et-reseaux-sociaux/lutte-contre-la-manipulation-de-l-information-declarations-des-operateurs-de-plateformes-en-ligne-et-questionnaires-de-larcom>

Le présent bilan porte une attention particulière au suivi des préconisations édictées l'année passée et aux moyens et mesures mis en place en conséquence. L'Arcom formule 24 préconisations renforcées, issues des éléments d'analyse de ce bilan comparé.

Enfin, l'Arcom souligne que l'analyse critique et circonstanciée qu'il conduit porte nécessairement sur les moyens et mesures portées à sa connaissance par les plateformes. Les opérateurs moins vertueux, dont les déclarations sont lacunaires ou superficielles, échappe ainsi à la critique autre que celle, fondamentale, de leur manque de coopération et de transparence.

I. Les plateformes déclarantes et leur approche en matière de lutte contre la manipulation de l'information

1. Présentation des plateformes au vu des déclarations reçues

1.1 Opérateurs déclarants pour l'exercice 2021

Dix opérateurs de plateforme en ligne soumis au devoir de coopération prévu par la loi du 22 décembre 2018 se sont livrés, pour la troisième année consécutive, à l'exercice de déclaration au régulateur des moyens et mesures mis en œuvre : [Dailymotion](#), [Google](#), [LinkedIn](#), [Meta](#), [Microsoft](#), [Snap](#), [Twitter](#), [Webedia](#), la [Fondation Wikimédia](#) et [Yahoo](#).

S'ajoutent [Pinterest](#) et [TikTok](#) qui ont remis une déclaration annuelle à l'Arcom pour la première fois.

1.2 Spectre d'activité des plateformes

Les services déclarés par ces douze opérateurs répondent aux deux types d'activités visées par la définition légale des plateformes en ligne⁴ : classement ou référencement de contenus, biens ou services mis en ligne par des tiers ([Bing](#), [Google Search](#), [Yahoo Media](#)) ou 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 ([Dailymotion](#), [YouTube](#), [LinkedIn](#), [Facebook](#), [Instagram](#), [Pinterest](#), [Snapchat](#), [TikTok](#), [Twitter](#), [Jeuxvideo.com](#) et [Wikipédia](#)).

Ces plateformes ont un spectre d'activité large : réseaux sociaux généralistes ([Snapchat](#), [Instagram](#), [Facebook](#), [TikTok](#), [Pinterest](#)) ou spécialisés ([LinkedIn](#)), plateformes de partage de vidéos ou d'audios différé ou en direct ([YouTube](#), [Dailymotion](#), [Twitch](#)), forums en ligne ([Jeuxvideo.com](#)), moteurs de recherche ([Bing](#), [Google Search](#), [Yahoo Media](#)) et encyclopédies participatives ([Wikipédia](#)).

1.3 Audience des plateformes

Plusieurs plateformes ([Google](#), [LinkedIn](#), [Microsoft](#) et [TikTok](#)) n'ont pas communiqué les audiences de leurs services. Certains opérateurs ([Meta](#), [Pinterest](#), [Yahoo Media](#)) l'ont fait à titre confidentiel. [Twitter](#) déclare évaluer son audience uniquement au niveau mondial et donne un chiffre fondé sur une mesure qui ne correspond pas à celle prévue réglementairement.⁵ La [Fondation Wikimédia](#) déclare 94 millions de visiteurs uniques mensuels pour [Wikipédia](#) en français, mais ces derniers ne sont pas nécessairement situés en France.

Seuls ont été déclarés de façon publique les chiffres relatifs à [Dailymotion](#) (44 millions de visiteurs uniques en moyenne mensuelle en France en 2020), [Snapchat](#) (21,9 millions) et [Jeuxvideo.com](#) (5,7 millions).

L'Arcom rappelle que la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information ne permet pas au régulateur d'exiger la collecte de ce type de données chiffrées, malgré leur importance pour évaluer la pertinence des moyens et mesures mis en œuvre au regard des audiences respectives des plateformes.

⁴ Article 111-7 du Code de la consommation.

⁵ L'opérateur ne déclare pas le nombre de visiteurs uniques mensuels mais le nombre d'utilisateurs actifs quotidiens monétisables au quatrième trimestre de 2020 (192 millions dans le monde).

1.4 Modèle économique, chiffre d'affaires et effectifs des opérateurs

Le modèle économique des services faisant l'objet d'une déclaration est basé pour tout ou partie sur les revenus publicitaires, à l'exception de la [Fondation Wikimédia](#) qui déclare être financée par les dons. La fondation rappelle par ailleurs qu'elle est établie exclusivement aux Etats-Unis et qu'elle ne dispose ainsi d'aucun établissement ni d'aucun salarié en France.

La majorité des opérateurs ont déclaré le chiffre d'affaires du service en France de manière confidentielle ([Dailymotion](#), [Yahoo](#), [Webedia](#), [Snap](#), [LinkedIn](#), [Pinterest](#)). Seuls deux opérateurs ont fourni une information publique : 11,8 millions d'euros pour [Twitter](#) et 879,3 millions d'euros pour la filiale française de [Meta](#)⁶ (ce dernier chiffre ne répondant pas exactement à la demande de l'Arcom qui souhaitait connaître le chiffre d'affaires du service en France et non de la filiale). [Google](#) a uniquement déclaré le chiffre d'affaires mondial (257,6 milliards \$ dont 15,1 milliards liés à la publicité sur [YouTube](#)).

En termes d'effectifs, 56 personnes composent les équipes éditoriales et produits de [Jeuxvideo.com](#), qui s'ajoutent aux équipes techniques, commerciales et fonctions supports mutualisées au sein du groupe [Webedia](#). [Meta](#) déclare employer 292 personnes (« effectif moyen »), [Pinterest](#) 48 équivalents temps plein (ETP) et [Twitter](#) 34 ETP. [Google](#) déclare uniquement le nombre d'employés au niveau mondial (156 500, incluant ceux de [YouTube](#)). Ces données reflètent les différences de modèle des services, en particulier le degré de recours à des entités extérieures (pour la modération et la vérification des faits (*fact-checking*), par exemple).

Cette année encore, plusieurs opérateurs ([Dailymotion](#), [Yahoo Media](#), [Snap](#), [LinkedIn](#)) ont déclaré les effectifs employés aux activités du service en France de manière confidentielle.

[TikTok](#) et [Microsoft](#) n'ont déclaré ni le chiffre d'affaires du service en France, ni les effectifs employés aux activités du service en France.

L'Arcom rappelle que les données relatives au nombre de visiteurs uniques, au chiffre d'affaires et aux effectifs des opérateurs sont nécessaires pour appréhender les spécificités des différents services. L'Arcom demande aux opérateurs de fournir des données complètes et publiques afin d'améliorer la transparence de l'information accessible à tous les publics.

Elle rappelle par ailleurs que les informations recensées ci-dessus ont été collectées sur une base uniquement déclarative, la loi du 22 décembre 2018 n'ayant pas donné au régulateur les moyens nécessaires pour exiger des précisions permettant de vérifier les informations déclarées.

2. Approches générales en matière de lutte contre la manipulation

2.1 Notion de manipulation de l'information des opérateurs

La loi du 22 décembre 2018 impose aux opérateurs de mettre en œuvre des mesures en vue de lutter contre la manipulation de l'information. L'article 11 de la loi caractérise le type de phénomènes contre lequel les plateformes doivent lutter : 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* ». L'Arcom relève pourtant que les opérateurs sont loin de recourir à une définition uniformisée de la manipulation de l'information (voir tableau ci-dessous).

⁶ Source vers laquelle renvoie l'opérateur dans sa déclaration : [Infogreffe.fr](#).

Cela peut notamment s'expliquer par les différents modèles des plateformes concernées, certains opérateurs se concentrant davantage sur la lutte contre des pratiques ou phénomènes pouvant porter atteinte à l'intégrité de leur service que sur les contenus en eux-mêmes.

L'Arcom reconnaît donc qu'il est malaisé de retenir une définition unique de la manipulation de l'information, d'autant plus que toute information fausse n'est pas illégale et que la publication d'une information fausse n'est pas nécessairement faite dans l'intention de tromper. Elle estime en outre que la lutte contre ces phénomènes doit prendre en considération les pratiques et les comptes problématiques et pas uniquement les contenus diffusés sur les plateformes en ligne. À cet égard, les notions retenues par les opérateurs (voir tableau ci-dessous) témoignent des efforts faits par certains d'entre eux pour tenir compte de ces nuances.

Pour autant, elle encourage à la constitution de notions communes et aisément compréhensibles et leur inscription dans les politiques générales des services, comme le prévoit le Code européen de bonnes pratiques renforcé contre la désinformation.

Opérateur	Notion(s) retenue(s)	Définition de la notion ou des notions retenue(s) par l'opérateur	Espace contenant les informations relatives aux interdictions
Dailymotion	Désinformation	« Nous considérons la désinformation comme englobant toutes informations dont le caractère faux ou trompeur est vérifiable. » ⁷	Politique relative aux contenus prohibés
Google	Désinformation	« Efforts délibérés pour mentir et tromper en utilisant la vitesse, la portée et les technologies d'internet. »	Non indiqué
LinkedIn	Fausse informations	« Affirmations spécifiques, présentées comme des faits, qui sont manifestement fausses ou essentiellement trompeuses. »	Politiques de la communauté professionnelle
Meta	Fausse informations	« Tout contenu faux ou trompeur. »	Centre de transparence
Pinterest	Mésinformation, désinformation ou information malveillante	« Contenu erroné ou trompeur susceptible de nuire au bien-être, à la sécurité ou à la confiance des utilisateurs ou du public. » ⁸	Guide pour la communauté
Snap	Fausse informations	« Usurpation d'identité, pratiques trompeuses et fausses informations. »	Règles communautaires
TikTok	Désinformation	« Contenu faux ou mensonger. »	Règles communautaires
Twitter	Informations trompeuses	« Informations trompeuses - déclarations ou affirmations dont le caractère faux ou trompeur a été démontré par des experts en la matière, tels que des autorités de santé publique.	Politique en matière d'intégrité civique Politique en matière d'informations trompeuses sur le COVID-19

⁷ [Dailymotion](#) ajoute deux conditions pour qualifier les contenus de désinformation : « Elles doivent être créées, présentées et diffusées dans un but lucratif ou dans l'intention délibérée de tromper le public ; et elles sont susceptibles de causer un préjudice public, au sens de menaces vis-à-vis des processus démocratiques (notamment électoraux) et dans l'élaboration de politiques démocratiques et/ou un préjudice aux biens publics, tels que la protection de la santé, de l'environnement ou la sécurité des personnes. »

⁸ [Pinterest](#) ne donne pas de définition mais fait état d'une liste de types de contenus : « les allégations concernant la santé non étayées par des preuves médicales, telles que les conseils anti-vaccination ou les faux traitements ; les contenus audios ou visuels manipulés, également désignés sous le vocable « deepfakes » ; des informations erronées ou trompeuses qui nuisent à l'intégrité des élections, qui mettraient en cause la légitimité des résultats d'élections ou entraveraient la participation à la vie civique ; ou encore la désinformation climatique. »

	Affirmations contestées	<i>Affirmations contestées - déclarations ou affirmations dont l'exactitude, la véracité ou la crédibilité sont contestées ou inconnues.</i>	
	Affirmations non vérifiées	<i>Affirmations non vérifiées - informations (qui peuvent être vraies ou fausses) qui ne sont pas confirmées au moment où elles sont partagées. »</i>	
Webedia	Fausse information	Définition donnée par le Conseil Constitutionnel : « la notion de fausse information doit s'entendre comme visant des allégations ou imputations inexacts ou trompeuses d'un fait telles que définies au paragraphe 21. » ⁹	Charte des forums
Fondation Wikimédia	Désinformation	« Informations fausses, inexacts ou trompeuses partagées avec l'intention probable de causer un préjudice. »	Non indiqué
Microsoft	Pas de définition ¹⁰		
Yahoo	Pas de définition		

2.2 Respect de la liberté d'expression

Plusieurs opérateurs ont fait état de la manière dont ils envisageaient l'articulation entre l'objectif de lutte contre la manipulation et le respect de la liberté d'expression.

À cet égard, [Microsoft](#) déclare que le blocage de contenus dans les résultats de recherche peut soulever d'importants problèmes en matière de droits fondamentaux, ce qui conduit [Bing](#) à ne supprimer que les contenus manifestement illégaux. [Webedia](#) indique être peu concerné et qu'un *topic* du forum ne saurait être supprimé que s'il constituait un trouble grave à l'ordre public.

[Snap](#) indique évaluer le risque de préjudice causé par un contenu pour garantir le caractère nécessaire et proportionné de sa suppression. [LinkedIn](#) affirme se concentrer sur les fausses informations qui menacent la sécurité de ses membres et accorder « *le bénéfice du doute* » lorsqu'un contenu n'enfreint pas de manière irréfutable ses politiques. [Dailymotion](#) déclare que pour prévenir le risque de retrait de contenus illégitimes, il ne modère pas « *les erreurs de citation, la satire, la parodie, les informations et commentaires partisans clairement identifiés, les contenus éducatifs, scientifiques ou journalistiques contextualisant et explicitant des informations fausses ou trompeuses.* »

[Meta](#) déclare se concentrer sur les fausses informations qui pourraient causer un « *dommage physique imminent* » et précise ne pas vouloir se faire « *arbitre de la vérité* ». C'est pourquoi, face à un contenu jugé faux par ses partenaires de vérification indépendants, il préfère en réduire la diffusion dans le fil d'actualité, afin de permettre un juste équilibre entre, d'une part, un service qui soit un « *lieu d'expression mis à la disposition des utilisateurs et, d'autre part, une communauté sûre et authentique* ».

[Twitter](#) indique que son rôle est de servir la conversation publique et préfère ainsi favoriser le contre-discours face à des déclarations inexacts ou des perceptions erronées.

⁹ CC, décision n° 2018-773 DC du 20 décembre 2018,

¹⁰ [Microsoft](#) déclare ne pas supprimer de contenu sauf situation exceptionnelle pour le contenu jugé illégal en vertu de la législation locale. L'opérateur indique ainsi ne pas supprimer les « fausses informations » de ses résultats de recherche car il considère que des utilisateurs pourraient souhaiter avoir accès à des contenus de mauvaise qualité (à des fins de recherche ou de journalisme) et ne propose ainsi pas de définition.

Enfin, [TikTok](#) apporte très peu d'informations en indiquant uniquement chercher un « *juste équilibre permettant de limiter la diffusion d'informations trompeuses* ». [Pinterest](#), [Google](#), la [Fondation Wikimédia](#) et [Yahoo](#) ne déclarent aucun élément sur ce volet.

L'Arcom constate que les opérateurs n'ont pas suffisamment fait état, au sein de l'ensemble de leur déclaration, des éventuelles atteintes à la liberté d'expression liées à une modération automatique accrue, comme il avait été constaté l'an passé.

3. Une transparence qui demeure insuffisante

L'Arcom rappelle que la transparence des opérateurs sur les mesures mises en œuvre pour lutter contre la manipulation de l'information et les moyens qu'ils y consacrent est une obligation légale, inscrite à l'article 11 de la loi du 22 décembre 2018¹¹.

Dans son précédent bilan, l'Arcom avait indiqué qu'elle appréhendait la transparence selon trois modalités :

- l'information fournie par l'opérateur à l'ensemble des utilisateurs de son service sur les moyens mis en œuvre pour lutter contre la manipulation de l'information, qui doit être claire, concise et intelligible, et aisément accessible ;
- la mise à disposition des citoyens et de l'ensemble de la société civile (journalistes, associations, agents publics, monde associatif, écosystème académique, etc.), de façon plus détaillée, de toutes les informations permettant l'analyse de la responsabilité et de l'impact des plateformes dans la dynamique informationnelle ;
- la fourniture d'informations complètes au régulateur, en lui transmettant l'ensemble des éléments lui permettant la meilleure compréhension possible des mesures déployées.

3.1 Transparence sur les indicateurs et données chiffrées

L'Arcom a préconisé aux opérateurs, dans son précédent rapport, de faire preuve de plus de transparence vis-à-vis du public en fournissant davantage de précisions chiffrées et d'éléments contextualisés, notamment dans les déclarations, et de lui communiquer toutes les informations permettant de mieux comprendre les mesures prises et leur impact.

Cette année, elle a invité les opérateurs à fournir des données chiffrées sur les fausses informations (ou notion équivalente) identifiées sur leurs services en France, en 2021, et sur les mesures prises en conséquence. Pour ce faire, elle leur a demandé de renseigner des indicateurs généraux tels que le nombre de contenus identifiés comme comportant une fausse information, la part de ces contenus détectés automatiquement et de ceux détectés par des signalements d'utilisateurs, le nombre de ces contenus ayant fait l'objet de mesures, leur audience et le nombre d'émetteurs. Or, seuls [Dailymotion](#), [LinkedIn](#), [Snap](#) et [Twitter](#) ont communiqué des données chiffrées¹².

Elle a également souhaité se voir communiquer des indicateurs relatifs à chacun des six domaines de la loi sur lesquels les opérateurs doivent mettre en place des moyens et mesures et faire preuve de transparence sur ceux-ci. Les déclarations sont sur ce point hétérogènes et au final peu fournies.

[Webedia](#) a communiqué plusieurs données chiffrées mais considère, au regard du modèle de son service, ne pas être concerné par une majorité d'entre eux. [Dailymotion](#) et [Snap](#) ont également déclaré quelques chiffres demandés par le régulateur mais de manière très lacunaire.

[Twitter](#) a uniquement fait état d'indicateurs généraux. Il en est de même pour [LinkedIn](#) et [Pinterest](#), mais une majorité de leurs indicateurs sont déclarés de manière confidentielle.

¹¹ « (...) Ces mesures, ainsi que les moyens qu'ils y consacrent, sont rendus publics (...). »

¹² Voir fiches opérateurs en annexe.

[Meta](#) renvoie vers ses rapports de transparence et a communiqué quelques chiffres sur les retraits de contenus et les faux comptes, dont certains de manière confidentielle.

[TikTok](#) renvoie également vers ses rapports de transparence mais ne déclare que des éléments chiffrés au niveau mondial.

[Microsoft](#) et [Google](#) n'ont communiqué aucun élément chiffré et ont renvoyé vers leurs rapports de transparence.

La [Fondation Wikimedia](#) et [Yahoo](#) ont considéré ne pas être concernés par les indicateurs demandés, au regard du modèle de leur service.

À l'appui de cette situation, [Meta](#) déclare ne pas suivre tous les types de données et d'indicateurs demandés par l'Arcom sur les fausses informations et les contenus. [Twitter](#) rappelle que certaines données demandées sont difficiles à récupérer ou à produire. En outre, il ressort des échanges avec plusieurs opérateurs que certains redoutent que ces données soient utilisées pour contourner les moyens mis en œuvre pour identifier et modérer les contenus et, par ailleurs, que l'identification de données à l'échelle d'un pays n'est pas aisée. Concernant les données déclarées de manière confidentielle, certains opérateurs ont invoqué le secret des affaires ou le risque d'affecter l'attractivité de leur service.

Sur ces différents points et justifications, l'Arcom entend formuler plusieurs remarques.

Tout d'abord, lorsque l'opérateur n'est pas en mesure de communiquer l'ensemble des données chiffrées demandées, la transparence devrait s'effectuer *a minima* sur des données chiffrées générales permettant une évaluation globale de la pertinence des moyens et mesures mis en œuvre. L'absence de communication d'informations telles que le nombre de fausses informations (ou notion équivalente) identifiées en France, le nombre d'émetteurs de ces contenus ou leur audience moyenne témoigne non seulement d'un manque de transparence à l'égard du public, mais soulève en outre la question de la capacité des opérateurs à identifier ces données et à évaluer leurs propres politiques en matière de lutte contre la manipulation.

Par ailleurs, l'Arcom estime que l'usage, par les opérateurs, de la confidentialité sous couvert de secret des affaires apparaît souvent peu justifié.

En conséquence, elle déplore vivement les efforts très insuffisants des plateformes en matière de transparence des données chiffrées relatives à l'ampleur du phénomène de désinformation et aux résultats des moyens qu'elles déploient pour le contrer.

3.2 Transparence sur les moyens et mesures mis en œuvre par les opérateurs

L'Arcom tient à noter une amélioration de la transparence de la majorité des opérateurs sur les moyens et mesures mis en œuvre pour lutter contre la manipulation de l'information.

Outre le fait que les déclarations soient plus fournies en informations précises, plusieurs opérateurs ([Google](#), [Meta](#), [Snap](#), [TikTok](#), [Twitter](#) et [Wikimédia France](#)) ont échangé avec le régulateur pour apporter des informations complémentaires à celles contenues dans leurs déclarations. [Dailymotion](#) a pour sa part étoffé la version publique de sa déclaration annuelle à la suite d'un échange portant sur l'étendue des données confidentielles.

Cependant, le niveau de transparence demeure très hétérogène d'un opérateur à l'autre.

[Yahoo](#) se distingue cette année encore par une déclaration lacunaire, tant sur la forme que sur le fond des éléments déclarés.

La déclaration de [TikTok](#) est particulièrement imprécise, avec peu d'informations relatives au service en France et pas d'éléments tangibles permettant une évaluation des moyens et mesures mis en place.

[Google](#), [TikTok](#), [LinkedIn](#) et [Pinterest](#) n'ont pas ou peu répondu aux interrogations qui leur étaient spécifiquement adressées dans le questionnaire de l'Arcom.

Les opérateurs restent globalement peu transparents sur des aspects majeurs du bilan que sont le fonctionnement des outils de détection de fausses informations ou d'identification de pratiques problématiques, les pratiques de désinformation identifiées sur leur service et le fonctionnement de leurs politiques publicitaires.

Face à ces lacunes, l'Arcom ne peut que réitérer sa conviction que l'efficacité de la lutte contre la manipulation de l'information et, plus largement, des désordres informationnels touchant l'espace numérique, ne peut aller sans un accès du public, de la société civile et des pouvoirs publics aux informations permettant de comprendre les phénomènes et d'évaluer les politiques déployées par les plateformes et leurs résultats.

Pour ce faire, il est nécessaire que soient mis en place des espaces dédiés facilement accessibles pour permettre une meilleure compréhension, par le public, des risques de manipulation de l'information relatifs à la plateforme mais également pour faire état de celles qui ont pu être identifiées.

Par ailleurs, une meilleure compréhension des risques systémiques sur les services des opérateurs passe par une approche ouverte impliquant la société civile, et notamment la communauté académique. À cette fin, l'accessibilité des données des plateformes en ligne est fondamentale. L'Arcom se félicite à cet égard du dispositif ambitieux et prometteur prévu par le DSA en la matière.

Le Code européen de bonnes pratiques renforcé contre la désinformation devrait permettre une avancée majeure de transparence, les signataires s'étant engagés à fournir publiquement des indicateurs qualitatifs et quantitatifs à l'échelle européenne et parfois, à celle des États membres. Mais cet effort ne portera ses fruits que si la société civile, l'Europe et ses États membres se saisissent de ces informations pour poursuivre et approfondir un dialogue exigeant avec les opérateurs.

Préconisations

Au vu de ces éléments d'analyse, l'Arcom formule les recommandations suivantes :

- ❖ faire preuve de plus de transparence vis-à-vis du public en fournissant davantage d'indicateurs chiffrés aussi bien sur les contenus, pratiques et phénomènes de manipulation de l'information identifiés que sur les moyens et mesures mises en œuvre pour lutter contre ceux-ci ;
- ❖ mettre en place davantage d'initiatives pour améliorer la transparence vis-à-vis de la société civile et de la recherche, notamment en matière d'accessibilité des données, afin de permettre une évaluation indépendante et plurielle des risques et des mesures déployées pour les contrer ;
- ❖ communiquer au régulateur davantage d'informations, de manière confidentielle si nécessaire, pour permettre une meilleure compréhension des moyens et mesures mis en œuvre pour lutter contre la manipulation de l'information.

II. Analyse des moyens déployés par les plateformes pour lutter contre la manipulation de l'information en 2021

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

La loi du 22 décembre 2018 dispose que les opérateurs de plateforme en ligne sont tenus de mettre en place un dispositif de signalement facilement accessible et visible permettant à leurs utilisateurs de signaler de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité des scrutins mentionnés au premier alinéa de l'article 33-1-1 de la loi du 30 septembre 1986¹³.

L'Arcom relève à cet égard que les signataires du code européen de bonnes pratiques renforcé contre la désinformation se sont engagés à mettre en place une fonctionnalité permettant aux utilisateurs de signaler des informations fausses ou trompeuses violant leurs conditions générales.

Les déclarations adressées à l'Arcom diffèrent peu sur ce point des exercices précédents, les opérateurs renvoyant aux informations fournies l'an passé (*Meta*, *Google*) ou ne faisant pas état de nouveautés notables (*Dailymotion*, *Twitter*, *LinkedIn*, *Webedia*). Cette année encore, la *Fondation Wikimédia* indique que le dispositif de signalement sur *Wikipédia* « *n'est parfois pas la solution la plus appropriée en ce qu'elle ne permet pas d'atteindre la meilleure efficacité possible pour lutter contre les fausses informations en ligne.* »

L'Arcom ne dispose pas des données chiffrées pertinentes pour évaluer l'effectivité du traitement de ces signalements et, *a fortiori*, analyser son évolution. Ces informations sont en effet soit inexistantes (*Microsoft*, *Twitter*, *Yahoo*), soit partielles (*Webedia*, *LinkedIn*, *Dailymotion*, *Snap*), soit agrégées à l'échelle mondiale (*Google*, *Meta*, *TikTok*, *Pinterest*), soit fournies à titre confidentiel (*Pinterest*). L'Arcom souligne une nouvelle fois le défaut de coopération des plateformes sur ce volet et regrette que la loi ne lui fournisse pas d'outil pour contraindre les opérateurs à lui communiquer ces données.

1.1 Une hétérogénéité dans l'accessibilité et la visibilité des dispositifs de signalement, notamment ceux des moteurs de recherche

Dans sa recommandation du 15 mai 2019, l'Arcom a invité les opérateurs à assurer la facilité d'accès et la bonne visibilité du dispositif de signalement, notamment en désignant clairement ce dernier et en le plaçant à proximité immédiate des contenus et comptes pouvant être signalés.

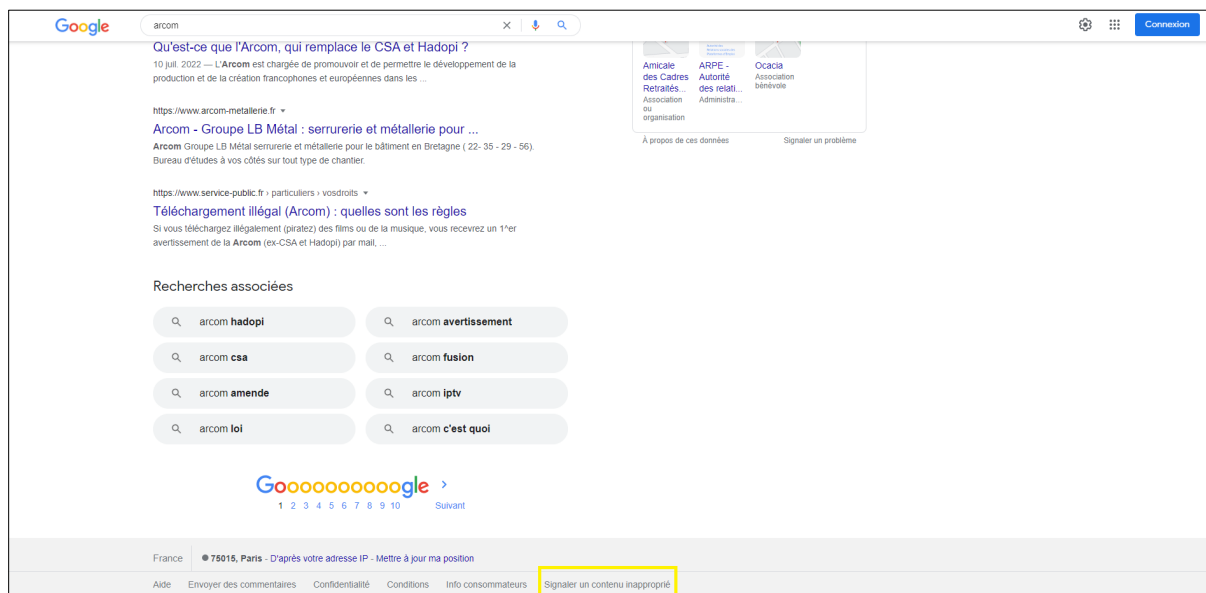
• Moteurs de recherche

Dans son précédent rapport annuel, l'Arcom a préconisé aux moteurs de recherche « *d'améliorer la visibilité et la facilité d'utilisation de leur dispositif de signalement* ». Pour l'établissement du présent document, l'Arcom a demandé aux moteurs de recherche de fournir les raisons pour lesquelles il n'est pas possible de signaler un contenu issu d'une recherche à partir d'un outil de signalement à proximité directe du résultat de recherche.

En réponse, *Google* indique qu'il considère que « *l'accessibilité au dispositif de signalement de fausses informations sur les moteurs de recherche doit être appréciée en tenant compte de la nature particulière du service (...) qui diffère des réseaux sociaux* ». Il affirme que les critères de visibilité et de facilité d'utilisation ont été pris en compte lors de la conception

¹³ 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.

du formulaire de signalement de *Google Search*, qui est disponible pour toutes les versions du service. Il précise néanmoins avoir pris en compte les préconisations formulées par l'Arcom dans les précédents bilans, en rendant disponible « *directement et aisément* » son dispositif et en ajoutant la catégorie « *fausses informations* »¹⁴ parmi les motifs de signalement.

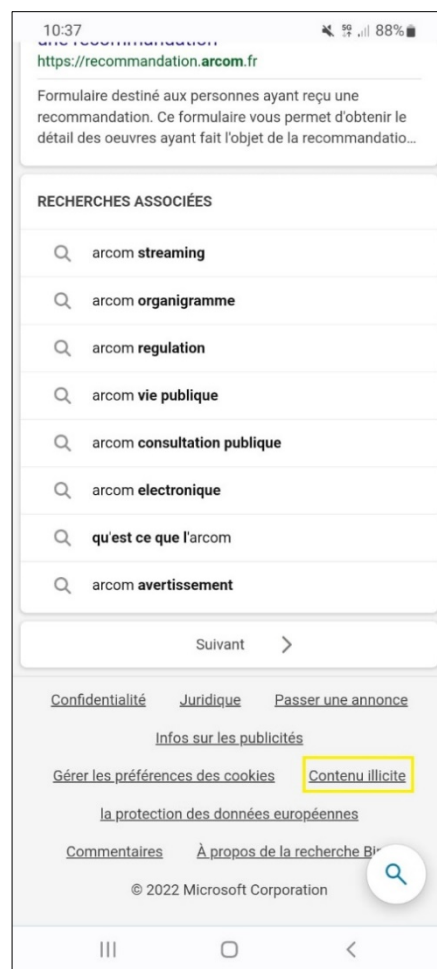
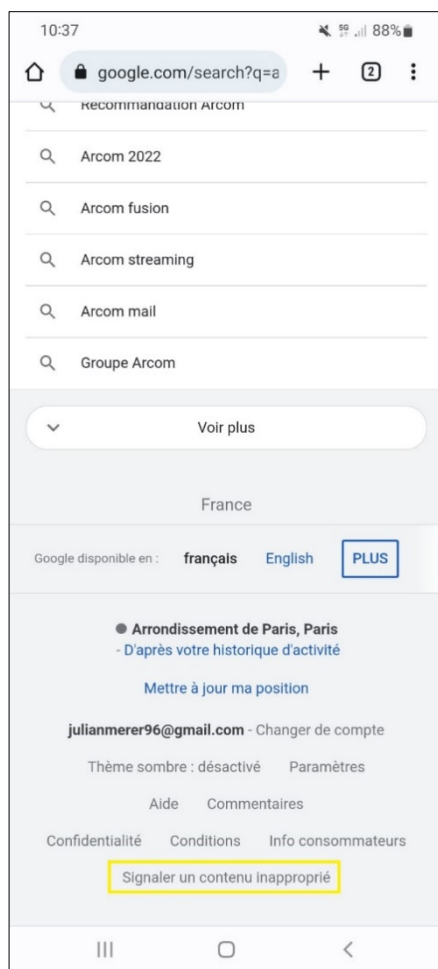


Accès au dispositif de signalement de Google Search sur ordinateur (navigateur : Chrome)¹⁵

Microsoft rappelle que l'utilisateur de *Bing* est informé de la possibilité de procéder à des signalements par des liens en bas de page renvoyant vers un formulaire et considère ce dispositif suffisant. Il ajoute que « *La transparence dans les possibilités offertes à l'utilisateur par des pages informatives [...] semble plus efficace car chaque élément ajouté sur la page de recherche a un impact sur la performance du moteur de recherche (vitesse d'affichage etc.).* »

¹⁴ L'ajout du motif « fausses informations » a été réalisé au cours de l'été 2022 selon *Google*.

¹⁵ Source : observations Arcom, octobre 2022.



Accès au dispositif de signalement de Google Search (à gauche) et Bing (à droite) sur téléphone (navigateur : Chrome)¹⁶

L'Arcom salue la prise en compte de ses préconisations par [Microsoft](#) et [Google](#) sur la facilité d'utilisation et la visibilité du dispositif mais estime qu'elle demeure insuffisante et appelle les deux opérateurs à fournir davantage d'efforts. Elle souligne par ailleurs que les versions mobile de ces services accessibles à partir d'un navigateur posent également des difficultés, car il est nécessaire de *scroller* plusieurs fois avant de pouvoir accéder au dispositif afin de signaler des contenus.

La déclaration de [Yahoo](#) ne portant pas sur le service [Yahoo Search](#), l'opérateur n'a pas fourni d'éléments en réponse aux remarques et interrogations de l'Arcom.

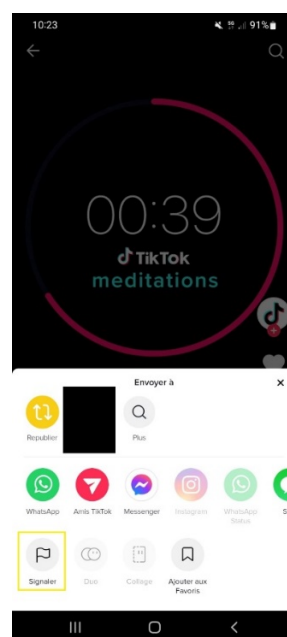
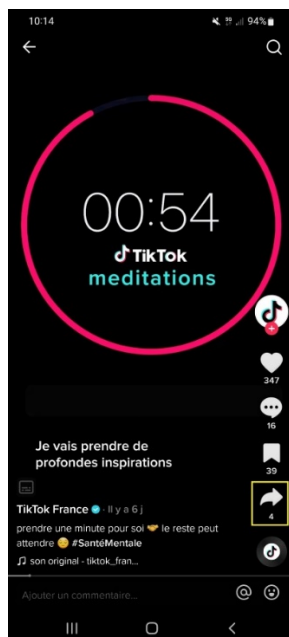
• *Autres plateformes*

Comme l'an passé, l'Arcom estime que la majorité des plateformes assurent une visibilité satisfaisante à leurs outils de signalement, qui sont pour la plupart situés à proximité du contenu et facilement accessibles.

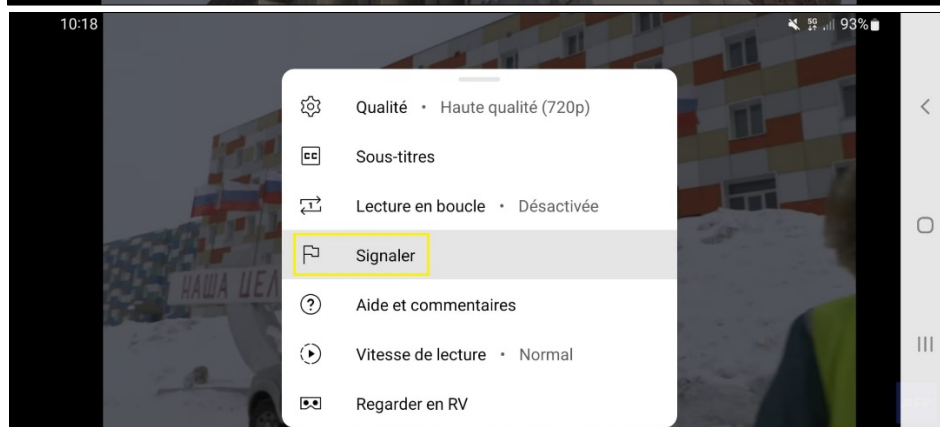
Néanmoins, le choix du pictogramme attaché à ces outils est parfois peu intuitif. Ainsi, sur [TikTok](#), le dispositif de signalement n'est visible qu'après avoir appuyé sur une icône en forme de « flèche » qui symbolise le repartage du contenu. Par ailleurs, le dispositif de

¹⁶ Id.

signalement de l'application [YouTube](#) n'est accessible que dans les paramètres et non pas à proximité immédiate du contenu.



Accès au dispositif de signalement de TikTok (application)¹⁷



Accès au dispositif de signalement de YouTube (application)¹⁸

¹⁷ Id.

¹⁸ Id.

En outre, l'Arcom rappelle que les opérateurs sont incités à fournir un outil de signalement commun à toutes les déclinaisons de leur service, quel que soit le mode d'accès.

1.2 Un manque de pertinence de certains motifs de signalement

- *Motifs et sous-motifs proposés par les opérateurs¹⁹*

Service	Motif relatif à la manipulation de l'information (ou notion équivalente) en 1 ^{er} niveau	Sous-motifs (2 ^e niveau)
Dailymotion	Désinformation	NA
Bing	J'ai un autre problème	Je demande la suppression de contenu conformément à la loi locale
Google Search	Autre problème d'ordre juridique : signaler du contenu pour un motif juridique non répertorié (diffamation, fausses informations, contenu incitant à la haine, etc.).	Problème d'ordre juridique : problème d'ordre juridique non répertorié.
YouTube	Spam ou contenu trompeur	Aucun
LinkedIn	Douteux, spam ou faux	Désinformation
Facebook	Fausses informations	Santé Politique Enjeu social Autre chose
Instagram		
Pinterest	Désinformation	Désinformation médicale Désinformation sur la participation civique Désinformation sur le climat Théorie du complot
Snapchat	Fausses informations	Politique Santé Problème de société Autre
TikTok	Désinformation nuisible	Élections Santé Contenu de marque non divulgué Autre désinformation nuisible
Twitter	Il induit en erreur au sujet d'élections	Fausses informations sur des élections ou l'inscription sur les listes électorales Fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité d'un des scrutins
Jeuvideo.com	Aucun	Aucun
Wikipédia	Pas de dispositif de signalement	
Yahoo Media		

Source : observations de l'Arcom, octobre 2022

¹⁹ Sur certains services, le dispositif de signalement des communications commerciales diffère (voir partie dédiée).

- **Analyse de la pertinence des motifs proposés**

Depuis 2018, plusieurs opérateurs ([Snap](#), [LinkedIn](#), [Dailymotion](#), [Google](#)) ont modifié leur dispositif de signalement pour y intégrer un motif relatif à la manipulation de l'information (ou notion équivalente). Toutefois, les choix proposés par les dispositifs de certaines plateformes ne couvrent pas le champ de fausses informations au sens de la loi.

La majorité des dispositifs de signalement des opérateurs proposent le terme « désinformation » ou « fausses informations » ([Dailymotion](#), [Google Search](#), [Facebook](#), [Instagram](#), [Pinterest](#), [Snapchat](#), [TikTok](#)) dès le premier niveau. Concernant les sous-motifs (second niveau), certains de ces mêmes opérateurs offrent une granularité intéressante permettant de couvrir différents types de risques et ce, en des termes clairs ([Facebook](#), [Instagram](#), [Pinterest](#), [Snapchat](#), [TikTok](#)).

À cet égard, [Pinterest](#) est le seul opérateur à proposer un motif de signalement spécifique à la désinformation sur le climat. L'Arcom salue la mise en place de cette option de signalement et invite l'opérateur à fournir davantage d'information sur ses résultats.

[LinkedIn](#) et [YouTube](#) proposent respectivement « douteux, spam ou faux » et « spam ou contenu trompeur », termes plus globaux évoquant moins directement la désinformation. [LinkedIn](#) y fait toutefois référence en second niveau, à l'inverse de [YouTube](#).

Le cas de [Twitter](#) est particulier : le motif de premier niveau couvre uniquement la désinformation en contexte électoral et ne mentionne la désinformation susceptible de troubler l'ordre public qu'en sous-motif, ce qui peut induire en erreur l'utilisateur, voire le décourager.

10:23 91%

← Signaler un problème

Aidez-nous à comprendre. En quoi ce Tweet pose-t-il problème ?

- Ce Tweet ne m'intéresse pas.
- Il est suspect ou publie du spam.
- Les propos tenus sont inappropriés ou dangereux.
- Il induit en erreur au sujet d'élections.
- Il exprime des intentions suicidaires ou autodestructrices.

[En savoir plus](#) sur le signalement des infractions à nos règles

10:23 91%

← Signaler un problème

En quoi ce Tweet est-il suspect ou propage-t-il du spam ?

- Le compte qui tweete est un faux compte.
- Il inclut un lien vers un site potentiellement dangereux, malveillant ou de hameçonnage.
- Les hashtags inclus ne semblent pas avoir de rapport.
- Il utilise la fonction Répondre pour spammer.

Autre chose

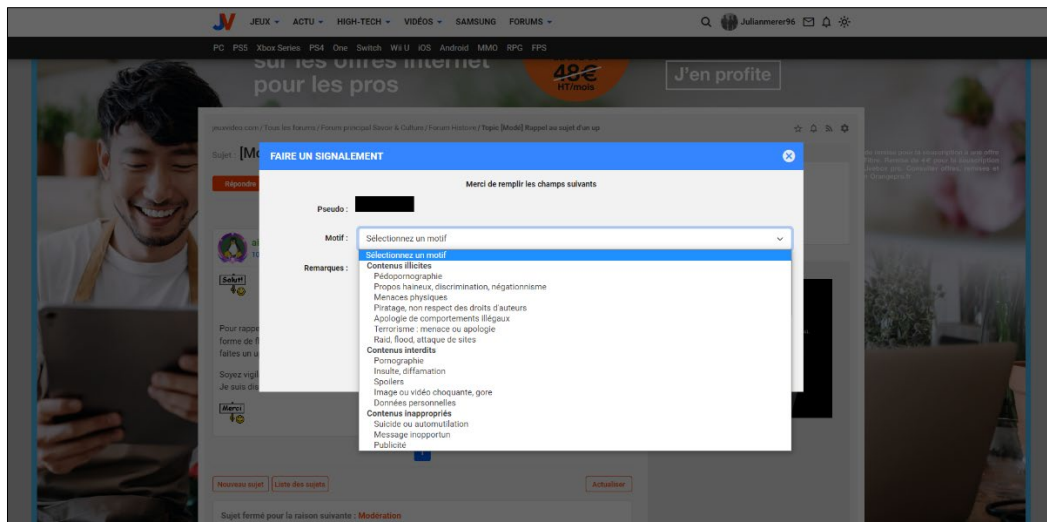
[En savoir plus](#) sur les infractions à nos règles relatives au spam

Motifs de signalement proposés sur Twitter (application)²⁰

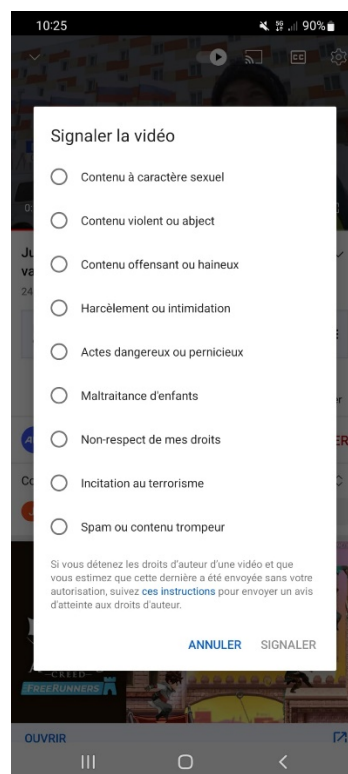
Les dispositifs de [Bing](#) et [Jeuxvideo.com](#) n'y font simplement pas référence.

L'Arcom invite les opérateurs dont les outils de signalement ne retiennent pas le critère de manipulation de l'information (ou notion équivalente) ou le font de façon ambiguë à remédier à cette situation.

²⁰ Id.



Motifs de signalement proposés sur Jeuxvideo.com²¹



Motifs de signalement proposés sur YouTube (application)²²

1.3 Un manque de transparence vis-à-vis des utilisateurs sur les processus d'instruction des signalements et les procédures de recours

Malgré les préconisations formulées l'année passée, certains opérateurs n'ont pas fourni davantage d'éléments sur les informations fournies aux utilisateurs pour mieux comprendre les processus d'instruction et les procédures d'appel possibles.

²¹ Id.

²² Id.

- **Un manque d'informations sur le degré d'automatisation du processus de déclenchement d'une action à l'égard d'un compte ou d'un contenu**

Dans son précédent bilan, l'Arcom a préconisé le maintien d'une intervention humaine dans le processus de décision d'une action à l'égard d'un contenu ou d'un compte, à l'exception de certaines pratiques aisément détectables automatiquement pour lesquelles des outils d'intelligence artificielle permettent de soulager le travail des modérateurs sans risque de sur-censure.

Invités à commenter cette préconisation et à fournir des données chiffrées, les opérateurs ont fait état de plusieurs approches quant au recours à la modération automatique.

[Dailymotion](#), [Snap](#) et [Webedia](#) ont indiqué que l'ensemble des décisions d'intervention étaient prises à l'issue d'une intervention humaine. Pour [Webedia](#), si l'automatisation lui apparaît appropriée pour certains contenus (pédopornographie, contenus à caractère terroriste, etc.) pour lesquels le caractère illicite ou préjudiciable ne peut être remis en cause, la décision finale doit être prise par un modérateur humain « *en mesure de contextualiser la portée du message et décider de l'action de modération à apporter* ». [Microsoft](#) indique s'inscrire dans la logique de la préconisation de l'Arcom mais n'apporte pas d'informations sur la place de l'intervention humaine dans son processus de modération et évoque uniquement les standards éthiques encadrant l'intelligence artificielle.

[Twitter](#) déclare que sa modération n'est pas axée « *sur les personnes, mais sur la technologie* ». Il explique qu'en raison du nombre de contenus diffusés simultanément sur son service, il se considère bien plus efficace dans la « *lutte contre les contenus préjudiciables en utilisant davantage la technologie* ». [Meta](#) a une approche différente, considérant lui-aussi la technologie comme essentielle dans son système de modération mais principalement en tant que soutien au travail des équipes. L'Arcom regrette que ces deux opérateurs ne lui aient pas fourni de données chiffrées sur la part des contenus détectés et traités de façon automatique par rapport à la part opérée par des humains. Tout en étant consciente de l'enjeu lié au nombre de signalements à traiter, elle manifeste son inquiétude quant à l'usage exclusif de l'intelligence artificielle dans la modération d'un contenu, *a fortiori* s'agissant de la manipulation de l'information.

[Google](#), [LinkedIn](#), [Yahoo](#), [TikTok](#) et [Pinterest](#) n'ont pas commenté la préconisation précitée, ce que l'Arcom regrette vivement. [TikTok](#) a néanmoins déclaré qu'il intervenait notamment avant signalement, en utilisant à la fois la technologie et la modération humaine. [Pinterest](#) a une approche hybride qui conjugue examen manuel et automatisé, sans donner davantage de précisions.

Seuls [Dailymotion](#), [Snap](#) et [Webedia](#) ont fourni les chiffres demandés par l'Arcom, qui permettent d'avoir un regard plus précis sur les moyens qu'ils déploient²³. Ainsi, le temps moyen de traitement d'un signalement relatif à un contenu de désinformation (ou notion équivalente) diffère d'une de ces trois plateformes à l'autre, en restant court : en moyenne, moins de deux heures pour [Dailymotion](#), 4,3 heures sur [Snapchat](#) et, tous types de signalements confondus, moins de 24 heures sur [Jeuxvideo.com](#). L'Arcom souligne que la comparaison de ces chiffres doit être faite avec précaution, de nombreux paramètres étant susceptibles d'entrer en jeu dans le temps de traitement d'un signalement (nature des messages, difficulté d'analyse, nombre de signalements reçus, etc.).

Par ailleurs, les trois opérateurs déclarent que plus de 9 contenus signalés sur 10 sont traités dans la journée (tous types de signalements confondus pour [Jeuxvideo.com](#) et

²³ [Pinterest](#) et [LinkedIn](#) ont également fourni de tels chiffres, mais de manière confidentielle.

Dailymotion, s'agissant spécifiquement des signalements pour fausses informations pour *Snapchat*).

L'Arcom souligne le besoin d'avoir des données comparatives permettant de rendre compte de la pertinence des politiques des opérateurs dans le traitement des signalements.

- ***Un manque de clarté dans les déclarations sur les informations apportées aux utilisateurs durant les procédures de recours***

L'an passé, l'Arcom avait demandé aux opérateurs d'être plus transparents sur la procédure de modération envers les auteurs de signalements et des utilisateurs ayant publié un contenu signalé et ce, à différentes étapes :

- en leur fournissant des informations sur l'avancée du traitement du signalement en cours ;
- en leur communiquant la décision prise à son issue dans un délai raisonnable ;
- en leur indiquant plus clairement et explicitement les voies de recours existantes.

De manière générale, les opérateurs n'apportent pas suffisamment d'éléments dans leurs déclarations permettant de comprendre comment les utilisateurs sont informés de l'avancée de la procédure en cours de traitement (alors même qu'il ressort des observations réalisées par les services de l'Arcom que certaines plateformes mettent ce type d'outil à disposition).

Concernant l'information communiquée aux utilisateurs sur la décision prise à l'issue du traitement du signalement, les déclarations des opérateurs sont parcellaires. L'Arcom note que *Meta*, *TikTok*, *LinkedIn*, *Webedia* et *Pinterest* fournissent proactivement une information à l'utilisateur affecté par une décision de modération, *Dailymotion* réservant cette faculté aux seuls utilisateurs professionnels. *TikTok* informe également l'auteur du signalement de l'issue de l'examen. En revanche, aucun d'entre eux ne fait état du délai dans lequel ces éléments sont communiqués.

L'ensemble des opérateurs propose désormais une procédure d'appel d'une décision de modération d'un compte ou d'un contenu (*Snap* depuis janvier 2022), à l'exception de *Yahoo* qui ne l'évoque pas et de la *Fondation Wikimédia* qui n'en propose pas étant donnée la nature de son service. L'information sur ces possibilités de recours, quand elles existent, sont généralement fournies dans la notification de la décision.

En revanche, la possibilité pour l'auteur du signalement de former un recours à l'issue de la décision semble une pratique encore peu développée : seul *Meta* affirme le lui permettre, ainsi que *Dailymotion* depuis 2022. *TikTok* précise que l'auteur du signalement est informé de l'action prise le cas échéant, mais que si le signalement n'a pas été pris en compte, il ne peut pas faire l'objet d'un appel²⁴.

Une fois le recours formé, *Microsoft* et *LinkedIn* précisent qu'ils informent l'utilisateur du lancement de la procédure (*Microsoft* envoyant automatiquement un accusé de réception à l'auteur du signalement). À l'issue de celle-ci, *Pinterest* et *Dailymotion* adressent un courriel sur le résultat du réexamen. *Microsoft* informe seulement certains demandeurs en fonction de la nature de la violation. *TikTok*, *Meta*, *Google*, *Webedia* et *Twitter* décrivent les procédures de recours contre les suppressions de contenus et de comptes sans expliciter les modalités d'information de leurs utilisateurs.

L'Arcom relève que les opérateurs signataires du code européen de bonnes pratiques renforcé contre la désinformation ont pris un engagement qui recouvre partiellement ses préconisations, consistant à mettre en place un mécanisme d'appel transparent à l'attention des utilisateurs dont le contenu ou les comptes ont fait l'objet de mesures de

²⁴ Source : audition de TikTok en septembre 2022.

modération. Le DSA va plus loin encore, prévoyant non seulement des mécanismes de signalement faciles d'accès et d'utilisation, mais aussi la communication des décisions motivées, d'une part, à l'auteur du signalement et, d'autre part, en cas de décision de modération restreignant la portée d'un contenu (retrait, déclassé, etc.), aux utilisateurs affectés, ainsi qu'une information sur leurs possibilités de recours, internes et externes.

L'Arcom appelle les opérateurs à accentuer leurs efforts pour garantir une meilleure information et une plus grande transparence vis-à-vis des utilisateurs, qu'il s'agisse des auteurs de signalements comme des personnes affectées par une décision de modération, quant aux différentes étapes du traitement de leurs demandes, aux décisions prises et aux motifs les expliquant, et aux possibilités de recours internes et externes. Ces mécanismes et leur accessibilité aux utilisateurs constituent des leviers essentiels pour permettre à chacun de contribuer activement au maintien d'un environnement numérique respectueux des droits des personnes, tout particulièrement de la liberté d'expression.

Préconisations

Au vu de ces éléments d'analyse, l'Arcom formule les recommandations suivantes :

- ❖ améliorer la visibilité et la facilité d'utilisation du dispositif de signalement en le plaçant à proximité immédiate du contenu (notamment pour les moteurs de recherche [Google Search](#) et [Bing](#)), et ce, sur toutes les versions du service (particulièrement pour les versions mobile de [TikTok](#) et [YouTube](#)) ;
- ❖ proposer, au sein des dispositifs de signalement, des motifs ou sous-motifs relatifs à la manipulation de l'information (ou notion équivalente) clairs et suffisamment précis (notamment pour [Twitter](#), [Jeuxvideo.com](#), [LinkedIn](#) et [YouTube](#)). Ces motifs doivent également couvrir la manipulation de l'information susceptible de troubler l'ordre public ou de nuire au bon déroulement des processus électoraux ;
- ❖ mieux informer les auteurs de signalement et utilisateurs ayant publié un contenu signalé de l'avancée des procédures de traitement des signalements en cours et leur en communiquer l'issue dans un délai raisonnable et de façon motivée ;
- ❖ permettre aux auteurs de signalement et aux utilisateurs affectés par une décision de modération de former un recours et leur fournir une information claire sur leurs possibilités de recours au moment de la notification de la décision prise.

2. Transparence des algorithmes

En matière de lutte contre la manipulation de l'information, deux types de systèmes algorithmiques entrent tout particulièrement en jeu : les systèmes de recommandation de contenus et les systèmes utilisés à des fins de modération. Les premiers permettent généralement de sélectionner, ordonnancer et prioriser de façon individualisée le contenu proposé à l'utilisateur tandis que les seconds visent à détecter ou traiter les contenus préjudiciables ou illégaux. Il apparaît essentiel que les utilisateurs puissent disposer des informations nécessaires sur le fonctionnement de ces outils pour exercer leur esprit critique sur les contenus qui leur sont proposés, et sur ceux qui ont été détectés, puis modérés.

Il convient de noter que la situation de [Webedia](#) et de la [Fondation Wikimedia](#) est peu analysée dans cette partie, leurs services n'utilisant pas ou peu d'algorithmes.

Certaines réponses des opérateurs diffèrent peu des informations déjà déclarées l'année passée ou accessibles sur leurs sites et ne permettent pas à l'Arcom d'apprécier de manière exhaustive l'efficacité des moyens mis en œuvre. Plusieurs opérateurs ne répondent à aucune des questions posées et se limitent à fournir des informations générales sur leurs outils ([LinkedIn](#), [TikTok](#), [Pinterest](#)). D'autres opérateurs se contentent d'apporter quelques informations sur leurs systèmes de recommandation sans expliquer le fonctionnement de leurs outils automatiques de modération ([Dailymotion](#), [Google](#), [TikTok](#)), alors même que l'Arcom avait précisé ses attentes sur ce point au regard des réponses peu satisfaisantes reçues l'an passé.

De manière générale, l'Arcom note une certaine amélioration des déclarations s'agissant des critères utilisés pour la recommandation de contenus et des critères entraînant une baisse de visibilité des contenus considérés comme étant de faible qualité, afin de prévenir des risques de manipulation de l'information.

Elle salue par ailleurs l'effort des opérateurs pour donner plus de précisions sur les processus d'évaluation des outils automatiques. Elle regrette cependant vivement qu'aucun des opérateurs ne fournisse d'indicateurs chiffrés permettant d'évaluer la performance de leurs outils de modération et de recommandation.

2.1 Une hétérogénéité des informations relatives aux algorithmes de recommandation

Dans sa recommandation du 15 mai 2019, l'Arcom appelait les plateformes à fournir une information claire, suffisamment précise et facilement accessible sur les critères ayant conduit à l'ordonnancement du contenu proposé à l'utilisateur et le classement de ces critères selon leur poids dans l'algorithme.

Dans le même esprit, les signataires du code européen de bonnes pratiques renforcé contre la désinformation se sont engagés à rendre transparents les principaux critères et paramètres utilisés pour hiérarchiser ou déclasser les contenus (engagement 19).

- ***La déclaration de nombreux critères utilisés pour la recommandation de contenus, sans mention de leur équilibre***

Afin d'effectuer la tâche ou série de tâches pour laquelle ils ont été programmés, les algorithmes de recommandation utilisent des données en entrée et produisent un résultat en sortie qui permet la recommandation de contenus. Ainsi, le choix des critères principaux utilisés par les opérateurs et leur équilibre peuvent être des facteurs expliquant une plus ou moins grande exposition à des fausses informations. À l'exception de [LinkedIn](#) et de

[Pinterest](#), les opérateurs concernés font état de nombreux éléments sur lesquels ils basent les recommandations.

Concernant **les résultats de recherche de contenus**, [Google](#) indique afficher les meilleurs résultats d'une recherche sur [Google Search](#) en fonction des mots-clés de la requête, puis de la recherche de correspondance ainsi que du classement des pages selon leur utilité. [Microsoft](#) génère les résultats de recherche sur [Bing](#) à l'aide de systèmes de classement relatifs aux mots sur la page ou au titre de la page mise en avant. En outre, les deux opérateurs indiquent se baser sur des critères de pertinence et de fiabilité de l'information pour empêcher la mise en avant de contenus relatifs à de la manipulation de l'information, mais sans préciser comment ils évaluent cette fiabilité.

Le risque de mise en avant de résultats de recherche indexant des contenus de désinformation peut aussi découler d'un phénomène appelé « vides de données » (*data voids*) qui correspond à l'absence d'informations fiables sur un sujet recherché. Pour y répondre, [Google](#) affiche un message informant l'utilisateur qu'il n'a pas été en mesure de trouver de sources fiables correspondantes. [Microsoft](#) a lancé des projets de recherche en la matière.

Les résultats de recherche sur [YouTube](#) sont basés sur trois notions : pertinence, engagement et qualité. [Google](#) indique s'appuyer sur des évaluateurs externes ne précise malheureusement pas les critères d'évaluation lui permettant de juger de la qualité d'un contenu.

Concernant **la recommandation sur le fil d'actualité ou la page d'accueil**, [Meta](#) déclare que la recommandation s'effectue en trois temps : un processus d'intégrité pour réduire le nombre de publications à recommander (sans donner davantage de précisions), la détermination d'un score de qualité et un processus de diversification du type de contenu (contenu écrit, vidéo, issu d'un groupe public ou privé, etc.). Le score de qualité est attribué à chaque contenu sur la base de « *milliers de critères* » (tels que la mise en avant de contenus non lus, la relation avec le compte partageant le contenu ou encore le type de publication)²⁵. [Meta](#) prend également en compte les types de « réactions » de l'utilisateur face aux contenus pour en pondérer la présentation²⁶.

Les contenus recommandés sur [Twitter](#) sont majoritairement issus de comptes suivis par l'utilisateur, mais l'opérateur ne précise pas les critères utilisés à l'exception de la popularité ou des interactions avec les autres utilisateurs.

Sur [TikTok](#), la recommandation est en premier lieu basée sur les intérêts exprimés par l'utilisateur lors de l'inscription, sur l'utilisation par ce dernier du bouton « *Pas intéressé(e)* » et sur les paramètres du compte. La personnalisation du fil se base ensuite sur l'activité de l'utilisateur, telle que les réactions sur les contenus, leur partage, les comptes suivis et les commentaires publiés. Les informations d'une vidéo (*hashtags*, légendes et son) sont également utilisées pour pondérer l'ensemble de ces données en fonction de la « valeur » d'un contenu, calculée selon la manière qu'ont les utilisateurs de le consommer²⁷.

Certains opérateurs ont également fait état de critères utilisés pour des **parties spécifiques de leur service**. Sur [Snapchat](#), les contenus sont affichés sur la *Map* en fonction de leur géolocalisation tandis que sur les parties *Discover* et *Spotlight*, la

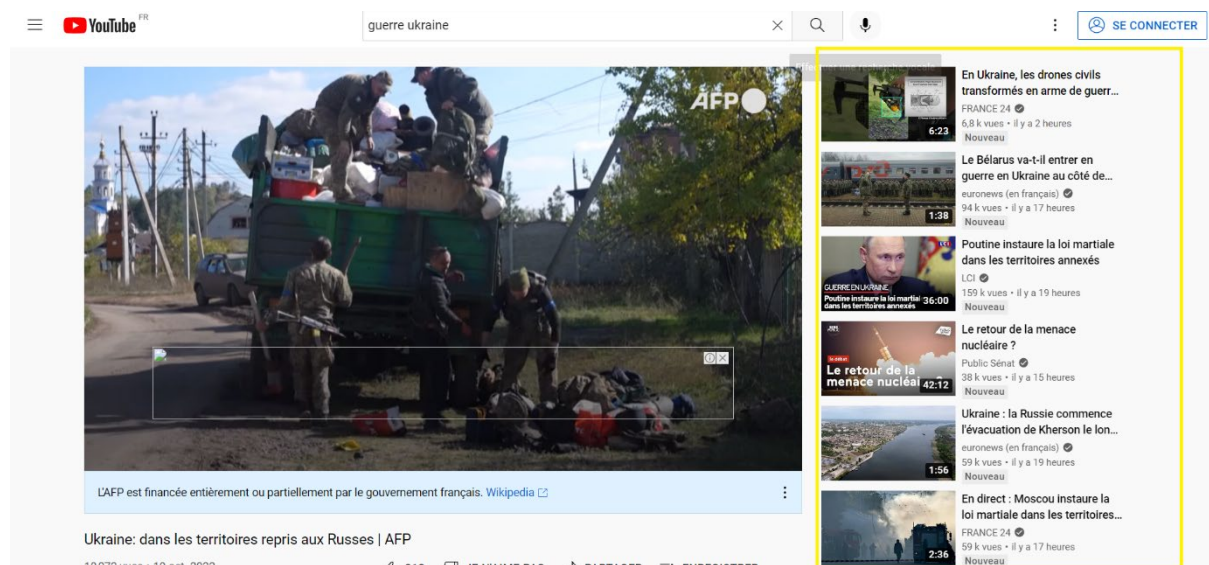
²⁵ [Meta](#) explique également que les données utilisées pour la recommandation sont relativement similaires pour son « fil » et les autres parties de son service, à l'exception du type de contenu proposé.

²⁶ Par exemple, si l'utilisateur a tendance à davantage commenter qu'à utiliser une option de réactions (like, partage, etc.), le fait de commenter un contenu deviendra un signal plus important pour lui recommander des contenus similaires les fois suivantes (Source : audition de Meta en septembre 2022).

²⁷ Il est intéressant de noter que ni le nombre d'abonnés, ni le fait que le compte ait déjà publié des vidéos performantes, ne constituent des facteurs directs dans le système de recommandation de [TikTok](#).

recommandation procède des recherches de l'utilisateur, des contenus consommés, des abonnements et des types de lieux visités.

Les recommandations faites à l'utilisateur en train de consommer un contenu sur *YouTube* (fonctionnalité « À suivre ») se fondent sur plusieurs indicateurs (historique, abonnements, contexte, etc.) mais également sur les usages des autres utilisateurs de la vidéo en cours de visionnage et sur les contenus qu'ils ont consommés par la suite.



Fonctionnalité « À suivre » sur *YouTube* ²⁸

L'Arcom s'interroge néanmoins sur les conséquences d'un mauvais équilibre de l'ensemble ces critères, faute d'informations permettant de comprendre leur hiérarchisation.

Par exemple, s'il peut apparaître pertinent en soi de recommander à un utilisateur des contenus similaires à un autre contenu qui l'a fait réagir, l'engagement initial peut être négatif, faisant peser le risque de continuer à mettre en avant des contenus pour que l'utilisateur continue d'interagir. En outre, cela peut avoir pour conséquence « d'enfermer » l'utilisateur sans lui proposer de contenus apportant de la contradiction. Si certains opérateurs ont expliqué mettre le critère d'engagement en balance avec une évaluation de la qualité ou de la pertinence, le manque d'informations et d'explications sur la hiérarchisation ne permet pas d'évaluer les éventuels risques qui peuvent résulter de ces systèmes. En effet, les nombreuses données en entrée mentionnées dans les déclarations ne contribuent pas toutes de manière égale au résultat, leur équilibre étant ajusté selon l'importance que les opérateurs leur donnent individuellement au regard du résultat qu'ils recherchent.

L'Arcom estime ainsi essentiel que les opérateurs poursuivent leurs efforts pour fournir davantage d'informations à leurs utilisateurs sur la manière dont ces critères sont mis en balance, afin de prévenir les risques liés à la recommandation.

²⁸ Source : observations Arcom, octobre 2022.

- ***Des mesures de réduction de la visibilité des contenus préjudiciables intéressantes bien que peu transparentes***

Les outils de recommandation peuvent également intégrer des critères conduisant à ne pas valoriser des contenus de faible autorité. *YouTube* s'appuie ainsi sur des milliers d'évaluateurs externes pour alimenter les modèles de ses systèmes afin d'identifier et limiter la propagation de contenus « *à la limite d'enfreindre ses règles* ». Sur *Bing* et *Google Search*, les résultats de recherche organiques sont classés de manière à faire apparaître en premier les informations fiables et faisant autorité et, à l'inverse, à priver de visibilité les contenus à faible autorité.

Meta explique que si un partenaire de vérification qualifie une photo, une vidéo, un texte ou un lien de « faux », celui-ci apparaît plus bas dans le fil. En outre, depuis 2021, les groupes politiques et civiques ne peuvent plus faire l'objet de recommandations dans ce dernier.

Twitter dispose d'une série d'outils qui permettent de limiter la visibilité des fausses informations. Sont exclus de sa recommandation certains types de comptes, notamment ceux ayant enfreint les CGU, ceux partageant du contenu trompeur ou violent ou des contenus sensibles ou provenant de médias affiliés à des Etats.²⁹

Ces actions permettent de faire baisser l'exposition et la viralité de fausses informations. On peut toutefois **s'interroger sur la transparence de l'information fournie aux utilisateurs** sur ces mesures de modération. A cet égard, les opérateurs ont des approches différentes. *Google* indique ne pas notifier l'utilisateur lorsque l'un de ses contenus sur *YouTube* est moins ou n'est plus recommandé.³⁰ *Twitter* fait de même afin de ne pas donner d'informations à des acteurs malveillants les incitant à contourner les mesures³¹. *Meta* fait état d'un système de notifications spécifiques, en lien avec les vérificateurs de faits : tous les utilisateurs qui ont publié un contenu dont la visibilité est abaissée sont prévenus de la mesure qui a été prise.³²

- ***Peu d'éléments déclarés quant aux moyens et fonctionnalités mis en place pour la compréhension et la personnalisation des recommandations de contenus***

Dans son précédent bilan, l'Arcom a préconisé de « *proposer aux utilisateurs des fonctionnalités leur permettant de comprendre, si possible de manière personnalisée et contextuelle, les effets des systèmes algorithmiques de recommandation et de modération.* »

Seul *Meta* a mis en place une fonctionnalité contextuelle pour comprendre la recommandation de contenus organiques³³, appelée « *Pourquoi est-ce que je vois ça ?* ». Proposé à proximité immédiate du contenu avec un libellé clair, cet outil apporte aux utilisateurs des informations les aidant à comprendre pourquoi un contenu en particulier apparaît sur leur fil, et propose un second niveau d'information en renvoyant vers des éléments plus détaillés sur la recommandation de contenus. Ainsi, les utilisateurs peuvent connaître la fréquence à laquelle ils interagissent avec des publications provenant d'un compte, la popularité de ce dernier et la fréquence à laquelle ils interagissent avec un type spécifique de publication (vidéos, photos ou liens, par exemple)³⁴. Néanmoins, l'Arcom note que ces informations ne sont pas disponibles pour les contenus relevant de

²⁹ Source : audition de Twitter en septembre 2022.

³⁰ Source : audition de Google en septembre 2022.

³¹ Source : audition de Twitter en septembre 2022.

³² Source : audition de Meta en septembre 2022.

³³ Certains opérateurs le proposent également mais uniquement pour les communications commerciales.

³⁴ Source : <https://about.fb.com/fr/news/2019/04/facebook-lance-la-fonctionnalite-pourquoi-est-ce-que-je-vois-ca-dans-le-fil-dactualite/>

« suggestions » de la plateforme (alors même qu’elles seraient particulièrement utiles) : sur [Facebook](#) et sur [Instagram](#), il est uniquement indiqué que les suggestions s’expliquent « en fonction de plusieurs facteurs, dont votre activité sur Instagram » (sans la fourniture d’informations personnalisées supplémentaires).

Plusieurs opérateurs rendent compte d’informations personnalisées ou plus générales sur le fonctionnement de leurs outils sur des pages dédiées ([Meta](#), [Snap](#), [Twitter](#), [Fondation Wikimédia](#)), dans le centre d’aide ou d’assistance ([Dailymotion](#), [Twitter](#)) ou dans les paramètres du service ([Snap](#), [Meta](#), [Google](#)). D’autres font état uniquement de critères généraux utilisés pour le classement de contenus ([Microsoft](#)).

Quelques opérateurs font mention de fonctionnalités pour personnaliser les recommandations de contenus³⁵. [Snap](#) propose aux utilisateurs de *Spotlight* d’indiquer « Je n’aime pas » sur un contenu et permet de personnaliser les catégories de style de vie qui contribuent à déterminer la mise en avant d’un contenu. [TikTok](#) et [Twitter](#) donnent la possibilité à l’utilisateur de cliquer sur « Pas intéressé(e) » pour l’un et « Ce tweet ne m’intéresse pas » pour l’autre, afin de recommander moins de contenus similaires. L’Arcom salue ces initiatives et fonctionnalités proposées à proximité immédiate du contenu et encourage les opérateurs à mieux expliquer à l’utilisateur les conséquences de cette personnalisation, au cas par cas.

2.2 Un manque de transparence quant au fonctionnement et aux usages des outils de détection

Comme l’année passée, l’Arcom relève que la majorité des informations déclarées par les opérateurs en matière de transparence des algorithmes concernent les outils de recommandation, et non ceux de détection. [Dailymotion](#), [Google](#), [LinkedIn](#), [TikTok](#) et [Yahoo](#) n’apportent pas ou peu d’informations à ce sujet.

Les outils de modération dans la lutte contre la manipulation de l’information visent à **détecter, prioriser le traitement voire traiter les fausses informations**. Ces systèmes automatiques viennent ainsi se substituer ou compléter le travail des modérateurs humains. Si la majorité des opérateurs ([Meta](#), [Pinterest](#), [Twitter](#), [Snap](#), [TikTok](#)) font mention dans leurs déclarations du recours à des outils automatisés pour identifier ou traiter des fausses informations, seul [Meta](#) apporte des précisions sur le fonctionnement des outils de détection. Il utilise des outils d’étiquetage de fausses informations, de reconnaissance de copies de contenus déjà modérés, et de détection automatique des variations de ce contenu. L’Arcom estime qu’il aurait été utile que l’opérateur fasse état de la performance de cet outil et de ses risques de faux positifs. En effet, l’identification de « variations de contenus » fait peser le risque d’une sur-modération.

La [Fondation Wikimédia](#) justifie ne pas utiliser d’outils automatiques de détection, la modération participative étant au cœur de son modèle. L’opérateur précise néanmoins le fonctionnement de deux outils de détection utilisés par d’autres communautés dans d’autres langues mais sur lesquels pourront bientôt se baser certains bénévoles francophones. D’une part, un programme informatique appelé « *ClueBot NG* » développé par des utilisateurs de [Wikipédia](#), utilise l’apprentissage automatique (à partir d’un jeu de données des modifications passées considérées ou non comme du vandalisme) pour identifier des pratiques malveillantes. D’autre part, le système « ORES » attribue un score à des articles et à des contributions afin de prédire leur qualité pour permettre aux contributeurs d’identifier rapidement celles contraires aux règles de l’encyclopédie.

³⁵ [Meta](#) a annoncé en octobre 2022 lancer une fonctionnalité similaire (« Voir plus » et « Voir moins ») pour augmenter ou réduire temporairement le score de classement d’un contenu et des publications similaires.

Par ailleurs, la [Fondation Wikimedia](#) fait état d'un outil en cours de développement qui prendra la forme d'une interface de programmation d'application permettant aux « vérificateurs » de détecter automatiquement les noms d'utilisateurs utilisés par une même personne pouvant être associés à des activités interdites sur l'encyclopédie. Il convient de souligner que l'intervention humaine reste centrale dans la modération, ces outils n'étant qu'une aide supplémentaire pour les bénévoles.

Ainsi, aucun opérateur à l'exception de [Meta](#) et la [Fondation Wikimedia](#) n'apporte d'informations quant au fonctionnement de leurs algorithmes de détection de contenus. Si plusieurs opérateurs, comme [Twitter](#), justifient ce choix par la volonté de ne pas donner de détails pour des raisons de sécurité et empêcher un détournement de leurs services, il subsiste un manque flagrant de transparence en la matière. Néanmoins, cela peut notamment s'expliquer par des approches différentes en matière de lutte contre la manipulation de l'information, certains opérateurs indiquant concentrer leur approche sur la détection de comportements ou phénomènes pouvant porter atteinte à l'intégrité des services. Les informations relatives au fonctionnement des algorithmes d'identification de ce type de pratiques sont analysées dans la partie II.4 du présent bilan.

L'Arcom regrette le défaut de transparence des opérateurs sur ce volet essentiel de leurs obligations. Elle rappelle le besoin impératif de disposer d'informations et d'indicateurs permettant d'évaluer la performance de ces outils, notamment pour les contenus francophones. Elle rappelle également que les opérateurs peuvent lui communiquer des informations de manière confidentielle, lorsqu'elles sont couvertes par le secret des affaires ou que leur publication nuirait à la sécurité de la plateforme.

2.3 Une amélioration de la transparence sur l'évaluation et la modification des outils automatiques

L'Arcom relève qu'en réponse à ses interrogations, plusieurs opérateurs ont abordé les enjeux éthiques liés à la mise en œuvre de leurs systèmes algorithmiques. En revanche, trois opérateurs ([Dailymotion](#), [Google](#) et [Yahoo](#)) n'ont pas apporté davantage d'informations en la matière cette année alors que les réponses apportées l'année dernière avaient été globalement jugées insuffisantes. [TikTok](#) ne répond pas non plus aux questions sur l'évaluation des algorithmes dans sa déclaration.

- ***La mise en place de processus internes ou externes d'évaluation des systèmes algorithmiques***

Pour améliorer leurs outils et procéder à une évaluation continue, certains opérateurs ont mis en place des équipes dédiées en interne ou développé des partenariats externes.

[Twitter](#) effectue des tests sur la précision de détection d'un outil avant son lancement. Pour ce faire, il l'évalue sur des corpus de contenus issus d'une base de référence en calculant la différence entre le nombre d'actions de modération attendues et le nombre d'actions proposées par l'outil. Après l'éventuelle modification et le lancement de ce même outil, une équipe dédiée interne à l'opérateur évalue la performance et effectue des diagnostics récurrents.

[LinkedIn](#) évalue ses outils en amont de leur mise en production et assure un suivi à la suite de leur déploiement. Sans donner davantage de précisions sur les modalités, [Pinterest](#) évoque une amélioration continue de ses modèles par des « analyses hors-ligne » et [Snap](#) une régulation humaine de ses outils par un échantillonnage aléatoire. Enfin, [Meta](#) déclare développer des outils et processus pour évaluer ses outils et prend pour exemple les

progrès réalisés en 2021 en matière de « traitement équitable des communautés » sur ses services (après des audits externes en 2020), exposés dans sa *Newsroom*.³⁶

La majorité des informations déclarées par les opérateurs concernent les processus d'évaluation mais aucun élément, à l'exception de ceux fournis par *Twitter* et *Meta*, ne portent sur les résultats de ces processus et sur les efforts réalisés en conséquence. L'Arcom estime par ailleurs qu'il est important, d'une part, qu'une partie de ces évaluations soit opérée par des acteurs indépendants, et, d'autre part, que les résultats de l'ensemble de ces processus soient rendus publics, afin de permettre aux utilisateurs de mieux comprendre le fonctionnement des systèmes automatiques. L'Autorité remarque également que peu d'éléments ont été déclarés quant à une évaluation spécifique des algorithmes de modération.

- **Plusieurs initiatives en matière de lutte contre les biais**

Plusieurs opérateurs font état de leurs actions de lutte contre les biais des systèmes algorithmiques. *Meta* déclare s'appuyer sur une équipe dédiée à l'intelligence artificielle responsable (« RAI ») pour améliorer la loyauté et l'explicabilité de ses systèmes et précise s'engager activement avec plusieurs acteurs, notamment au niveau européen³⁷, pour collaborer sur l'amélioration des méthodes de mesure de loyauté et des biais sociaux.

Twitter dispose également d'équipes dédiées et s'appuie particulièrement sur le groupe META (*Machine Learning Ethics, Transparency and Accountability*) qui a pour objectif d'apporter de la transparence sur le travail effectué pour améliorer les algorithmes de *machine learning* de l'opérateur. Pour ce faire, le groupe mène des analyses pour évaluer l'existence de préjudices potentiels, applique les enseignements de ces mêmes analyses pour modifier le fonctionnement des outils et « partage les connaissances » afin de travailler en collaboration avec le monde de la recherche en publiant les résultats de leurs études. L'Arcom considère que ce travail d'évaluation des risques et de transparence constituait une bonne pratique particulièrement notable ; aussi, elle prend note avec regret de l'annonce de la suppression de l'équipe META en novembre 2022, à la suite du rachat de la société.

L'Arcom préconise que ce type de travaux ne soient pas uniquement internalisées par les opérateurs. Dans cette perspective, garantir l'accès, par les chercheurs, aux données-clés permettant de comprendre les phénomènes de manipulation existant sur les plateformes apparaît indispensable. En outre, une telle collaboration entrerait en parfait résonance avec les dispositions du DSA en la matière.

LinkedIn propose une boîte à outils en *open source* (sur *Github*) pour mesurer l'équité dans les modèles de *machine learning* utilisés. Bien que les algorithmes utilisés sur son service aient pour seul objectif d'apporter une aide aux bénévoles, la *Fondation Wikimédia* déclare que ses outils font l'objet de recherches académiques spécifiques à la mesure d'éventuels biais pour les modifier par la suite. *Microsoft* mentionne des « bonnes pratiques » guidant le déploiement de ses technologies pour *Bing*, notamment la mise en œuvre de processus et outils pour identifier les idées préconçues dans l'ensemble des données et algorithmes, le recours à l'évaluation humaine et la recherche et mise en œuvre d'outils et techniques d'analyse.

Comme pour les processus d'évaluation, l'Arcom estime qu'il est nécessaire que soient rendus publics les études et progrès réalisés en la matière mais également que les opérateurs rendent leurs données plus accessibles³⁸ afin que des experts extérieurs puissent évaluer d'éventuels biais dans la recommandation ou la modération de contenus.

³⁶ <https://about.fb.com/news/2021/11/civil-rights-audit-progress-report/>

³⁷ Source : déclaration de Facebook en 2021.

³⁸ Voir partie II-6 sur l'éducation aux médias et à l'information et les relations avec le monde de la recherche.

- **Des approches différentes des définitions de loyauté, explicabilité et équité des algorithmes**

Dans son questionnaire, l'Arcom a proposé des définitions des notions de loyauté, d'explicabilité et d'équité, élaborées avec l'appui de son comité d'experts sur la désinformation en ligne. Comme l'année passée, l'Arcom a appelé les opérateurs à les commenter ainsi qu'à indiquer comment ils les mettaient en œuvre. [Meta](#) et [Snap](#) n'apportent pas ou peu d'éléments complémentaires par rapport à l'année passée. [Pinterest](#), [Google](#), [Tiktok](#) et [Yahoo](#), quant à eux, n'ont pas commenté ces définitions.

[Microsoft](#) fait état de six principes régissant l'intelligence artificielle responsable : équité, fiabilité et sécurité, confidentialité, inclusivité, transparence et responsabilité, principes que [LinkedIn](#) déclare également prendre en compte. [Microsoft](#) propose une liste de bonnes pratiques pour mettre en œuvre ces principes. [Dailymotion](#) déclare que les critères et définitions sont pertinents, tout en précisant que les problématiques qui peuvent découler du non-respect de ces principes concernent majoritairement les plateformes en ligne mettant en avant des contenus amateurs dont l'audience se situe sur le service. Enfin, [Twitter](#) considère que le terme équité n'est pas nécessairement le plus pertinent car il impliquerait un jugement moral. L'équité n'étant pas, selon l'opérateur, défini de la même manière dans plusieurs cultures ou géographies, il préfère le terme « *partialité* » des algorithmes.

Préconisations

Au vu de ces éléments d'analyse, l'Arcom formule les préconisations suivantes :

- ❖ faire preuve d'une transparence plus importante sur le fonctionnement des outils de détection de contenus, comptes et comportements utilisés à des fins de modération vis-à-vis du public et communiquer au régulateur davantage d'éléments permettant d'en évaluer la pertinence ;
- ❖ communiquer aux utilisateurs les critères utilisés pour la recommandation de contenus, et ce, de manière personnalisée et contextuelle à l'aide d'outils et fonctionnalités accessibles directement sur le service ;
- ❖ améliorer la transparence des processus et moyens dédiés à l'évaluation des outils automatiques en publiant régulièrement, dans un espace dédié, des éléments sur les modifications qui ont été réalisées à l'aide d'analyses internes ou externes ;
- ❖ instaurer une transparence accrue des politiques en matière de réduction de la visibilité des contenus.

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

En 2021, les mesures en faveur de la diffusion d'une information issue de sources vérifiées sont, globalement, identiques à celles déclarées pour l'exercice précédent. Dès lors, la majeure partie de l'analyse et des conclusions formulées dans les précédents bilans restent d'actualité³⁹.

Les nouvelles mesures déclarées et les nouveaux moyens mis en œuvre l'ont majoritairement été dans le contexte particulier de l'année 2021 (COVID-19 et proximité du début de la période électorale française) et n'ont pas nécessairement été pérennisés.

[Webedia](#) déclare, cette année encore, ne pas prendre de mesures en la matière en raison des caractéristiques intrinsèques du forum [Jeuxvideo.com](#) dont les utilisateurs sont, selon lui, uniquement des personnes privées et non des organismes (notamment de presse).

Deux autres opérateurs déclarent, comme l'an dernier, sélectionner en amont les partenaires médias dont les contenus sont disponibles au sein d'une partie de leur service ([Snap](#), pour la partie *Discover* de son service, et [Yahoo](#) pour la partie *Actualités*) et estiment moins nécessaire, de ce fait, de recourir à des dispositifs de vérification de l'information. De même, [Dailymotion](#) se considère comme « *peu exposé* » aux risques liés à la manipulation de l'information et ne dispose donc pas de dispositifs de *fact-checking*, car ses algorithmes promeuvent les seules sources « *de confiance* ».

L'Arcom attire l'attention des opérateurs sur le fait que ces partenariats ne sont pas infaillibles, d'autant plus lorsque les partenaires ne se limitent pas aux seuls médias d'information (ex. : divertissement). En outre, ces partenaires ne sont pas toujours les seules sources de contenus ; ainsi, sur l'onglet *Discover* de [Snapchat](#), on trouve de nombreux contenus de créateurs « *Snap stars* » et d'utilisateurs privés populaires⁴⁰. Par ailleurs, quand bien même des pratiques de manipulation de l'information ne seraient pas directement observables sur un service, la présence de contenus relatifs à la vérification de faits reste bénéfique.

S'agissant des deux opérateurs nouvellement déclarants, [TikTok](#) évoque succinctement trois dispositifs de promotion d'une information de source vérifiée, tandis que [Pinterest](#) ne traite que marginalement ce sujet.

De manière générale, l'Arcom déplore qu'aucun opérateur n'ait fourni l'ensemble des données chiffrées et indicateurs relatifs à son activité en la matière sur le territoire national en 2021. À titre d'exemple, l'absence de données chiffrées concernant le nombre de contenus soumis à la vérification de *fact-checkers* rend l'évaluation de l'engagement effectif des opérateurs dans la lutte contre la manipulation de l'information par l'Arcom et les lecteurs du présent bilan nécessairement partielle. L'Arcom invite l'ensemble des opérateurs à prendre en compte ces considérations en vue de la prochaine édition de ce bilan annuel et salue le fait que le code européen de bonnes pratiques renforcé contre la désinformation prévoit des engagements relatifs à la publication d'indicateurs et données chiffrés en matière de *fact-checking*⁴¹.

Enfin, l'Arcom déplore l'absence d'informations concernant la promotion d'une information de source vérifiée sur [YouTube](#) dans la déclaration transmise par [Google](#) pour l'année 2021, à l'instar des années précédentes.

³⁹ Le modèle de [Wikipédia](#) en fait un acteur à part, le sourçage de l'information étant obligatoire sur la plateforme et la vérification au cœur du travail des contributeurs.

⁴⁰ Les créateurs « *Snap stars* » sont « *certifiés* » par l'opérateur alors que les utilisateurs privés ne le sont pas.

⁴¹ Cf. engagements n°30 à 33.

3.1 Les nouvelles mesures relatives à l'identification des sources des contenus

L'identification des sources de contenus fournit aux utilisateurs des éléments de contexte (caractère professionnel ou amateur du contenu, informations importantes à connaître, etc.). Pour cela, les opérateurs procèdent de différentes manières, notamment par un dispositif de certification plus large que les seuls médias ou l'identification des médias spécifiques (ex. : médias affiliés à un État).

Dans un contexte marqué par une instabilité géopolitique forte, les efforts visant à une meilleure **identification des médias affiliés à un État**, entendu comme ceux ne bénéficiant pas des garanties d'indépendance éditoriale nécessaires vis-à-vis de l'État qui les contrôle, ont été amplifiés.

Depuis juin 2020, [Meta](#) identifie les médias « *contrôlés par l'État* », sur [Facebook](#) et [Instagram](#). Les organisations estimant avoir été marquées à tort peuvent envoyer une réclamation *via* un formulaire de réclamation spécifique.

[Twitter](#), pour sa part, avait dès août 2020 élargi les types de comptes politiques labellisés comme « *médias affiliés à un État* » et décidé de ne plus amplifier leur visibilité ni celle de leurs tweets par le biais de ses systèmes de recommandation. En outre, depuis cette date, ces comptes ne peuvent plus effectuer de publicité sur la plateforme. Néanmoins, l'opérateur souligne le risque d'extension injustifiée des labels d'affiliation à un État. Pour le minimiser, il indique être en train d'étudier un certain nombre de partenariats avec des entités tierces indépendantes.

Depuis avril 2019, les vidéos diffusées sur [YouTube](#) en France provenant de « *médias bénéficiant d'un financement public ou gouvernemental* » font l'objet d'une labellisation⁴². En revanche, aucun dispositif spécifique d'identification de ces mêmes médias n'existe, à cette date, sur [Google Search](#). Il en va de même selon leurs déclarations de [Dailymotion](#), [LinkedIn](#) et [Snap](#). [Microsoft](#), [Pinterest](#) et [Yahoo](#), quant à eux, n'ont pas répondu à l'objet de cette question.

Par ailleurs, [TikTok](#) indique travailler actuellement au développement d'une politique dédiée aux « *médias contrôlés par un État* ». Pour ce faire, l'opérateur a collaboré avec un panel d'experts pluridisciplinaires internationaux. Néanmoins, il ne souhaite pas partager davantage d'informations à ce stade sur le déploiement de cette nouvelle politique. L'Arcom a constaté, en septembre 2022, une mise en œuvre sélective de cette règle pourtant non finalisée. À titre d'exemple, le compte de *China Global Television Network Europe* (CGTN Europe) n'est pas identifié comme tel.

L'Arcom encourage vivement les opérateurs à mettre en place des dispositifs d'identification des médias qui ne bénéficient pas des garanties d'indépendance éditoriale nécessaires vis-à-vis de l'État qui les contrôle, et à les appliquer de façon homogène, transparente et selon des critères objectifs.

Des mesures visant à **mettre en avant les sources faisant autorité** ont également vu le jour.

[Google](#) détaille le fonctionnement des capsules d'informations qui apparaissent en haut à droite à la suite d'une requête sur le moteur de recherche. Les informations qui y sont présentées proviennent du *Google Knowledge Graph*, une base de données alimentée par diverses sources : données publiques, contrats de licence, transmissions directes par les propriétaires de contenus, etc. Le fait que les utilisateurs puissent signaler une erreur sur une information donnée *via* un dispositif dédié présentant un certain niveau de granularité

⁴² Source : <https://france.googleblog.com/2019/04/plus-de-moyens-pour-trouver-des.html>

est un moyen supplémentaire permettant d'assurer la vérification des informations fournies.

De son côté, *Twitter* a relancé en 2021, après consultation de ses utilisateurs⁴³, la possibilité pour tout utilisateur de faire une demande de certification de son compte par le biais d'un nouveau processus accessible à tous (ce dispositif de vérification de l'authenticité d'un compte ne se limite donc pas aux seuls médias). La demande fait l'objet d'un examen conjuguant des processus automatisés et humains. Les règles d'attribution et de retrait de la certification sont dorénavant publiques⁴⁴, ce qui est à saluer.

Dans son précédent bilan sur la lutte contre la manipulation de l'information, l'Arcom avait préconisé de maintenir une intervention humaine dans la vérification des informations fournies au sujet des organes de presse lorsqu'ils font l'objet d'une identification particulière. *Snap* et *Dailymotion* indiquent appliquer d'ores et déjà cette recommandation. *Meta* indique (i) que son processus d'évaluation implique déjà une intervention humaine, (ii) s'interroger quant à la pertinence de son imposition systématique au regard du volume de publications postées chaque jour sur ses plateformes et (iii) peiner à saisir ce que recouvre exactement le terme « intervention humaine ». En réponse, l'Arcom rappelle que la préconisation visait les informations fournies par la plateforme sur les organes de presse faisant l'objet d'une identification particulière, et non les informations fournies par ces organes de presse. L'Arcom regrette l'absence de commentaires émanant des autres opérateurs.

3.2 Les nouveaux dispositifs relatifs à la mise en avant des sources d'autorité et des contenus vérifiés

Trois grands axes se dessinent en la matière : la mise en avant de sources d'autorité, la labellisation des contenus ou des sources et les programmes de subventionnement et d'aide.

- ***La mise en avant des sources d'autorité et de contenus vérifiés***

S'agissant des moteurs de recherche tout d'abord, *Microsoft* a poursuivi le déploiement, initié fin 2017, d'un élément visuel sur *Bing*, le label « *Fact Check* », pour aider les utilisateurs à trouver des informations de *fact-checking* sur les actualités, les principaux articles et pages Internet dans les résultats de recherche⁴⁵. Après vérification sur un échantillon restreint de requêtes, il semblerait qu'à ce stade, ce label ne soit apposé que sur un nombre minoritaire de contenus en français.

Dans le même esprit, en avril 2022, *Yahoo* a ajouté un nouvel encart « *Fact Check* » sur la page d'accueil de son service en français avec des contenus dédiés à la vérification des faits fournis par ses partenaires, l'AFP et 20 Minutes. L'Arcom salue cette initiative.

Il convient de noter que *Google* proposait déjà de telles fonctionnalités (label et/ou espace et/ou outil dédiés) sur son moteur de recherche comme sur *YouTube*⁴⁶.

En octobre 2021, *Meta* a annoncé le lancement en France, effectif à partir du 15 février 2022, de « Facebook News », onglet qui rassemble les actualités provenant d'un éventail

⁴³ Ce dispositif de vérification de l'authenticité d'un compte avait été suspendu en 2017 car jugé « *arbitraire* » et « *déroutant* » par de nombreux utilisateurs (source : https://blog.twitter.com/fr_fr/topics/product/2020/certification).

⁴⁴ Source : <https://help.twitter.com/fr/managing-your-account/about-twitter-verified-accounts>

⁴⁵ Source : <https://blogs.bing.com/Webmaster-Blog/September-2017/Bing-adds-Fact-Check-label-in-SERP-to-support-the-ClaimReview-markup>

⁴⁶ Sources : <https://support.google.com/websearch/answer/7315336?hl=fr> et <https://toolbox.google.com/factcheck/explorer> et <https://support.google.com/youtube/answer/9229632?hl=fr>

de sources d'informations jugées fiables⁴⁷. Les articles sont sélectionnés par un mélange de moyens algorithmiques et humains. Dans ce cadre, *Meta* a noué un partenariat avec Media Services, filiale de l'AFP, pour procéder à cette sélection⁴⁸.

Ces différentes démarches semblent constituer un moyen efficace de lutte contre la désinformation. La part de contenus concernés reste néanmoins très largement minoritaire. De même, si la création d'espaces dédiés sur certains services s'inscrit dans une dynamique louable, l'absence de données chiffrées sur la fréquentation de ceux-ci altère considérablement l'analyse de l'efficacité d'une telle démarche.

- **La labellisation des contenus ou des sources**

En 2021, *Microsoft* a renouvelé pour trois ans et étendu son accord de licence avec *NewsGuard*. Ce partenariat a également abouti au lancement de quinze campagnes de sensibilisation contre la manipulation de l'information et pour une meilleure connaissance et utilisation des médias en ligne.⁴⁹ En outre, selon l'opérateur, ce partenariat lui permet de combler son manque de légitimité pour décider « du vrai et du faux » en matière d'identification et de labellisation des sources.

- **Les programmes de subventionnement et d'aide**

Meta a lancé en septembre 2021 un programme mondial de subventions doté d'un million de dollars destiné à des organisations œuvrant contre la manipulation de l'information climatique, le « *Climate Misinformation Grant* », en partenariat avec l'*International Fact-Checking Network* (IFCN). En France, *Science Feedback* a pu en bénéficier. Le projet de cette organisation consiste en la création d'un portail où les vérificateurs de faits pourront soumettre des articles liés au changement climatique afin d'obtenir de l'aide et de la revue d'experts⁵⁰.

En janvier 2021, *Google* a, quant à lui, annoncé l'ouverture d'un fonds mondial de trois millions de dollars visant à lutter contre la manipulation de l'information relative au vaccin contre la Covid-19. Dans ce cadre, *Google* a notamment sélectionné un projet de partenariat entre *Aleteia* (France) et *Verificat* (Espagne) visant à créer une base de données de vérifications des faits disponible en sept langues pour les médias catholiques internationaux.

Par ailleurs, afin de lutter contre la manipulation de l'information à l'approche de la période électorale française de 2022, *Google* déclare avoir lancé, en décembre 2021, le programme « Objectif Désinfox » en partenariat avec l'AFP afin de permettre aux journalistes d'être plus efficaces dans leurs recherches, leurs veilles et leurs mécanismes de vérification.

En outre, même si elle considère *Wikipédia* moins concernée par ces sujets que les autres plateformes (cf. *supra*), la *Fondation Wikimédia* met en œuvre des initiatives notables en la matière. Depuis 2019, elle développe ainsi le programme « *Knowledge Integrity* » afin de créer une synergie entre contributeurs et algorithmes. En 2021, celui-ci s'est notamment concrétisé par la création de « *WikiCheck* » (un prototype d'API pour une vérification automatique des faits), de « *Wiki-Reliability* » (ensemble d'articles *Wikipédia*

⁴⁷ Sources : <https://about.fb.com/fr/news/2021/10/facebook-france-et-lalliance-de-la-presse-de-linformation-sassocient-pour-renforcer-l'experience-de-l'actualite-pour-les-utilisateurs-et-les-editeurs-en-france/> et <https://about.fb.com/fr/news/2022/02/lancement-en-france-de-facebook-news-un-espace-dedie-a-l'actualite-sur-facebook/>

⁴⁸ Néanmoins, en octobre 2022, l'AFP a annoncé que *Meta* mettait fin à son contrat avec Media Services à compter du 1^{er} janvier 2023. Les contenus répertoriés par « Facebook News » seront sélectionnés selon une méthode de curation uniquement algorithmique. Les contrats passés avec les éditeurs de presse repris sur cet espace restent, quant à eux, effectifs jusqu'à la fin de l'année 2024.

Source : <https://www.france24.com/fr/info-en-continu/20221103-facebook-news-meta-va-cesser-sa-collaboration-avec-media-services-filiale-de-l-afp>

⁴⁹ Source : <https://www.newsquardtech.com/press/newsquard-expands-agreement-with-microsoft/>

⁵⁰ Source : <https://www.facebook.com/formedia/blog/climate-misinformation-grant>

annotés comme rencontrant divers problèmes liés à la fiabilité du contenu) et du « *Wikipedia Knowledge Integrity Risk Observatory* » (observatoire multidimensionnel). Par ailleurs, la *Fondation Wikimedia* a créé une « *bibliothèque Wikipédia* », en collaboration avec plus de 75 éditeurs (ex. : ProQuest, EBSCO, etc.). Celle-ci permet aux contributeurs actifs d'accéder à des sources d'informations fiables qui ne sont autrement disponibles que par le biais d'un abonnement payant.

3.3 L'exercice de fact-checking sur les plateformes elles-mêmes

Face à un contenu potentiellement problématique, les opérateurs ont schématiquement deux modes d'action sensiblement différents : certains font appel à des entités tierces indépendantes pour les aider à vérifier l'information (ex. : *Snap*, *LinkedIn*), d'autres les laissent intervenir directement sur leur service (ex. : *Meta*, sauf en présence d'un discours à caractère politique).

En revanche, la réduction de la visibilité et de la potentielle diffusion d'un contenu, voire sa suppression et le prononcé de sanctions à l'encontre du compte qui en est à l'origine, est une mesure communément mise en place par les opérateurs eux-mêmes lorsqu'il est avéré que celui-ci contient une fausse information.

Meta a développé (i) son programme de vérification des informations, initié en 2016, sur la base des travaux de l'IFCN et (ii) le « *Meta Journalism Project* », lancé en 2017 pour travailler plus étroitement avec cet écosystème⁵¹. Pour ce faire, *Meta* a poursuivi son association avec des partenaires de vérification tiers et indépendants à l'échelle mondiale, certifiés par l'IFCN.

LinkedIn indique soumettre dorénavant des questions spécifiques à des vérificateurs de faits externes tels que *Reuters*. Cette relation est encadrée contractuellement et donne lieu à une rémunération.

Snap, pour sa part, collabore avec le *Poynter Institute*. Toutefois, l'opérateur précise que qu'aucun contenu porteur de fausses informations n'a été constaté à ce stade en France dans le cadre de ce partenariat.

TikTok indique travailler avec treize organismes de vérification des faits (dont, en France, l'AFP et *Science Feedback*⁵²), dans 33 langues différentes et 64 pays à travers le monde. Tous ses partenaires sont accrédités par l'IFCN.

Pinterest déclare lutter contre la manipulation de l'information par une approche hybride conjuguant examens manuel et automatisé des contenus. De ce fait, l'opérateur travaille avec des acteurs du secteur ainsi qu'avec des organisations d'experts, des organismes gouvernementaux et des organismes sans but lucratif afin de bénéficier de leurs observations sur les moyens mis en œuvre sur son service.

Concernant *Twitter*, son recours aux tiers vérificateurs n'est toujours pas systématique. L'opérateur indique en effet s'associer à des organisations en période de « *risque aigu* » seulement (ex. : *EU DisinfoLab* en période électorale), en s'appuyant sur une définition fournie par l'ONU que l'opérateur n'a pas précisée⁵³.

Outre la dégradation de la visibilité du contenu et la mise en place d'interstitiels lorsque la vérification est en cours, *TikTok* a décidé d'informer ses utilisateurs quant à l'absence de vérification des faits exposés par un contenu, notamment lorsque celui-ci se rapporte à un événement en direct. Ainsi, la fonctionnalité « Message pour contenu non vérifié », lancée

⁵¹ Source : <https://www.facebook.com/formedia/mjp/timeline>

⁵² Source : audition de TikTok en septembre 2022.

⁵³ Id.

en mars 2021, a pour but d'informer l'utilisateur lorsqu'est identifiée une vidéo avec du contenu non corroboré et de l'inviter à reconsidérer son partage⁵⁴.

D'autres opérateurs testent également des fonctionnalités visant à inviter leurs utilisateurs à se questionner avant de partager et/ou interagir avec des contenus. Depuis 2020, *Twitter* affiche ainsi une bannière invitant l'utilisateur à reconsidérer son action (« retweet ») vis-à-vis d'un contenu qu'il n'aurait pas encore ouvert via *Twitter*.⁵⁵ Dans la même perspective, depuis mai 2021, certains utilisateurs de *Facebook* se voient présenter un message les invitant à reconsidérer leur action au moment de partager un contenu qu'ils n'auraient pas encore ouvert sur la plateforme⁵⁶.

L'Arcom salue et encourage la mise en place de telles mesures qui permettent de freiner la diffusion massive des fausses informations en ligne tout en incitant les utilisateurs à prendre davantage conscience de leur rôle dans cette viralité.

Enfin, concernant les actions futures et nouvelles fonctionnalités envisagées à l'avenir en la matière, *Microsoft* a indiqué travailler au lancement de *Search Coach* (outil de recherche intégrant les données de *NewsGuard* et qui vise à améliorer les compétences numériques des élèves) et à l'intégration des données de *NewsGuard* à MSN (son portail web).

Préconisations

Au vu de ces éléments d'analyse, l'Arcom formule les préconisations suivantes :

- ❖ mettre en place des dispositifs spécifiques d'identification des médias qui ne bénéficient pas des garanties d'indépendance éditoriale nécessaires vis-à-vis de l'État qui les contrôle et les appliquer de façon homogène, transparente et selon des critères objectifs ;
- ❖ mettre en place des dispositifs permettant d'ajouter des « frictions » dans la diffusion de contenus potentiellement problématiques ;
- ❖ développer le recours aux dispositifs (i) de fact-checking et (ii) de mise en avant des contenus en résultant, y compris hors période électorale.

⁵⁴ Source : <https://newsroom.tiktok.com/fr-fr/de-nouvelles-fonctionnalites-pour-encourager-les-utilisateurs-a-reconsiderer-leur-partage>

⁵⁵ Source : <https://twitter.com/TwitterSupport/status/1270783537667551233?>

⁵⁶ Source : <https://twitter.com/MetaNewsroom/status/1391816265891778560?s=20>

4. Lutte contre les comptes propageant massivement de fausses informations

La loi du 22 décembre 2018 invite les opérateurs de plateforme à prendre des mesures pour lutter contre les « *comptes propageant massivement de fausses informations* ». Cette notion couvre un ensemble de pratiques de manipulation de l'information à grande échelle. Elle est, par nature, protéiforme et son appréhension peut différer en fonction des caractéristiques de la plateforme.

Dans sa recommandation du 15 mai 2019, l'Arcom encourage les opérateurs à mettre en place des procédures appropriées et proportionnées⁵⁷ pour assurer leur détection, des dispositifs publics de suivi et de statistiques relatifs à la détection et au traitement de ces comptes et un espace d'information aisément accessible renseignant les utilisateurs sur les pratiques susceptibles d'entraîner une intervention.

4.1 Des différences d'approches rendant complexe l'analyse comparative des moyens mis en place

L'Arcom a demandé aux opérateurs de faire part dans leur déclaration, d'une part, de leur appréhension de la notion de « *compte propageant massivement des fausses informations* » et, d'autre part, et des différentes pratiques de manipulation de l'information observées sur leur service.

• Une appréhension hétérogène de la notion

L'Arcom relève une certaine frilosité chez une partie des déclarants qui évitent d'aborder pleinement ce sujet.

[TikTok](#)⁵⁸ et [Pinterest](#) se sont bornés à expliquer leurs politiques en la matière et ont occulté les questions portant sur les pratiques observées sur leur service, limitant ainsi fortement l'intérêt de leur réponse sur ce point.

Quatre autres opérateurs, s'estimant non concernés en raison de leur modèle, ne répondent à aucune des questions posées de l'Arcom, quand bien même certaines restent pertinentes.

[Snap](#) indique que la conception de sa plateforme empêche la diffusion massive de contenus. Il serait cependant utile que sa déclaration mentionne si des faux comptes de *bots* et/ou des *deepfakes* ont été identifiés sur le service, même de façon résiduelle.

[Microsoft](#) et [Yahoo](#) déclarent ne pas identifier de pratiques de manipulation de leur service, leurs utilisateurs ne pouvant publier du contenu. L'Arcom relève néanmoins que des pratiques de manipulation peuvent exister sur les moteurs de recherche, comme en atteste la déclaration de [Google](#) (voir ci-après).

[Dailymotion](#) n'estime pas « *nécessaire de mettre en place des mesures concernant le partage « massif » de fausses informations* » au motif de son choix stratégique de cibler davantage les professionnels, ce qui rendrait la plateforme, par essence, moins propice à la diffusion de fausses informations.

L'Arcom estime qu'un travail de définition commune des pratiques existantes de manipulation favoriserait la transparence sur ces phénomènes. Elle considère comme essentiel que les opérateurs poursuivent ou, pour certains, renforcent leurs efforts en la

⁵⁷ Sur le caractère proportionné des mesures prises pour lutter contre les comptes propageant massivement de fausses informations, le Conseil constitutionnel avait précisé que : « *Il appartiendra, en tout état de cause, au juge éventuellement saisi d'un litige sur les autres mesures complémentaires susceptibles d'être adoptées à cette même fin, notamment celles visant à lutter contre les comptes propageant massivement de fausses informations, d'examiner, dans chaque cas, si elles sont nécessaires, adaptées et proportionnées à l'objectif poursuivi* » (CC, décision n° 2018-773 DC du 20 décembre 2018, paragraphe 87).

⁵⁸ [TikTok](#) a uniquement mentionné des chiffres sur les « comptes spams » et faux comptes (au niveau mondial en décembre 2021) et a indiqué « améliorer sa rapidité et sa réaction face aux menaces qui évoluent. »

matière. À cet égard, l'Arcom note avec grand intérêt que les signataires du code européen de bonnes pratiques renforcé contre la désinformation se sont engagés à coopérer pour élaborer une liste exhaustive des tactiques, techniques et procédures (TTPs) et à revoir périodiquement celle-ci, afin de mettre en œuvre des politiques claires en réponse. La notion d'atteinte à l'intégrité du service employée dans le code est d'ailleurs de nature à éclairer le travail de définition que l'Arcom encourage.

- **Des descriptions de pratiques identifiées intéressantes mais qui gagneraient à être étoffées**

Seuls [Meta](#), [Twitter](#), [Google](#), [Webedia](#) (et dans une moindre mesure [LinkedIn](#) et la [Fondation Wikimédia](#)) font état de leur approche de la notion de « *comptes diffusant massivement de fausses informations* » ou des pratiques qu'ils observent sur leur service.

[Twitter](#) interdit d'utiliser les fonctionnalités de son service d'une manière destinée à amplifier artificiellement des informations ou d'adopter un comportement visant à manipuler les autres utilisateurs. Il définit la manipulation sur la plateforme comme « *toute utilisation à des fins d'activité massive, agressive ou trompeuse qui induit les autres en erreur et/ou perturbe leur expérience* ». Les pratiques qu'il observe sont le *spam* à motivation commerciale, les engagements inauthentiques qui simulent la popularité d'un compte ou contenu, les activités coordonnées pour influencer artificiellement les conversations et les activités coordonnées qui encouragent un comportement allant à l'encontre des règles d'utilisation.

Pratiques pouvant porter atteinte à l'intégrité des services identifiées par les opérateurs et mentionnées dans leur déclaration⁵⁹

Plateforme	Faux comptes	Bots	Deepfakes	Comportements inauthentiques coordonnés	Campagnes d'influence étrangères
Twitter					
YouTube					
Facebook					
Instagram					
LinkedIn					
Google Search					
Jeuxvideo.com					
TikTok		60			
Wikipédia					
Bing Daumotion Pinterest Snapchat Yahoo Media	Pas d'information pertinente				

⁵⁹ Les cases en bleu correspondent aux pratiques ou phénomènes détectés sur le service des opérateurs.

⁶⁰ Lors d'une audition menée en septembre 2022, l'opérateur a précisé n'avoir identifié aucun « *bot* malveillant ».

[Webedia](#) ne donne pas de définition des comptes propageant massivement de fausses informations mais déclare que la notion doit être appréciée en fonction des spécificités de chaque service. Sur [Jeuxvideo.com](#), il n'existe pas de critère spécifique pour déterminer la « diffusion massive » mais un faisceau d'indices (fil de discussion contenant des milliers de messages, thème abordé de manière répétitive, contenu diffusé à plusieurs reprises, etc.) pour identifier les pratiques problématiques. L'opérateur déclare que, dans les faits, son service en est peu victime.

[Meta](#) renvoie à sa déclaration de l'année passée où il constatait qu'un grand nombre de fausses informations étaient partagées dans l'objectif de réaliser des profits. En plus des comportements identifiés et interdits sur son service (faux comptes, spams, fausses informations contribuant à la pression sur les électeurs et vidéos manipulées), il distingue deux pratiques de diffusion massive :

- les comportements inauthentiques coordonnés dans le cadre de campagnes nationales non gouvernementales (groupes de comptes trompant sur leur identité et activités, en s'appuyant sur de faux comptes) ;
- les ingérences étrangères ou gouvernementales via des campagnes d'influence menées au nom d'une entité gouvernementale ou par un acteur étranger.

[Google](#) renvoie à sa déclaration de 2021 où il mentionnait les pratiques identifiées et interdites : usurpation d'identité, création abusive de comptes, tentatives de manipulation artificielle du taux d'engagement (comptes de vues sur [YouTube](#) ou manipulation des systèmes de classement sur [Google Search](#)) et campagnes d'influence.

• ***Focus : la diffusion de fausses informations à travers les fonctionnalités de diffusion de contenus en direct***

Dans la liste des pratiques identifiées sur son service, [Meta](#) a ajouté cette année la mention des fausses informations dans les contenus « en direct ». Ces contenus sont soumis aux standards de la communauté [Facebook](#) et aux règles de la communauté [Instagram](#) et sont ainsi susceptibles de recevoir les mêmes restrictions que les autres types de contenus : étiquetage, baisse de la visibilité ou retrait. L'outil de signalement est disponible en direct. En outre, deux fonctionnalités et outils sont offerts aux créateurs de contenus en direct : l'outil « *live moderator* » pour [Instagram](#), permettant de nommer des modérateurs durant les contenus *live*⁶¹ et la fonctionnalité « *moderation assist*⁶² » pour [Facebook](#)⁶³.

[Twitter](#) rappelle dans sa déclaration que les règles du service s'appliquent intégralement à tous ses produits et services, y compris Twitter Spaces (fonctionnalité de *live audio*). Il propose également des outils de modération pour les « hôtes » des *spaces*, dont le contenu peut être signalé en direct. Il précise que 80% à 90% des contenus problématiques sont modérés automatiquement⁶⁴.

[Google](#) n'a apporté aucune information à ce sujet pour les fonctionnalités *live* sur [YouTube](#).

À l'exception de [Meta](#) et [Twitter](#), l'Arcom note que les opérateurs n'ont pas fait état de moyens humains ni d'outils de modération automatique consacrés spécifiquement à la modération des contenus pendant leur diffusion en direct alors même que plusieurs événements tragiques ont démontré l'importance d'une telle modération.

⁶¹ https://help.instagram.com/1175145849979325/?helpref=related_articles

⁶² <https://www.facebook.com/business/help/545793280468485/>

⁶³ Précisions apportées lors de l'audition du 19 septembre 2022.

⁶⁴ Source : audition de Twitter, septembre 2022.

4.2 Une information partielle sur le fonctionnement des outils d'identification des pratiques de propagation massive

- ***Une transparence perfectible sur le fonctionnement des nombreux outils d'identification des pratiques de propagation massive***
 - **Concernant la lutte contre les faux comptes et bots**

Les moyens et mesures mentionnés ci-après concernent la détection des *bots* et des comptes trompeurs à l'origine d'une grande partie de contenus nuisibles sur les services. Il convient de préciser que parmi ces comptes, tous ne participent pas nécessairement à la propagation massive de la désinformation.

LinkedIn identifie les faux comptes grâce à des systèmes automatisés fonctionnant par analyse de similitudes avec des comptes et contenus déjà modérés, par détection d'anomalies et par suivi des séquences d'activités associées à une automatisation abusive.

Twitter utilise également des outils de détection automatisés s'appuyant sur des signaux comportementaux, à savoir la façon dont les comptes se comportent et réagissent les uns aux autres, l'absence d'adresse électronique confirmée, l'inscription simultanée de plusieurs comptes, des comptes qui publient et mentionnent de manière répétée des comptes qui ne les suivent pas, ou tout comportement qui pourrait indiquer une attaque coordonnée. Pour vérifier que certains comptes ne sont pas automatisés, l'opérateur précise également utiliser des outils de reCAPTCHA⁶⁵.

Webedia identifie la diffusion massive de fausses informations au moyen de la modération humaine et de l'utilisation systématique d'un « Captcha » pour confirmer qu'un utilisateur est bien humain lors de la publication de son premier message sur le forum.

TikTok déclare investir dans des dispositifs de défense automatisés pour détecter, bloquer et supprimer les comptes et engagements inauthentiques, sans apporter aucune information sur les moyens et outils utilisés.

Meta indique utiliser des méthodes de détection essentiellement automatisées mais apporte peu d'informations sur leur fonctionnement, à l'exception des mentions pour exemple, dans la déclaration de l'année dernière, du blocage de certaines adresses IP créant plusieurs comptes simultanément et de la détection d'actions suspectes telles que les adresses électroniques suspectes ou l'identification automatique des comportements similaires aux comptes déjà supprimés.

Il est intéressant de noter que la plupart des opérateurs déclarent que l'immense majorité des faux comptes sont détectés avant même un signalement d'un utilisateur (au niveau mondial, 99,7% pour *Meta*, 99% dès l'enregistrement pour *LinkedIn*).

L'Arcom relève ainsi que les opérateurs, à l'exception de *Webedia*, font principalement état d'outils automatisés de détection, sans nécessairement évoquer une intervention humaine dans ces processus d'identification.

Elle regrette vivement, là-encore, le manque d'informations et d'indicateurs permettant d'évaluer la pertinence des mesures et moyens mis en place. En effet, si certains chiffres ont été communiqués, notamment sur le nombre de comptes supprimés (au niveau mondial, 152 118 769 « comptes spams » pour *TikTok* au 4^e trimestre 2021, 16,4 millions de faux comptes pour *LinkedIn* au 2^e semestre 2021, 1,6 milliard au 1^{er} trimestre 2021

⁶⁵ Le reCAPTCHA est un système de détection automatisé proposé par Google. Il se base sur des tests de CAPTCHA qui sont des authentications automatisées correspondant à la demande à un utilisateur d'identifier des parties d'une image, de taper des mots affichés à l'écran ou à une demande de réinitialisation de mot de passe.

pour *Facebook* et *Instagram*), l'analyse comparative ne saurait être pertinente qu'avec une granularité plus fine et sur des échelles de temps équivalentes.

- **Concernant la lutte contre les *deepfakes***

Les hypertrucages ou *deepfakes* sont entendus comme les contenus (images, audio ou vidéo) modifiés à l'aide d'une intelligence artificielle. Ils peuvent être créés et/ou utilisés dans le but de tromper le public, mais ne sont pas nécessairement liés à des pratiques de manipulation de l'information.

L'Arcom remarque, d'une part, que les opérateurs n'ont pas apporté de réponse aux demandes de définition des *deepfakes* (sauf *Meta* dans sa déclaration de l'année passée) et, d'autre part, que l'ensemble des éléments déclarés en la matière concernent uniquement les hypertrucages vidéo. Par ailleurs, si certains opérateurs (*Microsoft*, *Meta*, *Google*⁶⁶) ont mentionné des initiatives prises pour contribuer à des jeux de données en open source permettant d'entraîner les algorithmes de détection, peu d'informations ont été données quant aux outils de détection eux-mêmes.

LinkedIn indique avoir établi des partenariats avec *Microsoft* pour développer des solutions et utiliser des sites tiers de vérification des faits pour traiter les signalements de contenus comme étant des hypertrucages.

Microsoft a développé deux outils : l'un, appelé *Microsoft Video Authenticator* et créé sur la base d'un ensemble de données publiques, fourni un pourcentage de chance de manipulation artificielle d'un contenu après son analyse. L'autre permet à un producteur de contenu d'ajouter une empreinte numérique pour que les utilisateurs, grâce à une extension, puissent en vérifier l'authenticité.

L'année passée, *Meta* expliquait détecter les hypertrucages en entraînant ses outils d'apprentissage sur des *deepfakes* créés par réseaux adversaires génératifs (GANs)⁶⁷.

- **Concernant la lutte contre les comportements inauthentiques coordonnés et les campagnes d'influence étrangère**

Google mentionne les actions de son groupe d'analyse des menaces (« Threat Analysis Group » - TAG) qui surveille plusieurs centaines de groupes d'acteurs malveillants soutenus par des gouvernements, se concentrant notamment sur l'identification et la résolution de problèmes tels que les campagnes d'hameçonnage et de piratage.

Twitter indique avoir également détecté de nombreuses campagnes d'influence, mais sans expliquer comment elles ont été identifiées ni les moyens qui sont alloués à cette détection. Il précise néanmoins les enseignements qui ont été tirés de ces campagnes (résumés dans un « document de position »⁶⁸) et mentionne le lancement, début 2022, du Twitter Moderation Research Consortium (TMRC)⁶⁹. Seul *Twitter* a indiqué le nombre de réseaux et comptes diffusant massivement de fausses informations détectés en France en 2021 (12 réseaux et 3 883 comptes).

Enfin, *Meta* déclare avoir constitué depuis cinq ans une équipe interdisciplinaire spécialisée dans la détection et la neutralisation d'opérations d'influence. Cette équipe de plus de 200 personnes a identifié, en 2021, 52 réseaux dans le monde qui se sont engagés dans des

⁶⁶ Source : déclaration de Google 2021.

⁶⁷ Aussi appelés réseaux antagonistes génératifs, ils correspondent à des modèles d'apprentissages profonds « génératifs » de contenus qui ont pour intérêt l'auto-évaluation permettant d'améliorer la qualité et le réalisme des trucages.

⁶⁸ <https://cdn.cms-twdigitalassets.com/content/dam/about-twitter/en/our-priorities/open-internet.pdf>

⁶⁹ Initiative définie comme la création d'« un groupe mondial d'experts issus du monde universitaire, de la société civile, des ONG et du journalisme, qui étudiera les questions de gouvernance des plateformes ».

campagnes coordonnées pour « *manipuler ou corrompre le débat public dans un but stratégique* ».

L'Arcom remarque que les opérateurs ne font pas systématiquement de différenciation, au sein de leurs déclarations, entre les moyens mis en œuvre pour lutter contre les campagnes d'influence étrangère et ceux déployés contre les comportements inauthentiques coordonnés, alors que ces derniers ne sont pas nécessairement liés à des campagnes d'influence étrangère. En effet, les usages et objectifs des comportements inauthentiques coordonnés sont divers : ils peuvent aussi bien recouvrir des pratiques purement lucratives (ex. : gonflement artificiel de la visibilité d'un contenu) qu'avoir pour objectif d'influencer l'opinion en période électorale (ex. : simulation d'un engagement spontané).

À cet égard, la déclaration de [Meta](#) définit deux « protocoles » impliquant une coordination d'utilisateurs sur un service tiers (souvent moins modéré) en vue d'une action conjointe sur la plateforme : le « *brigading* » (coordination en vue d'actions répétitives dans le but d'harcéler un utilisateur ou de le réduire au silence) et le « *mass reporting* » (coordination en vue de notifier simultanément un compte ou un contenu pour obtenir son retrait de manière injustifiée). L'opérateur ne fournit néanmoins pas d'informations complémentaires sur les moyens d'identification de ces pratiques, à l'exception de la mention de l'équipe interdisciplinaire précitée.

- ***Une absence d'éléments sur le lien entre ces pratiques et les recettes publicitaires potentielles***

L'Arcom déplore une absence d'informations en matière de recettes publicitaires liées aux pratiques portant atteinte à l'intégrité des services. Elle rappelle la préconisation de l'Arcom dans le précédent bilan annuel de « *faire preuve de davantage de transparence, à tout le moins à l'égard du régulateur, sur les recettes publicitaires potentielles générées par les comptes propageant massivement de fausses informations* ».

- ***Quelques mentions de coopérations avec d'autres opérateurs ou avec des tiers***

[Webedia](#) indique coopérer régulièrement avec les autorités administratives et judiciaires et ne pas avoir engagé de discussions avec d'autres opérateurs.

Outre son équipe d'experts mentionnée précédemment, [Meta](#) déclare discuter régulièrement avec le service Viginum et l'Ambassadeur pour le numérique.

[Twitter](#) précise rendre accessible l'archive des contenus relatifs aux opérations de désinformation étrangère afin de collaborer avec la société civile.

L'Arcom constate qu'aucune déclaration ne fait état d'éventuelles coopérations entre opérateurs.

L'Arcom note un manque de transparence sur les coopérations mises en place, plus particulièrement au niveau national. Elle renouvelle ses préconisations des années passées sur la nécessité de créer des canaux d'échanges entre les opérateurs, comme s'y sont d'ailleurs engagés les signataires du code européen de bonnes pratiques renforcé contre la désinformation, et de rendre publiques des informations en la matière.

4.3 Une hétérogénéité dans les informations communiquées aux utilisateurs

- ***De nombreuses informations quant aux pratiques interdites sur les services des opérateurs***

La recommandation de l'Arcom de 2019 encourage les opérateurs à mettre en place un espace d'information aisément accessible renseignant les utilisateurs de manière claire et précise sur les pratiques susceptibles d'entraîner une intervention de l'opérateur (création de comptes dans des volumes anormaux, partage de contenus à des fréquences anormales, utilisation de renseignements faux, volés ou trompeurs, etc.).

Tous les opérateurs informent leurs utilisateurs des pratiques de manipulation interdites sur le service ; en revanche, ils ne font pas état de la manière dont ils s'efforcent de rendre ces informations claires et aisément accessibles depuis le service. L'observation montre une accessibilité variable de ces informations : certains mentionnent ce type de pratiques dans les standards et règles de la communauté (ou conditions générales d'utilisation), d'autres dans des espaces dédiés à la lutte contre la manipulation de l'information (*TikTok*⁷⁰, *Dailymotion*⁷¹, *Facebook* et *Instagram*⁷², *YouTube*⁷³, *Twitter*⁷⁴, *Snap* uniquement en anglais⁷⁵).

En matière de transparence sur les pratiques identifiées sur le service, la majorité des opérateurs en font état dans leurs rapports de transparence respectifs.

Meta, *Twitter* et *Google* publient des rapports réguliers sur les opérations identifiées et les mesures prises en conséquence, et la *Fondation Wikimédia* y consacre des articles.

Cependant, aucun opérateur n'a indiqué si une information était délivrée aux utilisateurs ayant consommé des contenus à la suite d'une de ces pratiques de diffusion massive. Il serait intéressant et utiles de connaître les actions et projets des opérateurs pour informer ceux qui ont interagi avec des faux comptes (automatisés ou non) ou qui ont consommé voire partagé un *deepfake* malveillant.

- ***Focus : la transparence sur les opérations de propagation de fausses informations soutenues par des entités étatiques ou institutionnelles***

Dans son précédent bilan annuel, l'Arcom a préconisé aux opérateurs d'« *informer davantage les utilisateurs sur les pratiques coordonnées d'influence et les risques qui en découlent, tout particulièrement en période électorale.* »

Meta partage ses observations⁷⁶ sur les comportements inauthentiques coordonnés qui ont pour objectif de « *manipuler le débat public dans un but stratégique* ». L'entreprise évoque ainsi le démantèlement de plus de 150 réseaux d'opération d'influence dans plus de 50 pays.

La déclaration de *Google* fait état de l'ensemble des campagnes d'influence déjouées sur ses services au 4^e trimestre 2021. L'opérateur publie également chaque trimestre un

⁷⁰ <https://faq.dailymotion.com/hc/fr/articles/360020059680-R%C3%A8gles-relatives-%C3%A0-la-d%C3%A9sinformation->

⁷¹ <https://faq.dailymotion.com/hc/fr/articles/360020059680-R%C3%A8gles-relatives-%C3%A0-la-d%C3%A9sinformation->

⁷² <https://transparency.fb.com/fr-fr/policies/community-standards/misinformation>

⁷³ https://www.youtube.com/intl/ALL_fr/howyoutubeworks/our-commitments/fighting-misinformation/#:~:text=Comme%20indiqu%C3%A9%20dans%20le%20r%C3%A8glement,ensemble%20clair%20de%20faits%20v%C3%A9rifi%C3%A9s.

⁷⁴ <https://help.twitter.com/fr/rules-and-policies/platform-manipulation>

⁷⁵ <https://snap.com/en-US/safety-and-impact/post/our-approach-to-preventing-the-spread-of-false-information>

⁷⁶ <https://about.fb.com/news/tag/coordinated-inauthentic-behavior/>

rapport sur la lutte contre les campagnes d'influence coordonnées qui concerne *YouTube* et plusieurs parties de *Google Search* (tel que *Google News*)⁷⁷.

Twitter met à la disposition du public les archives des tweets et des médias qui résultent d'opérations de désinformation liées à un État⁷⁸.

Si l'Arcom salue ces initiatives, elle remarque néanmoins une granularité insuffisante des informations et l'absence de données chiffrées sur ces opérations en France.

Préconisations

Au vu de ces éléments d'analyse, l'Arcom formule les préconisations suivantes :

- ❖ établir une liste claire de l'ensemble des pratiques portant atteinte à l'intégrité de leur service afin de détailler plus clairement, avec des indicateurs chiffrés, les moyens et mesures mis en œuvre pour y répondre. Pour ce faire, les opérateurs sont encouragés à échanger entre eux sur les types de pratiques identifiées et à se nourrir de la typologie commune qui sera élaborée par les signataires du code européen de bonnes pratiques renforcé contre la désinformation ;
- ❖ créer un espace unique et facilement accessible sur le service listant les interdictions en matière de pratiques portant atteinte à l'intégrité des services, celles identifiées et les décisions prises en conséquence ;
- ❖ travailler à la mise en œuvre de fonctionnalités d'information des utilisateurs victimes de certaines pratiques de manipulation, de manière individualisée et contextuelle.

⁷⁷ <https://blog.google/threat-analysis-group/tag-bulletin-q4-2021/>

⁷⁸ <https://transparency.twitter.com/fr/reports/information-operations.html>

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

La publicité et les fonctionnalités de monétisation des contenus sur les plateformes sont sources de revenus pour des acteurs malveillants et offrent à ces contenus une exposition importante. C'est pourquoi elles sont utilisées pour participer à la manipulation de l'information. Il importe donc que les plateformes en ligne prennent des mesures pour contrer ces utilisations indésirables. Dans sa recommandation de 2019 et ses préconisations de 2020 et 2021, l'Arcom a formulé des orientations en matière d'information des utilisateurs sur l'accessibilité des politiques publicitaires, la gestion des préférences publicitaires et le financement de la manipulation de l'information par les communications commerciales.

À l'exception de la [Fondation Wikimédia](#), qui ne diffuse pas de communications commerciales, l'ensemble des opérateurs ont fourni à l'Arcom des éléments de réponse qualitatifs sur ces différents points. Toutefois, alors que [Google Search](#) et [YouTube](#) figurent parmi les trois plateformes les plus visitées en France sur l'année 2021⁷⁹, [Google](#) a choisi de ne pas répondre à de nombreuses questions, notamment celles concernant le fonctionnement des communications commerciales sur ses services ou sur la gestion des préférences publicitaires.

Les communications commerciales diffusées par les opérateurs s'inscrivent pour la plupart dans les définitions « d'annonces publicitaires », de « contenus sponsorisés » et de « contenus en partenariat » proposées par l'Arcom⁸⁰, bien que parfois désignées selon des termes propres à l'opérateur. Deux opérateurs disposent de supports particuliers : [LinkedIn](#) propose des « campagnes InMail sponsorisées » qui utilisent les messageries personnelles des utilisateurs lors de campagnes publicitaires et [Snap](#) diffuse des communications commerciales via les produits « Filtres » et « Lens », des dispositifs de réalité augmentée. En outre, trois opérateurs ([Webedia](#), [Twitter](#) et [Dailymotion](#)) indiquent que de la publicité peut être intégrée aux contenus diffusés en direct⁸¹. La vérification en temps réel de tels contenus est une gageure⁸². Des études complémentaires pourraient éclairer davantage ce phénomène.

Seul [Dailymotion](#) donne des **estimations chiffrées** du financement de la manipulation de l'information en France. Sur son service, près de 50% des contenus identifiés comme de la désinformation, soit sept contenus, ne sont pas monétisés ; ceux qui l'ont été ont eu 5407 impressions en moyenne (sans que l'opérateur ne précise si cette caractéristique en fait des contenus particulièrement viraux) et l'essentiel des revenus qu'ils ont générés est capté par l'opérateur (près de 90%).

Pour sa part, [Snap](#) ne précise pas si les 971 contenus d'utilisateurs considérés comme étant de fausses informations sur son service ont bénéficié ou non des revenus des communications commerciales adjacentes.

⁷⁹ Données Médiamétrie/NetRatings, Audience Internet Global.

⁸⁰ L'Arcom proposait les définitions suivantes :

- **Annonce publicitaire** : message publicitaire au sens strict, créé dans un format dédié à la publicité et publié via des outils spécifiques.
- **Contenu sponsorisé** : publication d'un utilisateur (quel qu'il soit : individu, institution, marque...) sur la plateforme dans un format « organique » dont la visibilité est augmentée contre rémunération versée à la plateforme.
- **Contenu d'utilisateur en partenariat** : contenu organique créé par l'utilisateur en partenariat avec un tiers (notamment une marque) contre rémunération ou avantage.

⁸¹ Au cours de son audition par le comité d'expert désinformation, [Meta](#) a indiqué que des publicités pouvaient être ajoutées sur les contenus ayant été diffusés en direct, après leur diffusion.

⁸² En 2019, l'attentat de Christchurch a été diffusé en direct pendant dix-sept minutes avant d'être interrompu.

Ces chiffres ne sont pas représentatifs de l'ensemble du secteur et doivent être maniés avec précaution. L'Arcom regrette vivement l'absence de données précises sur ces sujets de la part des principaux opérateurs. Lors des auditions menées en septembre 2022, les opérateurs concernés ont toutefois annoncé que des travaux internes étaient en cours afin de pouvoir fournir davantage de données chiffrées, de sorte à se mettre en conformité avec le code européen de bonnes pratiques renforcé sur la désinformation.

5.1 Des modèles de vente et d'enchère et de ciblage publicitaire peu transparents

- ***Des modalités de vente et d'enchères opaques malgré leur facilité d'accès***

Certains opérateurs ont développé des modules propriétaires de gestion de la publicité programmatique axée sur la demande, directement mis à disposition des utilisateurs. D'autres utilisent des services tiers. Aucun n'a mentionné le nombre de communications commerciales vendues en programmatique sur leur service sur le territoire français.

[Webedia](#) et [Dailymotion](#) ont fait part de ventes en gré à gré. Cette méthode de commercialisation implique la vérification manuelle de chaque annonceur et limite ainsi le financement de la manipulation de l'information par des moyens automatisés.

D'une manière générale, les modalités de vente, en particulier les méthodes d'enchères utilisées, ont été peu décrites, à deux exceptions. [Twitter](#) indique que ses enchères dépendent d'un score publicitaire comprenant résonance, pertinence et caractère récent. Sur [Microsoft Advertising](#), elles sont en fonction des montants, de la pertinence et de la performance. Les contenus trompeurs ayant un potentiel de viralité important qui pourrait être assimilé à de la « pertinence » ou de la « performance » par un algorithme. Il serait utile que ces opérateurs précisent si des mesures sont prises pour éviter un tel effet indésirable.

Sur les plateformes disposant d'un gestionnaire de publicité intégré au service, il semble relativement facile de s'inscrire sur les services, que l'utilisateur soit un particulier ou un professionnel, puisque le passage vers un compte « professionnel » se fait par la simple activation d'une option dans les paramètres généraux, et de dépenser pour la diffusion de publicités grâce aux gestionnaires de publicités ou « *Ads manager* ». Par ailleurs, quelques opérateurs indiquent avoir mis en place un suivi des flux monétaires accessible à leurs utilisateurs ([Meta](#), [Dailymotion](#)).

- ***Une gestion des préférences publicitaires hétérogène***

Les options de ciblage des opérateurs se déclinent en données individuelles (localisation, langue, appareil, âge, sexe) et catégories prédéfinies par leurs soins, qui sont d'ailleurs bien renseignées dans les déclarations. Les données de ciblage peuvent également être fournies par des tiers : exploitation de cookies tiers ([Dailymotion](#), [Webedia](#)), listes de mails fournies par les annonceurs ([Twitter](#)), partenariats avec des sociétés d'études de marché américaines ([Snap](#)), données de services tiers ([Twitter](#)). Si l'origine des annonceurs est parfois indiquée dans les dispositifs d'information à proximité des contenus ([Meta](#), [Pinterest](#)), l'identité des services tiers dont sont issues les données n'y figure pas. Les opérateurs font également part de restrictions totales ou partielles du ciblage pour certaines catégories d'annonces (paris sportifs, produits bancaires, etc.) ou concernant un public sensible. Ce faisant, ils rendent difficile le ciblage des personnes en situation à risque ou précaire, susceptibles d'être particulièrement fragiles face aux contenus proposés, à plus forte raison si ces contenus sont ambigus.

De plus, les opérateurs ont indiqué avoir développé des collaborations avec des partenaires tiers pour nourrir leurs propres ciblage mais également pour proposer des

communications commerciales en dehors de leurs services. Par exemple, [LinkedIn](#), [Twitter](#) et [Meta](#) disposent de « réseaux d'audience », des encarts publicitaires gérés par les opérateurs sur des sites éditeurs partenaires. Pour atteindre cette audience spécifique, [Twitter](#) utilise les données du service tiers [AppLovin](#). [Snap](#) utilise les données d'une solution de gestion d'accès multiplateforme nommée [Yext](#) pour alimenter son produit « Maps ».

Les opérateurs ont peu décrit l'accès des utilisateurs aux paramètres de préférences de ciblage. Les pratiques décrites sont hétérogènes : si la plupart des opérateurs disposent d'un menu dédié, [Dailymotion](#) et [Webedia](#) utilisent des modules d'information propres à la gestion des cookies accessibles via des encarts lors de l'accès à leurs sites. Deux dispositifs intéressants, en ce qu'ils facilitent l'utilisation de l'outil de gestion, sont à relever : [LinkedIn](#) rend disponibles ces paramètres de gestion des préférences depuis les contenus⁸³ et [Twitter](#) a établi des ponts entre ses paramètres de gestion et les paramètres de confidentialité des terminaux Apple et Android.

Généralement, les opérateurs distinguent les données fournies directement par l'utilisateur, les données déduites des contenus téléversés ou de l'activité sur le service et celles issues de prestataires extérieurs. En 2021, deux opérateurs ([TikTok](#), [Webedia](#)) offraient un choix binaire : l'utilisateur était contraint d'accepter ou de refuser toutes les publicités personnalisées. Toutefois, lors de son audition par le comité d'expert désinformation de l'Arcom, [TikTok](#) a annoncé avoir mis en place en 2022 des mesures de paramétrage supplémentaires, rapprochant son modèle des bonnes pratiques existantes. De plus, à l'exception de [Dailymotion](#), les services permettent aux utilisateurs de masquer des publicités, ce qui leur permet d'être actifs dans ce qu'ils ont envie ou non de voir, en particulier concernant des contenus qu'ils considéreraient comme ambigus sans qu'ils ne relèvent manifestement de la désinformation.

5.2 Une information sur le fonctionnement des politiques publicitaires à renforcer

- ***Une amélioration de l'accessibilité des politiques publicitaires***

Dans son précédent bilan, l'Arcom avait préconisé aux plateformes de « proposer des politiques publicitaires en français et comportant une partie relative à la manipulation de l'information ».

L'Arcom note que, globalement, les opérateurs ont mis en place de telles politiques et elle se réjouit que l'ensemble des opérateurs disposent de politiques publicitaires en français.⁸⁴ A cet égard, [Webedia](#) souhaiterait que l'Autorité produise un guide de bonnes conduites à destination du grand public qui pourrait être relayé au sein de sa politique publicitaire. Sur ce point, l'Arcom considère qu'il appartient aux opérateurs de construire leurs propres réflexions en s'appuyant sur la recherche et la société civile et que les politiques publicitaires sont destinées en priorité aux annonceurs.

Enfin, l'Arcom regrette que [Pinterest](#) et [TikTok](#) ne se soient pas prononcés sur les préconisations formulées par l'Arcom en matière de transparence des communications commerciales.

- ***De puissants dispositifs d'informations proches des contenus***

Le regard que porte un utilisateur sur un contenu publicitaire pensé pour l'inciter à agir (achat, dons, etc.) est nécessairement différent de celui qu'il porte sur un contenu organique (c'est-à-dire diffusé sans avoir recours à un format publicitaire). Aussi, il est important de lui apporter des clés pour comprendre s'il est confronté à un contenu

⁸³ Dans sa déclaration sur les mesures prises en 2020, [Meta](#) mentionnait un dispositif similaire.

⁸⁴ [Dailymotion](#) a confirmé par écrit que, contrairement à ce qui été indiqué dans sa contribution, sa politique publicitaire est bien disponible en français.

commercial ou un contenu organique, qui prennent parfois des formes similaires. C'est pourquoi l'Arcom avait également préconisé de mettre en place des outils d'information faciles d'accès.

Certains opérateurs proposent des pages d'information. [Dailymotion](#) indique qu'un travail de vulgarisation préalable de sa documentation lui est nécessaire. L'Arcom invite l'opérateur à produire de premiers éléments qui pourront être progressivement actualisés.

Certains proposent un dispositif d'information contextuel, à proximité de la communication commerciale. Les opérateurs identifient les communications commerciales autour de variations des termes « pub », « annonce » ou « sponsorisé » et donnent parfois des informations sur l'origine de la publicité ([Meta](#), [Pinterest](#)). Ces bonnes pratiques pourraient être approfondies, ces dispositifs offrant une porte d'entrée à proximité des contenus qui pourrait être largement enrichie. Par exemple, des opérateurs se servent de ces dispositifs pour diffuser des informations sur le ciblage ([Meta](#) et [Microsoft](#)), ou vers des pages d'information plus complètes ([LinkedIn](#)).

Si les partenariats rémunérés ne semblent pas soulever de problème majeur en matière de manipulation de l'information *stricto sensu*, les risques ne sont pas exclus, notamment en matière de santé publique. La loi française⁸⁵ pose le principe général d'une transparence de toute forme de publicité en ligne. Les conditions générales des plateformes prévoient que l'utilisateur mettant en ligne le contenu contenant un tel partenariat doit l'identifier ; des outils sont développés à cette fin (mention dans l'image ou dans la description, etc.), ainsi que des pages de sensibilisation dédiées et des guides de bonnes pratiques, les plateformes travaillant en partenariat avec l'Autorité de régulation professionnelle de la publicité (ARPP). En plus de ces bonnes pratiques, il serait intéressant que les opérateurs mentionnent explicitement cette infraction dans leurs conditions générales et précisent davantage, dans leur déclaration, comment l'absence d'identification de la part des utilisateurs est identifiée et sanctionnée en tant qu'infraction aux conditions générales du service⁸⁶.

En termes de transparence, les bases de données publicitaires sont des outils précieux. À cet égard, certains opérateurs proposent un registre des communications commerciales politiques. [Twitter](#) conserve les archives de celles diffusées sur son service lorsqu'elles étaient autorisées, entre juin 2018 et novembre 2019 ; son registre est intitulé « Centre de transparence ». [Snap](#) recense les publicités à caractère politique et militant ; le format brut « csv » de cette base présente l'avantage de faciliter les recherches, mais est moins accessible au grand public. De surcroît, cette base est difficilement accessible, à ce jour, depuis les pages d'informations de l'opérateur. [Google](#) dispose également d'une bibliothèque des communications commerciales politiques diffusées sur son service, même s'il n'en fait pas état dans sa déclaration. La base de [Meta](#) dispose d'un portail internet et d'un rapport synthétique à destination du grand public et d'une API, plus précise mais plus complexe à utiliser, à destination du public expert et de la recherche. De plus, l'opérateur a renforcé la transparence des publicités portant sur un enjeu social, électoral ou politique.

Concernant les autres contenus, seul [Meta](#) conserve les données des publicités actives. Pour continuer dans cette démarche de transparence, les données des publicités pourraient être conservées sur une période donnée après leur diffusion, puisque lorsque des publicités sont diffusées sur un temps très court, elles ne peuvent être identifiées a posteriori.

⁸⁵ Article 20 de la loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique : « Toute publicité, sous quelque forme que ce soit, accessible par un service de communication au public en ligne, doit pouvoir être clairement identifiée comme telle. Elle doit rendre clairement identifiable la personne physique ou morale pour le compte de laquelle elle est réalisée. »

⁸⁶ Seul [Snap](#) mentionne la difficulté d'identifier les utilisateurs ne se conformant pas à l'identification des partenariats commerciaux. Aussi, l'opérateur n'a pas mis de dispositif particulier et repose essentiellement sur les signalements des utilisateurs.

Ces dispositifs pourraient être des outils d'information très puissants, y compris pour les contenus ne relevant pas d'un débat d'intérêt général. Leur accès pourrait donc être davantage favorisé, en particulier de façon contextuelle, par un lien depuis les dispositifs d'information proche des contenus.

5.3 Des dispositifs de supervision des publicités encourageants

- ***Des bonnes pratiques de vérification de la conformité des communications commerciales qui pourraient être étendues***

Les opérateurs procèdent généralement à la vérification des communications commerciales en deux étapes : la vérification des annonceurs, puis celle des contenus. Ces vérifications ne sont pas systématiquement manuelles. Un contrôle supplémentaire peut intervenir une fois franchi un seuil élevé d'impression ou sur signalement d'un utilisateur. [TikTok](#) procède également à la vérification aléatoire d'échantillons de contenus pour s'assurer de leur conformité.

Comme indiqué plus haut, le passage vers un compte professionnel sur les plateformes disposant d'un gestionnaire de publicité intégré au service se fait très facilement, souvent via un bouton accessible dans les paramètres généraux du service. Cette facilité d'accès brouille les frontières entre annonceur professionnel et utilisateur final. Dans leurs déclarations, les opérateurs ont principalement décrit des dispositifs de sanctions des utilisateurs finaux, tout en précisant que des dispositifs différents existent en direction des annonceurs⁸⁷.

Quelques opérateurs ont fait part d'opérations de manipulation « *sophistiquées* » ([Microsoft](#)) ou « *coordonnées* » ([Google](#)). Par exemple, [Google](#) a attribué la création simultanée de milliers de comptes à une même opération. Ces phénomènes pourraient avoir touché d'autres opérateurs. Ainsi, dès lors qu'il existe ce type d'opérations effectuées par des moyens automatisés, il faut lire avec précaution l'indicateur du volume de comptes sanctionnés (ex. : suspension de plus de cinq millions de comptes d'annonceurs par [Google](#)) : il n'équivaut pas à un volume de personnes malveillantes.

Par ailleurs, un seul opérateur, [Webedia](#), indique vérifier les liens de redirection encapsulés dans les communications commerciales. Au vu des risques de renvoi depuis une communication commerciale en apparence anodine vers un site externe comportant des contenus de désinformation, l'Arcom invite les autres opérateurs à donner davantage d'informations sur leurs pratiques en la matière.

Il existe par ailleurs de nombreuses vérifications de l'environnement (contenus d'utilisateurs, sites d'éditeurs, etc.) dans lequel les communications commerciales sont diffusées, principalement effectuées par des entreprises spécialisées en sécurité des marques, américaines ou internationales.

- ***Les dispositifs de signalement des communications commerciales qui diffèrent du modèle général ne doivent pas être moins-disants***

L'article 11 de la loi du 22 décembre 2018 impose aux opérateurs de mettre en place « *un dispositif facilement accessible et visible permettant à leurs utilisateurs de signaler de [fausses] informations, notamment lorsque celles-ci sont issues de contenus promus pour le compte d'un tiers.* »

⁸⁷ Par exemple, [Microsoft](#) évoque des « *red flags* », [Google](#) supprime des comptes d'annonceurs, [Snap](#) indique prendre des mesures contre des entreprises.

Si l'ensemble des opérateurs permet le signalement des communications commerciales⁸⁸, parfois identiques aux outils de signalement généraux, certains ne respectent pas ou pas complètement cette disposition et ne proposent pas de motif « Fausses informations » ou « Trompeur ». C'est le cas de [YouTube](#) pour ses « annonces », de [Meta](#) pour un contenu sponsorisé sur [Instagram](#), ou de [Yahoo](#), qui estime ne pas être concerné car les contenus se rattachant à un débat d'intérêt général sont interdits sur son service. [Google Search](#), [Microsoft](#) et [Webedia](#) proposent chacun un signalement via un formulaire général qui ne comporte pas de motif « Fausses informations » mais permet à l'utilisateur de préciser l'objet de son signalement, y compris hors connexion. En outre, les formulaires ne se trouvent pas à proximité du contenu.

L'Arcom est très attentive à l'accessibilité de ces outils de signalement. Le dispositif spécifique aux contenus commerciaux sur [YouTube](#) apparaît moins accessible que le dispositif général : l'utilisateur doit cliquer sur une petite icône « i » en transparence, puis sur « Pourquoi cette annonce » pour signaler une publicité.

[TikTok](#) et [Twitter](#) ont fait part de plusieurs accès aux outils de signalement, à proximité des contenus ou par des voies complémentaires. L'Arcom encourage cette démarche en invitant [Twitter](#) à proposer cet outil en Français sur l'ensemble de son service, notamment l'espace *Moments*.

5.4 Des pratiques hétérogènes en matière de contenus promus se rattachant à un débat d'intérêt général

Les opérateurs ne disposent pas d'une définition commune relative aux publicités politiques et aux contenus promus se rattachant aux débats d'intérêt général. Les signataires du code européen de bonnes pratiques renforcé contre la désinformation se sont toutefois engagés à en adopter une. Actuellement, les définitions permettent une appréciation assez large, en particulier quand il est question d'intérêt général. Cas particulier, [Dailymotion](#) n'a pas fourni d'éléments sur son appréciation des publicités politiques ou se rattachant à un débat d'intérêt général. Par ailleurs, les politiques relatives aux publicités politiques ne sont pas toujours traduites entièrement en français ([Snap](#), [Yahoo](#)).

Quatre opérateurs autorisent la publicité politique, sauf en période électorale. Concernant les publicités se rattachant à un débat d'intérêt général, l'article L. 163-1 du code électoral impose aux opérateurs de fournir à l'utilisateur une information loyale, claire et transparente, agrégée au sein d'un registre mis à la disposition du public dans un format ouvert, sur les annonces diffusées pendant la période électorale et présentant un lien avec la campagne électorale (identité de l'annonceur, données personnelles utilisées, rémunérations versées). Si cinq opérateurs autorisent de tels contenus ([Meta](#), [Snap](#), [Google](#), [Yahoo](#), [TikTok](#)), deux seulement autorisent la diffusion des publicités se rattachant à un débat d'intérêt général en période électorale. Pour cela, [Meta](#) et [Snap](#) disposent tous deux d'une bibliothèque publicitaire (voir ci-dessus). De tels systèmes seront probablement adoptés par les signataires du code européen, qui prévoit l'utilisation de répertoires des publicités politiques et engagées.

L'identification des publicités politiques et se rattachant à un débat d'intérêt général est entièrement internalisée. La plupart des opérateurs ont mis en place des systèmes de vérification automatisés des contenus, de sorte à interdire ou à pré-labelliser les contenus, auxquels s'ajoutent parfois une vérification manuelle de la labellisation. Toutefois, l'absence de bibliothèque comprenant l'ensemble des publicités sur un temps long limite la possibilité, pour des tiers, de vérifier l'efficacité de cette labellisation.

⁸⁸ Le dispositif de signalement de [Dailymotion](#) ne précise pas si la vidéo signalée est bien la publicité ou le contenu de l'utilisateur.

Au vu de ces éléments d'analyse, l'Arcom formule les préconisations suivantes :

- ❖ l'ensemble des modèles d'enchères devraient faire l'objet de mesures de transparence renforcées, tel qu'un audit par des tiers indépendants, en particulier lorsqu'ils présentent des métriques propriétaires basées sur la performance qui pourraient être exploitées dans un but de manipulation de l'information ;
- ❖ dans la perspective d'une meilleure transparence des algorithmes de recommandation publicitaire, les dispositifs d'information contextuels devraient comprendre des informations sur le ciblage (de manière individualisée) et la gestion des paramètres et servir de lien vers d'autres outils d'information plus complets comme les bibliothèques publicitaires ;
- ❖ intensifier les efforts de vérification des liens de redirection encapsulés dans les communications commerciales, qui peuvent rediriger vers des sites tiers diffusant de la désinformation.
- ❖ L'Arcom rappelle en outre que la présence de désinformation dans les communications commerciales doit pouvoir être signalée par l'utilisateur via un dispositif facile d'accès et visible.

6. Éducation aux médias et à l'information et relations avec le monde de la recherche

6.1 Une hétérogénéité des actions en direction des utilisateurs face aux risques de manipulation de l'information

La notion d'éducation aux médias et à l'information (EMI) est ici appréhendée de manière large. Elle inclue autant les campagnes de sensibilisation qui permettent de développer des connaissances pour traiter des informations de manière critique que l'ensemble des ressources et fonctionnalités de « mise en capacité » des individus, leur permettant de mieux appréhender les informations qui leur sont proposées.

Les signataires du code européen de bonnes pratiques renforcé contre la désinformation se sont engagés, à cet égard, à mettre en œuvre des initiatives pour améliorer et renforcer les efforts en matière d'EMI et de développement de l'esprit critique, notamment en direction des publics les plus vulnérables.

Dans sa recommandation du 15 mai 2019, l'Arcom invite les opérateurs à sensibiliser les utilisateurs à l'influence de leurs propres contenus et à développer leur esprit critique. Pour ce faire, ils sont encouragés à développer des outils adaptés d'analyse de la fiabilité des sources d'information et à nouer des partenariats contribuant à l'EMI.

L'Arcom constate néanmoins que plusieurs opérateurs (*Dailymotion*, *Webedia*, *Pinterest*, *Yahoo*) n'ont pas fait état, dans leur déclaration, d'éléments relatifs à l'éducation aux médias et à l'information. Elle le déplore et appelle l'ensemble des opérateurs à s'engager avec détermination dans des actions d'EMI.

- ***Le lancement ou le soutien de plusieurs actions de sensibilisation en faveur de l'exercice de l'esprit critique face aux risques de manipulation de l'information***

Comme les années passées, seuls quelques opérateurs ont mis en place ou soutenu des campagnes de sensibilisation à l'attention des utilisateurs en lien avec la lutte contre la manipulation de l'information. *Dailymotion*, *Webedia*, *Pinterest*, *Yahoo*, *LinkedIn*, *Microsoft* et la *Fondation Wikimedia* n'ont pas mentionné de telles initiatives.

Si les formats et moyens déployés sont divers, la majorité des campagnes d'EMI déclarées visent à favoriser l'exercice, par les utilisateurs, de leur esprit critique pour permettre de mieux identifier les fausses informations. Il est intéressant de signaler qu'un certain nombre des campagnes déclarées sont construites de manière interactive en faisant participer les utilisateurs à l'aide de questionnaires ou en leur permettant de poser des questions en direct lors de la diffusion de contenus vidéo.

TikTok a par exemple créé un programme de sensibilisation⁸⁹ appelé « #EspritCriTik » sur les bons réflexes à adopter face à une information, en partenariat avec l'association Génération Numérique.

Twitter a publié un guide⁹⁰ à destination des utilisateurs pour mieux comprendre les fonctionnalités de son service.

Google, selon une démarche similaire, a publié un livret⁹¹, « *Le Vrai du Fake* », mettant en valeur des chaînes *YouTube* participant à « *aiguiser son esprit critique* ».

⁸⁹ <https://newsroom.tiktok.com/fr-fr/tiktok-lance-espritcritik-son-programme-de-sensibilisation-a-linformation>

⁹⁰ <https://about.twitter.com/content/dam/about-twitter/fr/tfg/download/twitter-digital-safety-playbook.pdf>

⁹¹ <https://kstatic.googleusercontent.com/files/acbcbf8cc76cb5b27d8f432ca72a6da04840def310cfd9c761e14afdbeb1918e1e6a8dfa8d1b3d5a00b802721d949c98d245d2609066e91191abab30d1fe143>

Par ailleurs, des partenariats ont été établis pour proposer des contenus pédagogiques aux utilisateurs, notamment dans le cadre de la crise sanitaire (*TikTok* avec Le Monde, *Google* avec Big Local News ou *Meta* avec l'OMS).

Certains opérateurs soutiennent des actions d'EMI initiées par des tiers. C'est le cas de *Meta*, via son Fonds pour le civisme en ligne, et de *Microsoft*, qui s'est associé à la société Newsguard pour organiser des événements en présence d'éducateurs, de bibliothécaires ou de parents ainsi que des ateliers en ligne accessibles à tous afin de promouvoir une « *meilleure compréhension des sources d'information fiables* ». Dans le même esprit, *Pinterest* a lancé un partenariat avec l'Association for Healthcare Social Media pour promouvoir leurs contenus destinés à sensibiliser les utilisateurs à l'importance de la vaccination. *Twitter* travaille avec l'UNESCO sur la liberté de la presse et *Google*, *Snap* et *Meta* ont soutenu plusieurs associations (notamment *Freeformers* pour *Meta*, Génération Numérique pour *Meta* et *Google* et e-enfance pour *Google* et *Snap*).

Enfin, l'Arcom relève une nouvelle fois la multiplication des initiatives à destination des journalistes, que cela soit *Meta* qui a financé un programme à destination des journalistes via son Fonds pour le Civisme en ligne, *Twitter* en offrant des crédits publicitaires à plusieurs associations qui œuvrent pour la liberté de la presse ou *Google* et sa *Google News Initiative* qui finance de nombreux médias à travers le monde.

- ***Le développement d'initiatives pour appeler l'utilisateur à la responsabilité et à une utilisation maîtrisée de la plateforme***

Dans sa recommandation du 15 mai 2019, l'Arcom encourage les opérateurs à sensibiliser les utilisateurs « *à l'utilisation maîtrisée des plateformes numériques et à la compréhension des enjeux, notamment démocratiques, qui s'y attachent* ». Si l'Autorité prend note des nombreuses actions en faveur du développement de l'esprit critique des utilisateurs, elle remarque que peu d'initiatives ont été lancées pour les accompagner dans une utilisation maîtrisée des plateformes, en particulier face aux risques spécifiques liés à la propagation de la désinformation.

Plusieurs opérateurs mettent à disposition des ressources générales dans des espaces dédiés pour améliorer les compétences et connaissances de leurs utilisateurs : *Snap* dans son centre de sécurité⁹², *Meta* dans une bibliothèque de ressources sur le numérique⁹³ et un portail pour les jeunes⁹⁴ ou encore *Google* (dans son livret et sur le site institutionnel de *YouTube*⁹⁵) et *Twitter* (dans son guide évoqué précédemment). Ces ressources ont l'avantage d'être pérennes même si elles sont encore assez peu accessibles ; les opérateurs ne font d'ailleurs pas état d'actions pour faire connaître ces dernières aux utilisateurs.

La *Fondation Wikimédia* déclare avoir initié le programme « *Wikipedia Education Program* » pour inciter les professeurs d'université à apprendre à maîtriser les outils de contribution sur son service.

L'Arcom relève avec intérêt l'existence de certaines fonctionnalités permettant d'inviter l'utilisateur à réfléchir au moment où il s'apprête à prendre connaissance ou à partager du contenu sur la plateforme.

Twitter envoie une notification aux utilisateurs sur le point de partager un contenu auquel est associé un lien qu'ils n'ont pas ouvert (l'opérateur précise que le nombre de personnes ouvrant des articles avant de les partager a augmenté de 33%).

⁹² <https://www.snap.com/fr-FR/safety/safety-center>

⁹³ <https://www.facebook.com/safety/educators>

⁹⁴ https://www.facebook.com/safety/youth?locale=fr_FR

⁹⁵ https://www.youtube.com/intl/ALL_fr/howyoutubeworks/our-commitments/fighting-misinformation/

Webedia rappelle, dans l'encart destiné à l'écriture des *posts*, que les messages doivent respecter sa charte d'utilisation de *Jeuxvideo.com*.

Dans le cadre de son partenariat avec NewsGuard, *Microsoft* propose une extension (uniquement disponible sur le navigateur *Microsoft Edge*) pour que les utilisateurs aient accès à une évaluation notée sur 100 de la fiabilité d'un site.

Afin de proposer des informations « fact-checkées » aux utilisateurs, *Google* rappelle dans sa déclaration avoir lancé un moteur de recherche appelé « Fact Check Explorer ».

Google est le seul opérateur à avoir déclaré une action de sensibilisation proposant des conseils face aux risques liés spécifiquement à *YouTube*, en partenariat avec Génération Numérique⁹⁶. Le *Digital Safety Book* de *Twitter* ou la page dédiée dans le centre de confidentialité de *Snap* s'en approchent mais mentionnent uniquement les solutions offertes aux utilisateurs.

- ***Toujours aussi peu de mesures et de transparence sur l'impact réel des actions d'EMI des plateformes***

Dans son dernier bilan annuel, l'Arcom a préconisé aux opérateurs d'évaluer l'impact des actions d'EMI qu'ils mènent sur le comportement de leurs utilisateurs et d'en rendre compte au public. L'Arcom remarque que la majorité des répondants n'ont pas fait état de telles évaluations.

Seuls *Twitter* (sur l'impact d'une de ses fonctionnalités), *Google* (sur le nombre de personnes qui ont suivi les formations financées) et *Meta* (sur les audiences de ses campagnes) ont fourni des données permettant de rendre compte de la portée de leurs initiatives en termes d'audience ou de compréhension de leurs conséquences.

Si l'Arcom estime qu'il est indispensable que les opérateurs fassent état de données chiffrées sur la réception et les impacts de leurs campagnes d'EMI, c'est parce qu'il est fondamental qu'ils puissent montrer que les moyens et mesures pris en la matière ont un impact réel pour les utilisateurs.

6.2 Une amélioration de la coopération avec le monde de la recherche mais un accès toujours très limité aux données

Dans sa recommandation de 2019, l'Arcom encourage les opérateurs à « *soutenir les initiatives indépendantes émanant de journalistes et de chercheurs destinées à mieux comprendre et mesurer le phénomène de la désinformation, notamment en leur donnant accès à leurs données dans le respect des règles en matière de protection des données personnelles, et dans le respect de l'impartialité de ces travaux* ». Dans la continuité de cette recommandation, l'Autorité a appelé les opérateurs, dans ses précédents bilans, à coopérer avec la société civile dans toutes ses composantes, aussi bien en soutenant ses initiatives qu'en facilitant l'accès aux données permettant d'analyser les phénomènes de manipulation de l'information.

L'Arcom constate que *Dailymotion*, *Yahoo* et *LinkedIn* n'ont pas déclaré d'éléments en la matière.

⁹⁶ https://www.youtube.com/intl/ALL_fr/howyoutubeworks/our-commitments/fighting-misinformation/

- **Quelques éléments déclarés sur les coopérations ou soutiens au monde de la recherche dans le domaine de la lutte contre la désinformation**

Certains opérateurs ont évoqué plusieurs initiatives en collaboration avec des chercheurs. *Pinterest* a déclaré coopérer avec le cabinet *Graphika* pour s'informer des tendances en matière de manipulation de l'information sur les autres plateformes en ligne afin d'anticiper les éventuels dangers sur son service.

La *Fondation Wikimedia* a lancé deux initiatives au niveau international : la *Wikimedia Research & Technology Fund*, un fonds pour subventionner des chercheurs, et le *Knowledge Integrity Program* pour mener des recherches sur les usages des contributeurs et développer des algorithmes pour appuyer le travail des modérateurs.

Microsoft indique, comme l'année passée, apporter un soutien financier à de nombreuses initiatives du monde de la recherche : le Trust Project de l'Université de Washington sur les phénomènes de désinformation, les travaux de Princeton sur les interférences étrangères, l'Institut Internet d'Oxford pour déployer une équipe spécialisée sur l'étude des phénomènes en période électorale ou encore des projets pour développer des outils d'identification des *deepfakes* ou des *data voids*.

Google annonce s'être lancé dans une collaboration de plusieurs années avec l'Institut universitaire européen, l'Observatoire européen des médias numériques et la Fondation Calouste Gulbenkian pour développer des projets de lutte contre la manipulation de l'information.

Les opérateurs ont fait état de nombreuses coopérations ; néanmoins, la majorité concernent des partenariats à l'étranger, et particulièrement aux Etats-Unis. Le développement de partenariats ou collaborations aux niveaux français et européen serait précieux.

- **Une accessibilité des données qui demeure très limitée**

Dans son précédent bilan annuel, l'Arcom appelle les opérateurs à intensifier « les collaborations avec le monde de la recherche, notamment en mettant en œuvre les conditions permettant la mise à disposition et l'exploitation des données à grande échelle ». Il apparaît en effet crucial de donner à la communauté académique les moyens d'étudier les dynamiques à l'œuvre sur les plateformes en ligne, notamment en matière de désinformation, et de développer des outils et approches indépendants afin d'éclairer ces phénomènes.

Pourtant, seuls quatre opérateurs (*Dailymotion*, *Meta*, *Twitter*, *Google*, *TikTok*) ont mentionné s'associer avec des chercheurs pour étudier les phénomènes à l'œuvre sur leur service ou avoir mis en place des initiatives d'ouverture de leurs données.

Dailymotion collabore ponctuellement avec ce type d'acteurs notamment sur le profilage algorithmique.

Meta indique travailler avec l'*European Digital Media Observatory (EDMO)* sur un cadre évolutif du partage des données et évoque le lancement, début 2022, de la « Researcher Platform »⁹⁷ afin de favoriser la recherche indépendante reposant sur l'analyse des données à grande échelle de ses services. Il indique également avoir lancé *Meta Data for Good* pour s'associer avec la *Social Progress Imperative* dans le but de développer un outil de cartographie des risques liés à la crise sanitaire.

⁹⁷ <https://research.facebook.com/data/>

Google et TikTok ont annoncé durant l'été 2022⁹⁸ le lancement d'initiatives pour une meilleure accessibilité de leurs données : un site dédié permettant aux chercheurs de demander un accès spécifique à l'API YouTube⁹⁹ et une API pour la recherche sur le système de modération de TikTok.¹⁰⁰

Twitter se distingue particulièrement dans ce domaine. Il offre aux chercheurs et développeurs la possibilité d'exploiter une partie de ses données grâce à une API ouverte depuis 2006. En 2021, une mise à jour a permis d'offrir aux chercheurs un accès gratuit à l'historique des conversations publiques, avec des niveaux d'accès plus élevés et des fonctionnalités de filtrage plus précises. Actuellement, il travaille en collaboration avec le Stanford Internet Observatory et l'Australia Strategic Policy Institute afin de rendre ses données brutes plus facilement compréhensibles. Il est toutefois important de signaler que Twitter a déclaré avoir lancé, début 2022, le Twitter Moderation Research Consortium pour étudier les questions de gouvernance des plateformes et auquel pourront adhérer des groupes ou des personnes « ayant fait leurs preuves en matière de recherche sur la modération de contenu et les questions d'intégrité. »

Certains opérateurs comme Meta font mention des difficultés auxquelles se heurte la mise en place de ce type de pratiques : normes d'éligibilité des chercheurs, normes techniques d'anonymisation des données, garanties techniques quant à l'accès aux données, garanties organisationnelles, modalités de recherche, etc. Néanmoins et malgré ces obstacles, comme l'a rappelé l'Arcom lors du lancement de la consultation publique sur l'accès aux données des plateformes pour la recherche¹⁰¹, il est essentiel que soit élaboré un cadre permettant l'étude des comportements en ligne et leurs effets. Ces initiatives permettent notamment de contribuer à la protection et au renforcement de l'indépendance, de l'autonomie et de la capacité d'analyse propres à la recherche.

À cet égard, le DSA qui entrera en application dans quelques mois, contribuera à changer significativement la donne. Elle établit en effet un cadre permettant un meilleur accès des chercheurs aux données clés des très grandes plateformes en ligne afin que le monde académique contribue à une meilleure compréhension de l'évolution des risques sur ces dernières.

Préconisations

Au vu de ces éléments d'analyse, l'Arcom formule les recommandations suivantes :

- ❖ mettre en place des actions d'éducation aux médias et à l'information sur les risques en matière de manipulation de l'information liés à l'utilisation de leur service et proposer des fonctionnalités permettant aux utilisateurs de prendre en compte ces risques lors de cette utilisation ;
- ❖ développer davantage la mesure de l'impact réel des actions d'EMI des opérateurs sur les utilisateurs et en rendre compte ;
- ❖ coopérer davantage avec la société civile à des initiatives communes permettant une meilleure compréhension des phénomènes en ligne, et plus particulièrement avec les organismes de recherche sur une meilleure accessibilité des données.

⁹⁸ Source : auditions de Google et TikTok, septembre 2022.

⁹⁹ Source : <https://research.youtube/>

¹⁰⁰ Source : <https://newsroom.tiktok.com/fr-fr/renforcer-notre-engagement-en-matiere-de-transparence>

¹⁰¹ <https://www.arcom.fr/consultations-publiques/consultation-publique-sur-lacces-aux-donnees-des-plateformes-en-ligne-pour-la-recherche>

7. Mesures complémentaires prises face à des situations spécifiques

7.1 L'impact de la crise de la COVID-19 sur les phénomènes de manipulation de l'information et les réponses apportées par les plateformes

Le prolongement de la crise sanitaire tout au long de l'année 2021 et la généralisation de l'accès aux vaccins se sont accompagnés de la persistance de risques majeurs concernant la manipulation et la diffusion de fausses informations.

Les opérateurs de plateforme ont poursuivi, mis à jour, voire amplifié les actions permettant de répondre à ces risques. Si certaines de ces mesures sont déployées depuis 2020, les éléments inédits recueillis cette année témoignent de la réactivité et de l'esprit de coopération dont ont fait preuve les opérateurs pour s'adapter au contexte né du prolongement de la crise.

Bien que certains acteurs aient indiqué ne pas avoir constaté de profonds changements dans leur modération (*Fondation Wikimédia*, *Dailymotion*, *Webedia*), la crise de la COVID-19 a eu un réel impact pour les plateformes (notamment les réseaux sociaux et les moteurs de recherche) au vu des nombreuses mesures mises en place et de la multiplication des collaborations avec les acteurs institutionnels, les professionnels de santé et le monde de la recherche académique.

• **Présentation des mesures prises par les opérateurs**

Mesures d'information	générales	Création et/ou mise à jour d'un site, d'une page, et/ou d'un encart dédié : <i>Fondation Wikimédia</i> , <i>Google</i> , <i>YouTube</i> , <i>LinkedIn</i> , <i>Meta</i> , <i>Microsoft</i> , <i>Pinterest</i> , <i>TikTok</i> , <i>Twitter</i> , <i>Yahoo</i> Publication de rapports périodiques de transparence : <i>Microsoft</i>
Traitement des signalements		Recours accru à l'apprentissage automatique et à l'automatisation pour la modération : <i>Twitter</i>
Politique de modération des contenus		Adaptation continue de la politique de modération : <i>YouTube</i> , <i>Meta</i> , <i>TikTok</i> , <i>Twitter</i> Clarifications supplémentaires aux modérateurs : <i>TikTok</i> , <i>Webedia</i> Formation intensive des membres des équipes « Confidentialité et Sécurité », reposant notamment sur des entités tierces indépendantes : <i>Snap</i>
Promotion des contenus issus de sources faisant autorité		Mise en avant des contenus <i>fact-checkés</i> : <i>Google</i> , <i>Meta</i> Mise en avant d'informations provenant d'organismes publics (gouvernement, autorités de santé, etc.) : <i>Google</i> , <i>Meta</i> , <i>Microsoft</i> , <i>Pinterest</i> , <i>TikTok</i> , <i>Twitter</i> , <i>Snap</i> , <i>LinkedIn</i> Lancement d'un fonds destiné à soutenir des initiatives en matière de vérification des informations liées aux vaccins contre la COVID-19 : <i>Google</i>
Collaboration avec des experts		Actions en partenariat avec des experts en matière de santé : <i>TikTok</i> Soutien aux bases de données collaboratives : <i>Google</i>

Lutte contre la diffusion massive de fausses informations	Réduction de la visibilité et/ou de la diffusion des contenus allant à l'encontre de la politique de modération en la matière : Meta , Twitter
Contenus publicitaires en lien avec la COVID-19	Restriction des publicités contenant une référence implicite ou explicite : Twitter Suspension d'annonceurs menant des activités frauduleuses : Google Blocage des annonces reposant sur une liste de mots-clés : Microsoft Mise à jour des règles publicitaires : Meta
Education aux médias (EMI)	Lancement de campagnes d'éducation aux médias : Meta , Microsoft , TikTok
Collaboration avec le monde de la recherche	Soutien (notamment financier) de travaux de recherche : Google , Meta , TikTok

- ***Une information actualisée et transparente des utilisateurs au cœur des dispositifs mis en place par les opérateurs***

L'Arcom souligne l'engagement renforcé des opérateurs dans la sensibilisation des utilisateurs aux risques entourant la manipulation de l'information dans le cadre de la crise sanitaire.

La très grande majorité des déclarants indique avoir œuvré en faveur de la promotion d'une information vérifiée dans le cadre de la lutte contre l'épidémie. À l'exception de [Dailymotion](#), tous indiquent ainsi disposer d'un encart, d'une page, voire d'un site dédié aux informations provenant de sources faisant autorité sur le sujet de la crise sanitaire. [Webedia](#) envisage de mettre en place un bandeau de redirection vers des sites « *contenant une information fiable* » (ex. : institutionnels) lorsque cela sera jugé nécessaire par ses modérateurs.

À l'instar de [TikTok](#) et [Pinterest](#) dès l'année dernière, certaines plateformes ([Google](#), [YouTube](#), [Bing](#), [Yahoo](#), [Snapchat](#)) ont déployé des fonctionnalités permettant de mettre en avant les contenus émanant de sources officielles ou d'étendre la portée des communications officielles.

[Microsoft](#) publie, au niveau européen, un rapport bimensuel sur la désinformation liée à la pandémie de la COVID-19 observée sur son service.

Les opérateurs ont mis en place de nouvelles initiatives de sensibilisation des utilisateurs aux dangers liés à la crise. Ainsi, en 2021, [Google](#) a lancé l'initiative « Restez informés » sur [Google](#) et [YouTube](#), [Meta](#) a mené une campagne européenne « Unis contre les fausses informations sur le COVID-19 », [Microsoft](#) a co-sponsorisé la campagne « VaxFacts » lancée par [NewsGuard](#) et [TikTok](#) a organisé des *lives* avec des experts en matière de santé. En parallèle, [Google](#), à travers son programme *Google News Initiative* notamment, s'est montré particulièrement actif dans la promotion d'une information d'autorité en 2021 ; il a formé 33 000 journalistes européens à la vérification numérique et au journalisme de données depuis le début de la pandémie, via sa formation *News Lab*.

S'agissant des collaborations avec le monde de la recherche en 2021, [Meta](#) a annoncé l'alliance de *Meta Data for Good* avec *Social Progress Imperative* pour développer un outil de cartographie des « *risques auxquels différentes populations sont confrontées au niveau sous-national* » face à la COVID-19 en Allemagne et en France. Cet outil combine des données issues du Sondage sur les tendances et l'impact de la COVID-19 du programme *Data for Good* et des indicateurs socio-économiques et sanitaires supplémentaires provenant d'autres sources. [Google](#) a ouvert un fonds mondial doté de trois millions de

dollars visant à financer des initiatives en matière de lutte contre la manipulation de l'information relative au vaccin contre la COVID-19 et a soutenu la création, par les plateformes *Big Local News* et *Pitch Interactive* de l'université de Stanford, d'un outil de cartographie des cas de COVID-19 à l'échelle mondiale. *TikTok* s'est associée à l'*Institut COVID-19 Ad Memoriam* pour créer une plateforme numérique nommée « *histoires de crise* » afin de rassembler des contenus liés à la pandémie de la COVID-19 et d'effectuer des recherches et des analyses sur la pandémie et ses conséquences¹⁰².

- ***L'approfondissement du travail de collaboration avec les acteurs centraux de la crise***

Afin de contrer les fausses informations circulant sur leur service, les opérateurs ont poursuivi, voire développé leur collaboration avec les acteurs institutionnels gouvernementaux (Service d'information du Gouvernement, ministère de la Santé), des organisations internationales (UNICEF et UNESCO notamment) et des autorités de santé (Santé publique France, l'OMS).

Plusieurs opérateurs (*Google* et *Twitter* notamment), dont les règles d'utilisation restreignaient considérablement voire interdisaient la promotion de contenus relatifs à la COVID-19, ont admis une exception lorsque cette promotion est réalisée par des organismes publics (ex. : entités gouvernementales, OMS, etc.).

- ***La recherche persistante d'un équilibre entre modération et liberté d'expression en temps de crise***

Le prolongement de la crise sanitaire n'aura pas été sans conséquence sur le nombre de fausses informations en circulation sur les plateformes. Pour y faire face, un certain nombre d'opérateurs relève la nécessité d'un travail accru de modération sur leur service (*Twitter*, *Google*, *Pinterest* et *Snap* notamment). Les déclarations montrent que la question du délicat arbitrage entre modération et liberté d'expression en temps de crise, déjà exposée dans le précédent bilan de l'Arcom, reste particulièrement prégnante.

À titre d'exemple, sur *Twitter* en 2021, 353 018 comptes ont été mis en cause, 3 455 ont été suspendus et 63 786 contenus uniques en lien avec la COVID-19 ont été supprimés.

Twitter a expliqué que la persistance de la crise avait eu des conséquences directes sur les équipes de modération de son service et a indiqué avoir effectué une hiérarchisation des signalements selon leur motif lors de leur traitement. Néanmoins, en raison de cette approche, l'opérateur indique avoir « *constaté des ralentissements et des allongements de délais de traitement dans d'autres domaines* ».

Globalement, les autres opérateurs expliquent que les mesures développées en 2020 lors de la survenue de la crise sanitaire leur ont permis de disposer d'outils et de dispositifs adaptés en matière de modération des contenus liés à la COVID-19 en 2021.

Outre la détection proactive, la suppression des contenus problématiques et la suspension des comptes qui en sont à l'origine, certaines mesures visent à réduire la visibilité ou la diffusion de contenus problématiques. *Twitter* affirme réduire la visibilité des tweets ou des comptes lorsqu'il est certain qu'ils violent sa politique en matière de fausses informations sur la COVID-19. *Meta* prend également des mesures de réduction de la diffusion de ces contenus, dès que ses systèmes de détection les identifient comme tels.

Twitter procède à la restriction des contenus publicitaires comportant une référence implicite ou explicite à la COVID-19 et *Microsoft* bloque les annonces publicitaires en lien avec la COVID-19, qu'il détecte en s'appuyant sur une liste de mots-clés. *Google* intervient également à l'encontre des annonceurs s'adonnant à une activité frauduleuse dans le cadre

¹⁰² Voir également parties II.3 et II.6 du présent bilan.

de la crise sanitaire : il a ainsi bloqué plus de 106 millions de publicités liées à la COVID-19 depuis le début de la pandémie et le nombre global de suspensions de comptes d'annonceurs a triplé entre 2020 et 2021.

Snap indique avoir été confronté à une problématique nouvelle liée à la crise sanitaire de la COVID-19 en France : la lutte contre la fraude aux « pass » sanitaires sur son service.

7.2 Les phénomènes de manipulation de l'information pendant la période électorale de 2022 et les réponses apportées par les plateformes

À l'approche de la période électorale française de 2022 (élections présidentielle et législatives), la coopération de l'Arcom avec les pouvoirs publics – Commission nationale de contrôle de la campagne électorale en vue de l'élection présidentielle (CNCCEP) et Viginum – et avec les opérateurs de plateforme s'est renforcée. Un mode opératoire inédit a été adopté avec l'instauration d'un dialogue nourri avec les opérateurs et la création d'un espace d'information¹⁰³ sur le site de l'Arcom pour faciliter l'accès du public aux mesures générales et spécifiques prises par les plateformes durant la période électorale.

Une synthèse des nombreuses initiatives des opérateurs ainsi que des questions soulevées est présentée ci-dessous. Le lecteur est invité à se référer au rapport de l'Arcom sur les campagnes électorales de 2022¹⁰⁴, publié le 2 novembre 2022, pour une analyse détaillée de ces éléments.

- **Les initiatives de transparence des plateformes**
 - **Relatives aux mesures générales prises par les opérateurs pour lutter contre la manipulation de l'information**

Certains opérateurs ont mis à disposition du public un d'espace d'information dédié à la période électorale, notamment aux mesures prises pour lutter contre la manipulation de l'information (*TikTok*, *Webmedia*, *Meta*, *Twitter*). Il s'agit d'une bonne pratique que l'Arcom avait préconisée en janvier, estimant important qu'ils communiquent de manière proactive, claire et accessible sur les mesures particulières prises dans le contexte électoral et fassent œuvre de transparence sur les raisons pouvant expliquer le retrait d'un contenu (ce qui n'a pas toujours été le cas, à l'exception de *Google* et *Wikimédia France*).

- **Relatives aux rémunérations en contrepartie de la promotion de contenus d'information se rattachant à un débat d'intérêt général**

Si *Google*, *Microsoft*, *TikTok*, *Meta* et *Snap* autorisent en toute période la diffusion de contenus promus se rattachant à un débat d'intérêt général, seuls les deux derniers la permettent en période électorale.

À cet égard, la loi impose aux opérateurs autorisant la diffusion de tels contenus en période électorale une obligation de transparence¹⁰⁵. *Meta* a mis à disposition les données relatives aux publicités de nature sociale, électorale ou politique autorisées sur *Facebook* et

¹⁰³ <https://www.arcom.fr/lutte-contre-la-manipulation-de-linformation-en-periode-electorale-les-mesures-prises-par-les-plateformes-en-ligne>

¹⁰⁴ <https://www.arcom.fr/nos-ressources/etudes-et-donnees/mediatheque/rapport-sur-les-campagnes-electorales-2022-election-la-presidence-de-la-republique-et-elections-legislatives>

¹⁰⁵ En application de l'article L. 163-1 du code électoral, les opérateurs concernés doivent fournir à l'utilisateur une information loyale, claire et transparente, agrégée au sein d'un registre mis à la disposition du public dans un format ouvert, sur les contenus promus se rattachant à un débat d'intérêt général diffusés pendant les trois mois précédant le premier jour du mois d'élections générales et jusqu'à la date du scrutin, et présentant un lien avec la campagne électorale (Conseil constitutionnel, décision n° 2018-773 DC du 20 décembre 2018, point 8). Ces informations concernent l'identité de l'annonceur, les données personnelles utilisées, les rémunérations versées lorsque le montant est supérieur à un seuil déterminé).

[Instagram](#) en période électorale. [Snap](#) propose une bibliothèque dédiée aux « publicités politiques ».

- ***Les initiatives à destination des équipes de campagne ou partis politiques***

Deux types de mesures ont été prises en la matière.

[Wikimédia France](#), [Google](#) et [Meta](#) ont publié des ressources destinées aux partis politiques et équipes politiques.

Les opérateurs ont par ailleurs désigné des points de contact, lors de l'élection présidentielle, pour les équipes de campagne. L'Arcom note que seuls [Meta](#) et [Google](#) ont proposé une adresse générique publique pour les deux périodes électorales.

- ***Les mesures d'éducation aux médias et à l'information et d'incitation au vote***

Les opérateurs ont multiplié les initiatives visant à informer leurs utilisateurs sur les scrutins et à les inciter à la participation au vote, telles que :

- 1) la création de sites, espaces ou fonctionnalités dédiés en partenariat avec le Service d'information du Gouvernement (SIG) ou avec des entreprises et agences de presse ([Google](#), [Meta](#), [Twitter](#), [Webedia](#), [TikTok](#), [Snap](#), [LinkedIn](#) notamment) ;
- 2) la valorisation des candidats et de leurs comptes officiels et la mise en avant des débats télévisés ([Google](#), [Snap](#)) ;
- 3) l'envoi de notifications pour inciter au vote et la mise en avant des résultats ou des contenus relatifs à la soirée électorale, le jour des différents scrutins ([Snap](#), [Meta](#), [Google](#), [Microsoft](#), [Dailymotion](#)).

Plusieurs opérateurs ont mené des campagnes d'EMI sur les risques de manipulation de l'information en période électorale ([Google](#) et [Meta](#) en partenariat avec l'AFP, [Snap](#), [TikTok](#), [Webedia](#), [LinkedIn](#)).

- ***Des campagnes qui n'ont pas rencontré de difficultés majeures mais qui ont soulevé des interrogations nouvelles***

L'ensemble des opérateurs ont relevé que la période de la campagne électorale avait été plus calme sur leur service en comparaison de précédents scrutins, en France ou à l'étranger. Ce constat, partagé par l'Arcom, peut s'expliquer par plusieurs facteurs (prolongement de la crise sanitaire, invasion de l'Ukraine par la Russie, absence de grand débat, robustesse des modalités de vote et écarts dans les résultats des premier et second tours de l'élection présidentielle notamment).

Des pratiques, phénomènes ou tendances problématiques ont été identifiés lors des deux campagnes mais sans conséquences avérées. Il convient par exemple de souligner des pratiques de détournement des services, la diffusion de narratifs relatifs à la délégitimation du scrutin ou encore l'émergence de problématiques en matière de respect de la propriété intellectuelle en période électorale.

Enfin, de nouvelles interrogations ont émergé quant à l'application des règles électorales sur les plateformes en ligne, s'agissant notamment des comptes utilisés par les candidats à l'élection présidentielle et du contrôle du respect des articles L. 49 et L. 52-2 du code

électoral relatifs, respectivement, à la période de réserve durant le week-end de scrutin¹⁰⁶ et à la diffusion de résultats¹⁰⁷.

¹⁰⁶ L'article L. 49 du code électoral interdit de diffuser des messages ayant un caractère de propagande électorale à partir de la veille du scrutin à zéro heure (soit le vendredi à minuit) par tout moyen de communication au public par voie électronique.

¹⁰⁷ L'article L. 52-2 du code électoral prévoit qu'aucun résultat, partiel ou définitif, ne peut être communiqué au public par quelque moyen que ce soit avant la fermeture des derniers bureaux de vote, soit avant 20 heures en métropole. Le même article prévoit, par dérogation, que lorsque la République forme une circonscription électorale unique, comme pour l'élection présidentielle par exemple, aucun résultat partiel ou définitif de l'élection ne peut être communiqué au public par quelque moyen que ce soit avant la fermeture du dernier bureau de vote sur le territoire métropolitain.

Synthèse des préconisations de l'Arcom

Sur la transparence en général

- ❖ **Préconisation n° 1** : faire preuve de plus de transparence vis-à-vis du public en fournissant davantage d'indicateurs chiffrés aussi bien sur les contenus, pratiques et phénomènes de manipulation de l'information identifiés que sur les moyens et mesures mises en œuvre pour lutter contre ceux-ci.
- ❖ **Préconisation n° 2** : mettre en place davantage d'initiatives pour améliorer la transparence vis-à-vis de la société civile et de la recherche, notamment en matière d'accessibilité des données, afin de permettre une évaluation indépendante et plurielle des risques et des mesures déployées pour les contrer.
- ❖ **Préconisation n° 3** : communiquer au régulateur davantage d'informations, de manière confidentielle si nécessaire, pour permettre une meilleure compréhension des moyens et mesures mis en œuvre pour lutter contre la manipulation de l'information.

Sur le dispositif de signalement

- ❖ **Préconisation n° 4** : améliorer la visibilité et la facilité d'utilisation du dispositif de signalement en le plaçant à proximité immédiate du contenu (notamment pour les moteurs de recherche [Google Search](#) et [Bing](#)), et ce, sur toutes les versions du service (particulièrement pour les versions mobiles de [TikTok](#) et [YouTube](#)).
- ❖ **Préconisation n° 5** : proposer, au sein des dispositifs de signalement, des motifs ou sous-motifs relatifs à la manipulation de l'information (ou notion équivalente) clairs et suffisamment précis (notamment pour [Twitter](#), [Jeuxvideo.com](#), [LinkedIn](#) et [YouTube](#)). Ces motifs doivent également couvrir la manipulation de l'information susceptible de troubler l'ordre public ou de nuire au bon déroulement des processus électoraux.
- ❖ **Préconisation n° 6** : mieux informer les auteurs de signalement et utilisateurs ayant publié un contenu signalé de l'avancée des procédures de traitement des signalements en cours et leur en communiquer l'issue dans un délai raisonnable et de façon motivée.
- ❖ **Préconisation n° 7** : permettre aux auteurs de signalement et aux utilisateurs affectés par une décision de modération de former un recours et leur fournir une information claire sur leurs possibilités de recours au moment de la notification de la décision prise.

Sur la transparence des algorithmes

- ❖ **Préconisation n° 8** : améliorer la transparence des processus et moyens dédiés à l'évaluation des outils automatiques en publiant régulièrement, dans un espace dédié, des éléments sur les modifications qui ont été réalisées à l'aide d'analyses internes ou externes.
- ❖ **Préconisation n° 9** : communiquer aux utilisateurs les critères utilisés pour la recommandation de contenus, et ce, de manière personnalisée et contextuelle à l'aide d'outils et fonctionnalités accessibles directement sur le service.

- ❖ **Préconisation n° 10** : instaurer une transparence accrue des politiques en matière de réduction de la visibilité des contenus.
- ❖ **Préconisation n° 11** : faire preuve d'une transparence plus importante sur le fonctionnement des outils de détection de contenus, comptes et comportements utilisés à des fins de modération vis-à-vis du public et communiquer au régulateur davantage d'éléments permettant d'en évaluer la pertinence.

Sur la promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle

- ❖ **Préconisation n° 12** : mettre en place des dispositifs spécifiques d'identification des médias qui ne bénéficient pas des garanties d'indépendance éditoriale nécessaires vis-à-vis de l'État qui les contrôle et les appliquer de façon homogène, transparente et selon des critères objectifs.
- ❖ **Préconisation n° 13** : mettre en place des dispositifs permettant d'ajouter des « frictions » dans la diffusion de contenus potentiellement problématiques.
- ❖ **Préconisation n° 14** : développer le recours aux dispositifs (i) de *fact-checking* et (ii) de mise en avant des contenus en résultant, y compris hors période électorale.

Sur la lutte contre les comptes propageant massivement de fausses informations

- ❖ **Préconisation n° 15** : établir une liste claire de l'ensemble des pratiques portant atteinte à l'intégrité de leur service afin de détailler plus clairement, avec des indicateurs chiffrés, les moyens et mesures mis en œuvre pour y répondre. Pour ce faire, les opérateurs sont encouragés à échanger entre eux sur les types de pratiques identifiées et à se nourrir de la typologie commune qui sera élaborée par les signataires du code européen de bonnes pratiques renforcé contre la désinformation.
- ❖ **Préconisation n° 16** : créer un espace unique et facilement accessible sur le service listant les interdictions en matière de pratiques portant atteinte à l'intégrité des services, celles identifiées et les décisions prises en conséquence
- ❖ **Préconisation n° 17** : travailler à la mise en œuvre de fonctionnalités d'information des utilisateurs victimes de certaines pratiques de manipulation, de manière individualisée et contextuelle.

Sur les mesures de lutte contre les fausses informations en matière de communications commerciales et de promotion de contenus d'information se rattachant à un débat d'intérêt général

- ❖ **Préconisation n° 18** : l'ensemble des modèles d'enchères devraient faire l'objet de mesures de transparence renforcées, tel qu'un audit par des tiers indépendants, en particulier lorsqu'ils présentent des métriques propriétaires basées sur la performance qui pourraient être exploitées dans un but de manipulation de l'information.
- ❖ **Préconisation n° 19** : dans la perspective d'une meilleure transparence des algorithmes de recommandation publicitaire, les dispositifs d'information contextuels devraient comprendre des informations sur le ciblage (de manière

individualisée) et la gestion des paramètres et servir de lien vers d'autres outils d'information plus complets comme les bibliothèques publicitaires.

- ❖ **Préconisation n° 20 :** intensifier les efforts de vérification des liens de redirection encapsulés dans les communications commerciales, qui peuvent rediriger vers des sites tiers diffusant de la désinformation.
- ❖ **Préconisation n° 21 :** l'Arcom rappelle en outre que la présence de désinformation dans les communications commerciales doit pouvoir être signalée par l'utilisateur via un dispositif facile d'accès et visible.

Sur l'éducation aux médias et à l'information ainsi que sur les relations avec le monde de la recherche

- ❖ **Préconisation n° 22 :** mettre en place des actions d'éducation aux médias et à l'information sur les risques en matière de manipulation de l'information liés à l'utilisation de leur service et proposer des fonctionnalités permettant aux utilisateurs de prendre en compte ces risques lors de cette utilisation.
- ❖ **Préconisation n° 23 :** développer davantage la mesure de l'impact réel des actions d'EMI des opérateurs sur les utilisateurs et en rendre compte.
- ❖ **Préconisation n° 24 :** coopérer davantage avec la société civile à des initiatives communes permettant une meilleure compréhension des phénomènes en ligne, et plus particulièrement avec les organismes de recherche sur une meilleure accessibilité des données.

Annexe – Points-clés par plateformes

Bing	71
Dailymotion	73
Facebook et Instagram	75
Google Search et YouTube	77
Jeuxvideo.com	79
LinkedIn	81
Pinterest	83
Snapchat	85
TikTok	87
Twitter	89
Wikipédia	91
Yahoo Media	93

Bing

Opéré par [Microsoft](#), Bing est un moteur de recherche permettant à ses utilisateurs de trouver des ressources en ligne à l'aide de systèmes de classement basés sur des mots-clés.

Points-clés du bilan 2022

- ➔ Déclaration relativement exhaustive sur les mesures de lutte contre la manipulation de l'information.
- ➔ Absence de transparence sur les données chiffrées relatives à la lutte contre la manipulation de l'information en France justifiée, pour certains, par le modèle du service.
- ➔ Commentaire d'une partie des précédentes préconisations de l'Arcom mais pas de mention de la manière de les prendre en compte par l'opérateur.
- ➔ Dispositif de signalement peu visible, peu accessible et peu lisible (particulièrement sur la version mobile).
- ➔ Pas de suppression de contenus, sauf situation exceptionnelle pour le contenu jugé illégal en vertu de la législation locale.
- ➔ En matière de lutte contre les biais algorithmiques, mention de nombreux éléments concernant les « bonnes pratiques » guidant le déploiement de ses technologies et les processus d'évaluation qui en découlent.
- ➔ Intégration du module *Newsguard* notant la fiabilité des sites d'information (uniquement disponible sur le navigateur *Microsoft Edge*).
- ➔ Fonctionnalités d'identification des contenus provenant d'éditeurs de presse qui sont labellisés grâce à des initiatives de *fact-checking*.
- ➔ Pas d'information sur d'éventuelles pratiques de manipulation sur le service, l'opérateur s'estimant non concerné au regard du modèle de son service.
- ➔ Recherches menées pour lutter contre les vides de données (absence d'informations fiables sur un sujet recherché).
- ➔ Financement des programmes d'éducation aux médias et à l'information de *Newsguard*.
- ➔ Mise en place de nombreux partenariats de recherche pour développer des outils permettant l'identification de pratiques ou phénomènes de manipulation de l'information.
- ➔ Mise à disposition de données issues de son algorithme en *open source*.

Bing : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Microsoft Ireland Operations Limited (Irlande)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	Non déclaré
Chiffre d'affaires du service en France	Non déclaré
Effectifs employés aux activités du service en France	Non déclaré
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	Microsoft, au regard du modèle de son service, ne s'estime pas concerné par ces indicateurs ¹⁰⁸
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	Non déclaré
Délai moyen de traitement d'un signalement	Non déclaré
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	Non déclaré
Nombre de comptes détectés propageant massivement de fausses informations	Microsoft, au regard du modèle de son service, ne s'estime pas concernée par ces indicateurs
Nombre de faux comptes automatisés détectés	
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	Non déclaré

¹⁰⁸ L'opérateur rappelle que « la politique de Bing n'est pas de supprimer les fausses informations des recherches, mais plutôt d'améliorer le classement afin que le contenu le plus pertinent et de haute autorité apparaisse en premier dans les résultats de recherche. »



Dailymotion

Dailymotion est une plateforme d'hébergement et de partage de vidéos, en direct ou en différé, autour de quatre thématiques principales : les actualités, la musique, le divertissement et le sport.

Points-clés du bilan 2022

- ➔ Efforts de transparence et de coopération : réponse à la majorité des questions de l'Arcom, commentaires de la mise en œuvre des précédentes préconisations de l'Arcom, fourniture de davantage de données chiffrées et recours plus faible à la confidentialité.
- ➔ Toutefois, peu d'informations sur le fonctionnement des outils de recommandation et pas d'éléments concernant les systèmes de modération et les processus d'évaluation des algorithmes.
- ➔ Existence d'un dispositif de signalement avec un motif « désinformation » dont la notion est définie dans sa politique relative aux contenus prohibés.
- ➔ Canal de traitement dédié à la désinformation ; les signalements sont traités en moins de 24 heures.
- ➔ En période d'urgence et en période électorale, priorisation de l'examen des contenus signalés pour désinformation et désindexation de ces contenus le temps de l'instruction.
- ➔ Intervention humaine maintenue dans toute décision d'intervention sur un contenu.
- ➔ Recommandation algorithmique : uniquement des contenus publiés par les partenaires professionnels.
- ➔ Labellisation des comptes des partenaires professionnels vérifiés avec un badge clairement identifiable.
- ➔ Pas de recours à des dispositifs de *fact-checking* internes ou externes.
- ➔ Pas de mesure de lutte contre les comptes propageant massivement de fausses informations.
- ➔ Annonces publicitaires vérifiées par les équipes de l'opérateur ou ses partenaires.
- ➔ Aucune mesure d'éducation aux médias et à l'information.
- ➔ Travail ponctuel avec des chercheurs et volonté de l'intensifier dans le cadre du DSA.

Dailymotion : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Dailymotion SA (France)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	44 000 000
Chiffre d'affaires du service en France	<i>Déclaré de manière confidentielle</i>
Effectifs employés aux activités du service en France	<i>Déclaré de manière confidentielle</i>
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	13
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	0 et 13
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	12
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	2 comptes supprimés
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	3 800 vues
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	574
Délai moyen de traitement d'un signalement	Moins de 2 heures
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	<i>Non déclaré</i>
Nombre de comptes détectés propageant massivement de fausses informations	« Dailymotion n'a pas estimé nécessaire de mettre en place des mesures concernant le partage « massif » de fausses informations. »
Nombre de faux comptes automatisés détectés	
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	0

Facebook et Instagram

La déclaration de [Meta](#) concerne deux de ses services : [Facebook](#) et [Instagram](#). [Facebook](#) est un réseau social permettant à des utilisateurs publier des contenus, de créer des groupes privés et de réagir aux publications d'autres comptes. [Instagram](#) est un réseau social de partage de photos et vidéos. Les deux services proposent une messagerie privée et permettent de publier des contenus éphémères ou en direct.

Points-clés du bilan 2022

- Une déclaration claire, relativement exhaustive et répondant à chacune des interrogations de l'Arcom.
- Pas de données chiffrées relatives à la lutte contre la manipulation de l'information en France.
- Commentaire d'une partie des précédentes préconisations de l'Arcom, avec des explications sur les raisons pour lesquelles certaines ne sont pas suivies (ou pas encore suivies).
- Dispositif de signalement accessible et visible proposant plusieurs motifs et sous-motifs relatifs à la manipulation de l'information.
- Aucune mention de la manipulation de l'information dans le dispositif de signalement des contenus promus sur [Instagram](#).
- Information fournie à l'utilisateur affecté par une décision de modération
- Possibilité de faire appel offerte à l'auteur du signalement une fois la décision prise.
- De nombreux éléments déclarés concernant la recommandation de contenus sur la page d'accueil des deux services et quelques éléments relatifs aux outils de détection de contenus.
- Sur ses deux services, mise en place d'une fonctionnalité (« Pourquoi est-ce que je vois ça ? ») pour comprendre, contextuellement et individuellement, les principaux motifs de recommandation du contenu.
- Évaluation des outils algorithmiques s'appuyant notamment sur une équipe dédiée à l'IA responsable.
- Identification des médias « contrôlés par l'État » sur les deux services.
- Lancement de l'espace « Facebook News » rassemblant des sources d'information jugées fiables¹⁰⁹.
- Travaux avec des chercheurs sur l'analyse des données des deux services à grande échelle et lancement de la *Researcher Platform* (en **2022**).

¹⁰⁹ Néanmoins, en octobre 2022, l'AFP a annoncé que Meta mettait fin à son contrat avec Media Services à compter du 1er janvier 2023. Les contenus répertoriés par « Facebook News » seront sélectionnés selon une méthode de curation uniquement algorithmique. Les contrats passés avec les éditeurs de presse repris sur cet espace restent, quant à eux, effectifs jusqu'à la fin de l'année 2024.

Facebook et Instagram : données chiffrées déclarées		
Informations générales	Facebook	Instagram
Nom et pays d'établissement de la société opératrice	Meta Platforms Ireland Limited (Irlande)	
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	<i>Déclaré de manière confidentielle</i>	<i>Déclaré de manière confidentielle</i>
Chiffre d'affaires du service en France	<i>Non déclarés¹¹⁰</i>	
Effectifs employés aux activités du service en France	292 (« effectif moyen »)	
Indicateurs généraux (en 2021 en France)	Facebook	Instagram
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i>	<i>Non déclaré</i>
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	<i>Déclaré de manière confidentielle</i>	<i>Déclaré de manière confidentielle</i>
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	<i>Non déclaré</i>	<i>Non déclaré</i>
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i>	<i>Non déclaré</i>
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	Facebook	Instagram
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	<i>Non déclaré</i>	<i>Non déclaré</i>
Délai moyen de traitement d'un signalement	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de comptes détectés propageant massivement de fausses informations	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de faux comptes automatisés détectés	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	<i>Non déclaré</i>	<i>Non déclaré</i>

¹¹⁰ L'opérateur renvoie, dans sa déclaration, au site Infogreffe.fr pour le chiffre d'affaires de sa filiale française (879,3 millions d'euros) et non celui de ses services en France.

Google Search et YouTube

La déclaration de [Google](#) concerne deux de ses services : [Google Search](#) et [YouTube](#). [Google Search](#) est un moteur de recherche permettant à ses utilisateurs de trouver des ressources en ligne. [YouTube](#) est une plateforme de partage de vidéos, en direct ou en différé, sur laquelle les utilisateurs peuvent interagir dans les espaces « commentaires ».

Points-clés du bilan 2022

- Déclaration pédagogique avec des réponses apportées aux questions spécifiques à [Google](#) et [YouTube](#), mais pas de réponse à certaines interrogations de l'Arcom, omettant ainsi plusieurs enjeux.
- Aucune donnée chiffrée relative à la lutte contre la manipulation de l'information en France et peu d'informations sur les moyens développés en 2021. De manière générale, manque d'informations relatives à [YouTube](#).
- Absence de commentaires sur la mise en œuvre des précédentes préconisations de l'Arcom.
- Dispositif de signalement de [Google Search](#), peu visible, peu accessible et peu lisible malgré l'ajout d'une catégorie « fausses informations ».
- Dispositif de signalement de [YouTube](#) peu visible (en particulier sur téléphone mobile) et sans motif ou sous-motif clairement relatif à la manipulation de l'information.
- Sur les deux services, outil de signalement des communications commerciales peu accessible.
- Action de réduction de la visibilité des contenus de mauvaise qualité (sans que l'utilisateur ayant mis en ligne le contenu n'en soit informé).
- Aucune information sur les processus d'évaluation des algorithmes.
- Aucune information sur les mesures particulières pour lutter contre les risques de manipulation qui découlent de la diffusion de contenus en direct (sur [YouTube](#)).
- Liste des campagnes d'influence déjouées en 2021 par le groupe d'analyse des menaces (« *Threat Analysis Group* »).
- Détection d'actions de fraude coordonnées impliquant simultanément des milliers de comptes.
- Campagne éducative sur les risques liés à l'utilisation de [YouTube](#).
- Lancement, **en 2022**, d'un site dédié pour permettre aux chercheurs de demander un accès spécifique à l'API de [YouTube](#).

Google Search et YouTube : données chiffrées déclarées		
Informations générales	Google Search	YouTube
Nom et pays d'établissement de la société opératrice	Google Ireland Limited (Irlande)	
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	<i>Non déclaré</i>	<i>Non déclaré</i>
Chiffre d'affaires du service en France	<i>Non déclaré</i>	<i>Non déclaré</i>
Effectifs employés aux activités du service en France	<i>Non déclaré</i>	<i>Non déclaré</i>
Indicateurs généraux (en 2021 en France)	Google Search	YouTube
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	0 ¹¹¹	0 ¹¹²
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	<i>Non déclaré</i>	<i>Non déclaré</i>
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i>	<i>Non déclaré</i>
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	Google Search	YouTube
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	<i>Non déclaré</i>	<i>Non déclaré</i>
Délai moyen de traitement d'un signalement	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de comptes détectés propageant massivement de fausses informations	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de faux comptes automatisés détectés	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	<i>Non déclaré</i>	<i>Non déclaré</i>
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	<i>Non déclaré</i>	<i>Non déclaré</i>

^{111 8} L'opérateur déclare que « aucun signalement sur le fondement de la n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information n'a été enregistré » sur ses deux services.



Jeuxvideo.com

Le site internet [Jeuxvideo.com](https://jeuxvideo.com) opéré par la société [Webedia](https://webedia.fr) est consacré à l'écosystème du jeu vidéo et de la culture « geek » en général. Outre une partie éditée en propre par l'opérateur, il comporte un forum de discussion sur des sujets divers qui constitue une plateforme en ligne et fait l'objet de la déclaration.

Points-clés du bilan 2022

- ➔ Déclaration répondant à une majorité des questions du régulateur (sauf domaines sur lesquels l'opérateur ne s'estime pas concerné au regard du modèle du service).
- ➔ Transparence plus importante que les années précédentes sur les données chiffrées et recours plus faible à la confidentialité.
- ➔ Commentaire d'une partie des précédentes préconisations de l'Arcom.

- ➔ Dispositif de signalement accessible et visible.
- ➔ Pas de motif de signalement spécifique aux fausses informations.
- ➔ Ensemble des décisions d'intervention prises à l'issue d'une intervention humaine.

- ➔ Aucune utilisation d'algorithmes de recommandation sur le service, le tri des contenus s'effectuant de manière chronologique ou antéchronologique.

- ➔ Pas de mesures de promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle au regard du modèle du service.

- ➔ En cas de « *flood* » (pratique consistant à la multi-publication d'un même contenu pour le propager un maximum), les comptes s'y adonnant peuvent être avertis, bannis temporairement ou définitivement.
- ➔ Utilisation de « Captcha » pour prévenir de la diffusion massive de fausses informations en plus de la modération humaine.

- ➔ Gestion des paramètres publicitaires qui repose sur la technologie des cookies tiers.

- ➔ Absence de mesures d'éducation aux médias et à l'information.

Jeuxvideo.com : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Webedia (France)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	5,7 millions (2021) ¹¹³
Chiffre d'affaires du service en France	<i>Déclaré de manière confidentielle</i>
Effectifs employés aux activités du service en France	56 personnes ¹¹⁴
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i>
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	<i>Non déclaré</i> ¹¹⁵
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	<i>Non déclaré</i>
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	<i>Non déclaré</i>
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i>
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	<i>Non déclaré</i>
Délai moyen de traitement d'un signalement	Moins de 24 heures
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	15 428 recours tout motif de signalement confondu
Nombre de comptes détectés propageant massivement de fausses informations	0
Nombre de faux comptes automatisés détectés	0
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	0
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	0

¹¹³ L'opérateur n'a pas précisé, dans sa déclaration, le nombre de visiteurs uniques par mois en **2020**.

¹¹⁴ L'opérateur précise que ces 56 personnes composent les équipes éditoriales et produit en France (soit hors équipes techniques, commerciales et fonctions support, ces dernières étant mutualisées au sein du groupe Webedia).

¹¹⁵ L'opérateur précise ne pas disposer d'outils automatisés pour identifier ces contenus.



LinkedIn

LinkedIn est un réseau social axé sur la vie professionnelle et l'économie, avec une partie « premium » permettant d'accéder à davantage de fonctionnalités.

Points-clés du bilan 2022

- ➔ Déclaration qui se distingue par sa technicité et fait état des principales mesures de lutte contre la manipulation de l'information mais ne répond pas aux interrogations spécifiques de l'Arcom, omettant ainsi plusieurs enjeux (notamment sur la transparence des algorithmes).
- ➔ Communication d'une majorité des données chiffrées de manière confidentielle, permettant une meilleure compréhension des enjeux par le régulateur mais témoignant d'une faible transparence auprès du public.
- ➔ Absence de commentaires sur la mise en œuvre des précédentes préconisations de l'Arcom.
- ➔ Dispositif de signalement visible et accessible, proposant un motif « Douteux, spam ou faux » peu clair et un sous-motif relatif à la « désinformation ».
- ➔ Appel des vérificateurs de faits spécialisés pour des situations ou contenus spécifiques.
- ➔ Absence de transparence concernant le fonctionnement des algorithmes de recommandation et de modération.
- ➔ Évaluation des outils automatiques en amont de leur mise en production et suivi de leur performance après leur déploiement.
- ➔ Identification des faux comptes grâce à des systèmes automatisés dès leur enregistrement.
- ➔ Aucune mesure d'éducation aux médias et à l'information.
- ➔ Aucune collaboration avec le monde de la recherche déclarée.

LinkedIn : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	LinkedIn Ireland Unlimited Company (Irlande)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	Non déclaré
Chiffre d'affaires du service en France	Déclaré de manière confidentielle
Effectifs employés aux activités du service en France	Déclaré de manière confidentielle
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	33 849
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	Déclaré de manière confidentielle
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	Déclaré de manière confidentielle
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	Déclaré de manière confidentielle
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	Déclaré de manière confidentielle
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	55 576
Délai moyen de traitement d'un signalement	6 h
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	304
Nombre de comptes détectés propageant massivement de fausses informations	Déclaré de manière confidentielle
Nombre de faux comptes automatisés détectés	Déclaré de manière confidentielle
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	Non déclaré
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	Non déclaré



Pinterest

[Pinterest](#) est un réseau social et moteur de recherche visuel permettant à ses utilisateurs de partager leurs centres d'intérêt et passions à travers des albums de photographies.

Points-clés du bilan 2022

- ➔ Déclaration qui ne répond pas aux interrogations spécifiques de l'Arcom, omettant ainsi de nombreux enjeux.
- ➔ Très peu d'informations relatives à l'éducation aux médias et à l'information, à la transparence des algorithmes et à la promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle.
- ➔ Communication d'une majorité des données chiffrées de manière confidentielle, permettant une meilleure compréhension des enjeux par le régulateur mais témoignant d'une faible transparence auprès du public.
- ➔ Dispositif de signalement accessible et visible proposant plusieurs motifs et sous-motifs relatifs à la manipulation de l'information.
- ➔ Un motif de signalement spécifique à la désinformation sur le climat.
- ➔ Processus de décision d'une action à l'égard d'un contenu ou d'un compte basé sur une approche hybride conjuguant examen manuel et automatisé.
- ➔ Absence d'informations sur les outils de recommandation et informations très générales sur la modération automatique de contenus.
- ➔ Manque de transparence concernant les pratiques ou phénomènes pouvant porter atteinte à l'intégrité de son service.
- ➔ Origine des annonceurs indiquée dans les dispositifs d'information à proximité des contenus.
- ➔ Aucune mesure relative à l'éducation aux médias et à l'information.
- ➔ Aucune collaboration avec le monde de la recherche déclarée.

Pinterest : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Pinterest Europe Limited (Irlande)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	<i>Déclaré de manière confidentielle</i>
Chiffre d'affaires du service en France	<i>Déclaré de manière confidentielle</i>
Effectifs employés aux activités du service en France	48 équivalents temps-plein
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Déclaré de manière confidentielle</i>
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	<i>Déclaré de manière confidentielle</i>
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	<i>Déclaré de manière confidentielle</i>
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	<i>Déclaré de manière confidentielle</i>
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Déclaré de manière confidentielle</i>
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	<i>Déclaré de manière confidentielle</i>
Délai moyen de traitement d'un signalement	<i>Déclaré de manière confidentielle</i>
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	<i>Non déclaré</i>
Nombre de comptes détectés propageant massivement de fausses informations	<i>Non déclaré</i>
Nombre de faux comptes automatisés détectés	<i>Non déclaré</i>
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	<i>Non déclaré</i>
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	<i>Non déclaré</i>



Snapchat

Snapchat est une application proposant un service de messagerie éphémère, une plateforme fermée avec des partenaires média (*Discover*) et des espaces de partage de photos et vidéos entre les utilisateurs (*Map*, *Spotlight*)

Points-clés du bilan 2022

- ➔ Des réponses apportées à une majorité des interrogations de l'Arcom.
- ➔ Transparence plus importante que les années précédentes sur les données chiffrées et recours plus faible à la confidentialité.
- ➔ Commentaire des précédentes préconisations de l'Arcom et explications sur les changements qui ont eu lieu en conséquence le cas échéant.
- ➔ Dispositif de signalement accessible et visible proposant plusieurs motifs et sous-motifs relatifs à la manipulation de l'information.
- ➔ Décisions d'intervention prises à l'issue d'une intervention humaine.
- ➔ Peu d'informations relatives aux critères de recommandation de contenu sur la partie *Spotlight* du service.
- ➔ Base de données publicitaire permettant de recenser les publicités politiques ou se rattachant à un débat d'intérêt général.
- ➔ Procédures de vérification des annonceurs et des publicités en amont, par une équipe dédiée.
- ➔ Mise à disposition de ressources générales dans des espaces dédiées pour appeler l'utilisateur à une utilisation maîtrisée de son service.
- ➔ Aucune collaboration avec le monde de la recherche déclarée.

Snapchat : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Snap Inc. (États-Unis)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	21 881 000
Chiffre d'affaires du service en France	<i>Déclaré de manière confidentielle</i>
Effectifs employés aux activités du service en France	<i>Déclaré de manière confidentielle</i>
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	971
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	0 et 971
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	971
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	971
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	Taux d'affichage : ~0,09 % ¹¹⁶
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	35 617
Délai moyen de traitement d'un signalement	4,3 h
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	0
Nombre de comptes détectés propageant massivement de fausses informations	<i>Snap, au regard du modèle de son service, ne s'estime pas concerné par ces indicateurs</i>
Nombre de faux comptes automatisés détectés	
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	Non déclaré
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	<i>Déclaré de manière confidentielle</i>

¹¹⁶ L'opérateur précise ne pas différencier les catégories des contenus non-conformes dans le calcul de son taux d'affichage.



TikTok

TikTok est une plateforme centrée sur les vidéos au format court. Les utilisateurs peuvent visionner des contenus et des *lives*, interagir avec d'autres utilisateurs par messagerie privée ou créer et partager des vidéos.

Points-clés du bilan 2022

- ➔ Déclaration particulièrement imprécise avec peu d'informations relatives au service en France et aucun élément tangible permettant une évaluation des moyens mis en place.
- ➔ Pas de réponse aux interrogations spécifiques de l'Arcom, omettant ainsi plusieurs enjeux centraux (étapes de l'instruction d'un signalement, fonctionnement des outils de modération, pratiques de manipulation identifiées sur le service, etc.).
- ➔ Pas d'éléments chiffrés relatifs à la France.
- ➔ Dispositif de signalement accessible mais peu visible (accessible via une icône en forme de « flèche » symbolisant le repartage du contenu).
- ➔ Plusieurs motifs et sous-motifs relatifs à la manipulation de l'information.
- ➔ Des éléments d'intérêt mais très généraux concernant les critères utilisés pour la recommandation de contenus.
- ➔ Absence d'informations quant aux processus d'évaluation des algorithmes.
- ➔ Recours à des partenaires de *fact-checking* accrédités par l'*International Fact-Checking Network*.
- ➔ Développement en 2022 d'une politique dédiée aux « médias contrôlés par un Etat »¹¹⁷.
- ➔ Aucune information sur les pratiques de manipulation de l'information identifiées sur le service.
- ➔ Mention de dispositifs de défense automatisés pour détecter, bloquer et supprimer les comptes et engagements inauthentiques mais peu d'éléments permettant d'apprécier leur pertinence.
- ➔ Vérifications supplémentaires des communications commerciales lorsqu'elles dépassent un certain seuil d'impression et examen d'un échantillon aléatoire de publicités pour évaluer leur conformité.
- ➔ Plusieurs actions d'éducation aux médias et à l'information lancés par l'opérateur ou en partenariat avec des tiers, sur le développement de l'esprit critique de l'utilisateur.

¹¹⁷ L'Arcom a constaté, en septembre 2022, une mise en œuvre sélective de cette règle pourtant non finalisée. À titre d'exemple, le compte de *China Global Television Network Europe* (CGTN Europe) n'est pas identifié comme tel.

TikTok : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	TikTok Technology Limited (Irlande)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	<i>Non déclaré</i>
Chiffre d'affaires du service en France	<i>Non déclaré</i>
Effectifs employés aux activités du service en France	<i>Non déclaré</i>
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i>
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	<i>Non déclaré</i>
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	<i>Non déclaré</i>
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	<i>Non déclaré</i>
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i>
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	<i>Non déclaré</i>
Délai moyen de traitement d'un signalement	<i>Non déclaré</i>
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	<i>Non déclaré</i>
Nombre de comptes détectés propageant massivement de fausses informations	<i>Non déclaré</i>
Nombre de faux comptes automatisés détectés	<i>Non déclaré</i>
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	<i>Non déclaré</i>
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	<i>Non déclaré</i>



Twitter

Twitter est un réseau social permettant aux utilisateurs d'interagir par des messages limités à 280 caractères, pouvant comprendre des images ou des vidéos.

Points-clés du bilan 2022

- Des réponses apportées à une majorité des interrogations de l'Arcom malgré des éléments peu précis concernant le fonctionnement des outils automatiques.
- Une transparence très relative concernant les données chiffrées.
- Dispositif de signalement accessible et visible mais dont le motif général relatif à la manipulation de l'information couvre uniquement la désinformation en contexte électoral (mention du trouble à l'ordre public uniquement en sous-motif).
- Modération basée majoritairement sur la technologie.
- Pas de précisions sur les critères utilisés pour la recommandation de contenus à l'exception d'informations relativement générales, ni sur le fonctionnement des outils de détection de fausses informations (ou notion équivalente).
- Action de réduction de la visibilité de certains contenus, notamment ceux partageant du contenu trompeur ou provenant de médias affiliés à des États.
- De nombreux éléments déclarés en matière de processus d'évaluation des algorithmes réalisés en internes (équipe META : « *Machine Learning Ethics, Transparency and Accountability* »)¹¹⁸ et par des experts externes.
- Recours au *fact-checking* limité aux périodes de « risque aigu ».
- Déclaration de nombreux éléments quant aux pratiques et phénomènes pouvant porter atteinte à l'intégrité du service identifiées sur ce dernier mais des informations lacunaires sur le fonctionnement des outils de lutte à leur encontre.
- Publication de rapports réguliers sur les opérations d'influence identifiées et mise à disposition de l'archive des contenus relatifs aux opérations de désinformation étrangère afin de collaborer avec la société civile.
- Absence de campagnes d'éducation aux médias et à l'information à l'exception de la publication d'un *Digital Safety Book*.
- API ouverte permettant l'accessibilité d'une partie des données
- Travaux menés en collaboration avec des organismes de recherche internationaux.

¹¹⁸ L'Arcom considère que ce travail d'évaluation des risques et de transparence constituait une bonne pratique particulièrement notable ; aussi, elle prend note avec regret de l'annonce de la suppression de l'équipe META en novembre 2022, à la suite du rachat de la société.

Twitter : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Twitter International Unlimited Company (Irlande)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	<i>Non déclaré</i> ¹¹⁹
Chiffre d'affaires du service en France	11 820 329
Effectifs employés aux activités du service en France	34 équivalents temps-plein
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	63 786 relatifs au Covid-19 et 593 relatifs à l'intégrité civique
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	<i>Non déclaré</i>
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	Covid-19 : • 353 018 comptes mis en cause • 3 455 comptes suspendus • 63 786 contenus uniques supprimés Intégrité civique : • 581 comptes ayant fait l'objet de mesures • 23 comptes suspendus • 593 contenus uniques supprimés
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions etc.)	
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	<i>Non déclaré</i> ¹²⁰
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	219 905
Délai moyen de traitement d'un signalement	<i>Non déclaré</i>
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	<i>Non déclaré</i>
Nombre de comptes détectés propageant massivement de fausses informations	12 réseaux et 3 838 comptes
Nombre de faux comptes automatisés détectés	<i>Non déclaré</i>
Nombre de <i>deepfakes</i> sur lesquels des mesures ont été prises	<i>Non déclaré</i>
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	<i>Non déclaré</i>

¹¹⁹ L'opérateur précise ne pas calculer le nombre de visiteurs uniques mensuels mais uniquement le nombre d'utilisateurs actifs quotidiens monétisables au niveau mondial (192 millions pour le quatrième semestre 2020).

¹²⁰ L'opérateur précise que s'il ne récolte pas ces données : « Du 1^{er} janvier 2021 au 30 juin 2021, Twitter a supprimé 4,7 millions de Tweets qui enfreignaient les Règles de Twitter. Parmi les Tweets supprimés, 68% ont reçu moins de 100 impressions avant leur suppression, et 24% ont reçu entre 100 et 1 000 impressions. Seuls 8% des Tweets supprimés ont reçu plus de 1 000 impressions. Au total, les impressions des Tweets en infraction représentaient moins de 0,1 % de toutes les impressions de tous les Tweets pendant cette période. »



Wikipédia

Wikipédia est une encyclopédie en ligne à but non lucratif. Opérée par la [Fondation Wikimédia](#), elle fonctionne grâce à des contributeurs et des administrateurs élus pour « *protéger l'encyclopédie collaborative* » qui est régie par des règles et principes fondateurs établis par la communauté de volontaires.

Points-clés du bilan 2022

- ➔ Déclaration claire et didactique, dans un format différent du fait de la spécificité du modèle.
- ➔ Absence d'éléments chiffrés permettant d'évaluer le niveau de risques en matière de manipulation de l'information.
- ➔ Absence de commentaires concernant les préconisations du précédent bilan.
- ➔ Pas d'outil de signalement en raison de l'approche spécifique de la plateforme : les informations qui y sont communiquées font l'objet de vérifications collaboratives.
- ➔ Mise en place de « patrouilles » de veille constante (équipes bénévoles) pouvant prendre des mesures allant jusqu'à empêcher un utilisateur de contribuer de manière temporaire ou permanente (en cas de « vandalisme »).
- ➔ Pas d'outils automatique de modération actuellement, du fait de la spécificité du modèle et de la modération collaborative, mais un outil de détection automatique des noms d'utilisateurs utilisés par une même personne est en cours de développement.
- ➔ Outils automatiques de modération qui font l'objet de recherches académiques spécifiques à la mesure de leurs éventuels biais.
- ➔ Publication d'articles relatifs aux campagnes d'influence identifiées sur le service.
- ➔ Mention de quelques opérations identifiées sur son service en contexte électoral, du travail des équipes *Trust and Safety* mené en conséquence et de la transparence sur les mesures prises en réponse.
- ➔ L'encyclopédie ne permet pas de payer pour référencer, supprimer ou promouvoir un article.
- ➔ Mise en place du *Wikipedia Education Program* pour inciter les professeurs d'université à apprendre à maîtriser les outils de contribution sur son service.

Wikipédia : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Wikimedia Foundation Inc. (États-Unis)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	Non déclaré ¹²¹
Chiffre d'affaires du service en France	Non déclaré
Effectifs employés aux activités du service en France	0
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	La Fondation Wikimedia, au regard du modèle de son service, ne s'estime pas concernée par ces indicateurs ¹²²
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	La Fondation Wikimedia, au regard du modèle de son service, ne s'estime pas concernée par ces indicateurs
Délai moyen de traitement d'un signalement	
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	
Nombre de comptes détectés propageant massivement de fausses informations	
Nombre de faux comptes automatisés détectés	Non déclaré
Nombre de deepfakes sur lesquels des mesures ont été prises	La Fondation Wikimedia, au regard du modèle de son service, ne s'estime pas concernée par ces indicateurs
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	« Wikipédia ne comporte pas d'encarts publicitaires. »

¹²¹ L'opérateur déclare 94 millions d'utilisateurs non pas pour 2020 mais pour l'année 2021, et non pas en France mais de son service en français.

¹²² L'opérateur précise néanmoins que « en 2021, environ 5 300 protections d'articles ont été mises en œuvre sur Wikipédia en français. »

Yahoo Media

Le service [Yahoo Media](#) propose des contenus produits par une équipe éditoriale dédiée et effectue de la curation de contenus d'actualités produits par des tiers. Le service regroupe différentes fonctionnalités (*Yahoo Homepage, Yahoo News, Yahoo Style, Yahoo Sport* et *Yahoo Finance*).

Points-clés du bilan 2022

- ➔ Déclaration légèrement plus fournie que l'année passée mais une transparence qui reste très insuffisante aussi bien dans la forme (en anglais) que dans la quantité d'informations déclarées.
- ➔ Une absence totale de transparence concernant les données chiffrées.
- ➔ Absence d'un dispositif de signalement à l'exception des communications commerciales.
- ➔ Pas de définition des fausses informations.
- ➔ Le contenu provient uniquement de sources sélectionnées et non d'utilisateurs privés.
- ➔ Interdiction des publicités se rattachant à un débat d'intérêt général.
- ➔ Absence de mesures d'éducation aux médias et à l'information.

Yahoo Media : données chiffrées déclarées	
Informations générales	
Nom et pays d'établissement de la société opératrice	Yahoo EMEA Limited (Irlande)
Nombre de visiteurs uniques par mois du service en France et dans le monde, calculé sur la base de l'année précédant l'année d'exercice (moyenne mensuelle en 2020)	Déclaré de manière confidentielle
Chiffre d'affaires du service en France	Déclaré de manière confidentielle
Effectifs employés aux activités du service en France	Déclaré de manière confidentielle
Indicateurs généraux (en 2021 en France)	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	Yahoo, au regard du modèle de son service, ne s'estime pas concerné par ces indicateurs
Parmi ces contenus, nombres détectés automatiquement et à la suite de signalements utilisateurs	
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures	
Nombre d'émetteurs de ces contenus identifiés et part qui a fait l'objet d'une mesure (avertissement, restrictions, etc.)	
Audience moyenne des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	
Indicateurs spécifiques aux domaines de la loi (en 2021 en France)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	Yahoo, au regard du modèle de son service, ne s'estime pas concerné par ces indicateurs
Délai moyen de traitement d'un signalement	
Nombre de recours reçus par l'opérateur à l'issue de la procédure de signalement	
Nombre de comptes détectés propageant massivement de fausses informations	
Nombre de faux comptes automatisés détectés	
Nombre de deepfakes sur lesquels des mesures ont été prises	
Nombre de communications commerciales porteuses de fausses informations (ou notion voisine adoptée)	Non déclaré