

# **Twitch – Déclaration annuelle à l'ARCOM en application du titre III de la loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information**

## **Introduction**

Twitch est un service de partage de vidéo en direct (« streaming »), axé principalement sur l'univers du jeu vidéo, mais proposant également des contenus liés à la musique, aux loisirs créatifs et culinaires. La grande majorité du contenu diffusé sur Twitch n'est disponible qu'en direct. Cette caractéristique implique de penser la sécurité et le bien-être de notre communauté différemment par rapport aux services reposant sur le contenu préenregistré mis en ligne de manière durable. Cela signifie également que la désinformation ne se propage pas sur Twitch comme sur d'autres services.

Le modèle de diffusion de Twitch n'est pas propice en soi à la propagation de désinformation ou aux campagnes de désinformation à grande échelle. En effet, il est extrêmement difficile pour un nouvel utilisateur d'attirer une large audience en direct. Développer une audience sur Twitch prend du temps. Qui plus est, la plupart des contenus sont éphémères et diffusés uniquement en direct, ce qui limite fortement le potentiel de partage et de viralité du contenu.

Si la désinformation n'est pas répandue sur Twitch, nous sommes toutefois conscients qu'elle pourrait causer des dommages importants si nous l'autorisons. C'est pourquoi nous avons travaillé avec une équipe d'une douzaine de chercheurs et d'experts afin d'analyser les mécanismes de diffusion de la désinformation préjudiciable en ligne et de nous assurer de l'efficacité de notre stratégie de prévention des risques eu égard à notre modèle et à nos spécificités en tant que service. Il est apparu que la grande majorité de la désinformation préjudiciable et largement démentie qui circule en ligne est propagée par des agents de désinformation préjudiciable ayant en commun les trois caractéristiques suivantes: leur présence en ligne, sur Twitch ou ailleurs sur internet, est dédiée au (1) partage continu (2) de sujets de désinformation préjudiciable (3) largement discrédités et partagés en masse, telles que des théories conspirationnistes incitant à la violence. Twitch exclut les agents de désinformation préjudiciable correspondant à ces trois critères, lesquels représentent l'essentiel du risque de préjudice, et notamment le risque d'incitation à des comportements dommageables et dangereux dans la vie réelle.

Nos règles contre les « [Agents de désinformation préjudiciable](#) » sont entrées en vigueur en mars 2022. Entre mars et décembre 2022, nous avons suspendu un total de 111 agents de désinformation préjudiciable à l'échelle du service. Ces chiffres sont relativement bas et reflètent la présence limitée de la désinformation sur Twitch.

Bien que la désinformation préjudiciable ne soit pas actuellement répandue sur Twitch, nous savons que cela peut changer. Nous faisons constamment évoluer notre approche en fonction des recommandations des experts et des tendances observées auprès de notre communauté d'utilisateurs, et continuons de nouer des partenariats avec le monde universitaire et la société civile. De même, en tant que signataire du Code de bonnes pratiques de l'Union européenne en matière de désinformation, nous continuerons à travailler de concert avec les pouvoirs publics et les acteurs du secteur. Nous nous engageons à lutter contre la désinformation sur Twitch de façon efficace et ciblée, en conciliant protection de notre communauté d'utilisateurs d'une part, et liberté d'expression d'autre part.

## **Questions liminaires : présentation du service**

- **Nom du service** : Twitch
- **Nom, raison sociale et domiciliation de la société opératrice**: Twitch Inc., 350 Bush Street, 2nd Floor, San Francisco, CA 94104 (États-Unis)
- **Identité et fonction du représentant légal exerçant les fonctions d'interlocuteur référent sur le territoire français au sens de l'article 13 de la loi du 22 décembre 2018** : Pour les besoins de ce rapport uniquement, la personne référente est Olaf Cramme, vice-président, Global Public Policy (crammeoc@twitch.tv).

- **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 2022) :** 6,41 millions (en France)
- **Modalités de mise à disposition du service auprès du public français (web, applications, assistants vocaux, etc.) :** Le service est disponible en français sur [www.twitch.tv](http://www.twitch.tv) ou via l'application Twitch (iOS et Android).
- **Langue du service proposée par défaut sur le territoire français :** La langue par défaut est le français pour les utilisateurs situés en France.
- **Le service disponible sur le territoire français constitue-t-il une version spécifique du service ?** Il s'agit de la version française de Twitch, qui est accessible à tous les utilisateurs disposant d'une adresse IP.
- **Description du service (objet, population(s) ciblée(s), types de contenus publiés et consommés, thématiques, etc.) :** Twitch est un service de partage de vidéo en direct (« streaming »), axé principalement sur l'univers du jeu vidéo, mais proposant également des contenus liés à la musique, aux arts et loisirs créatifs et culinaires. Twitch propose à sa communauté d'utilisateurs un contenu de haute qualité et attrayant, créé par une grande diversité de streamers et streameuses, des fonctionnalités innovantes et performantes, et une expérience en ligne sûre qui sont un facteur d'attractivité auprès des utilisateurs du service – spectateurs comme streamers. Twitch permet à titre principal de visionner des contenus en direct autour d'un centre d'intérêt commun. Les créateurs et créatrices de contenu peuvent mettre à disposition un nombre restreint d'enregistrements ou d'extraits de leurs streams sur une période limitée. Twitch emploie 8 personnes pour les besoins de son activité en France.
- **Modèle économique du service (type d'activités rémunératrices pour l'opérateur tirées de l'exploitation du service en France) :** Les activités lucratives de Twitch comprennent la vente d'abonnements mensuels pour soutenir les streamers et streameuses à titre individuel, les « Bits » (que l'audience peut utiliser pour montrer son appréciation) et la publicité.

## Questions générales

Twitch a pour mission de développer des communautés d'utilisateurs accueillantes et interactives, au sein desquelles les internautes du monde entier se sentent en sécurité et à l'aise pour s'exprimer. Nous mettons en place un environnement favorable pour les streamers, accueillant et divertissant pour les spectateurs, et mettons tout en œuvre pour limiter les expériences négatives sur le service. En particulier, Twitch s'efforce de dissuader les comportements nuisibles en fournissant aux streamers les lignes de conduite, les outils, la technologie et toutes les informations nécessaires à la modération de leur communauté.

Chaque jour, des internautes se réunissent sur Twitch autour de différents sujets, passions et talents. Nous sommes fiers du rôle positif que joue notre service en permettant à ceux qui partagent le même centre d'intérêt de se retrouver. Les individus qui détournent le service pour y répandre des informations fausses et préjudiciables n'ont pas leur place au sein de notre communauté d'utilisateurs. Bien que la présence de ces individus ne soit pas prédominante sur Twitch, ils peuvent causer des dommages importants si nous les autorisons à utiliser notre service. C'est pourquoi Twitch a mis en place des règles interdisant notre service aux agents de désinformation.

## Lutte contre la manipulation de l'information sur le service

Définition des fausses informations (ou de toute notion voisine adoptée) retenue par l'opérateur

Nous luttons contre les agents de désinformation préjudiciable sur Twitch et en dehors du service afin de protéger notre communauté d'utilisateurs et le public, tout en préservant le dialogue ouvert entre les streamers et leur audience. Nous suspendons ainsi les utilisateurs dont la présence en ligne est dédiée au (1) partage continu (2) de sujets de désinformation préjudiciable (3) largement discrédités et partagés en masse.

Ces efforts ciblent en premier lieu les utilisateurs Twitch qui partagent des contenus de désinformation préjudiciable de façon répétée. Elle ne s'applique pas sur la base de discussions ou déclarations ponctuelles des utilisateurs. Pour déterminer si un utilisateur enfreint les règles du service, nous examinons son comportement sur la plateforme et en dehors de celle-ci.

Nos règles relatives aux agents de désinformation préjudiciable couvrent les sujets suivants (nous continuerons à mettre à jour cette liste en fonction des facteurs de risques émergents) :

- La désinformation ciblant les personnes en fonction de critères ethniques ou raciaux, sociaux, d'origine, de religion, de sexe, de genre ou d'identité de genre, d'orientation sexuelle, de handicap, d'état de santé ou encore de statut d'ancien combattant. La discrimination est par ailleurs prohibée par nos règles en vigueur contre les comportements haineux et le harcèlement.
- La désinformation préjudiciable en matière de santé, les théories complotistes liées à des traitements médicaux ou encore les fausses informations concernant la COVID-19 et le vaccin contre la COVID-19
  - La désinformation sur les traitements dont les dangers sont avérés
  - Les fausses informations concernant la COVID-19, ainsi que toute autre USPP (urgence de santé publique de portée internationale) déclarée par l'Organisation mondiale de la santé
- La désinformation soutenue par des réseaux conspirationnistes violents et/ou incitant à la violence
- La désinformation nuisant à l'intégrité d'un processus civique ou politique
  - La mise en avant de fausses allégations liées au résultat d'un processus politique contrôlé
- En cas d'urgence (ex : catastrophes naturelles, fusillades), nous prenons des mesures contre toute action de désinformation susceptible de porter atteinte à la sécurité publique.

Cette définition diffère-t-elle selon les zones géographiques de mise à disposition du service ?

NON

Interdiction des fausses informations (ou notion voisine) sur le service ?

OUI (comme expliqué ci-dessus)

Le cas échéant, où l'utilisateur peut-il trouver l'information relative à cette interdiction ?

Les règles relatives à la désinformation et à la prohibition de cette dernière sur le service sont détaillées dans les [Lignes de conduite de la communauté Twitch](#).

Remplir le tableau suivant :

Les sanctions prévues par Twitch dans le cadre de la lutte contre les agents de désinformation sur le service consistent à suspendre définitivement tous les agents de désinformation préjudiciable et leurs comptes, ainsi qu'à supprimer tous les contenus associés. Les chiffres ci-dessous indiquent donc les comptes supprimés (et non les contenus supprimés dont les comptes étaient basés en France).

Comme mentionné en introduction, les chiffres des sanctions liés à la désinformation sont relativement faibles sur Twitch pour deux raisons : 1) Le modèle de diffusion de Twitch, principalement en direct, n'est pas propice à la propagation de désinformation ou aux campagnes de désinformation à grande échelle. Il est extrêmement difficile pour un nouveau streamer ou une nouvelle streameuse d'attirer un grand nombre de spectateurs simultanés. En effet, développer une audience sur Twitch prend du temps. Qui plus est, la plupart des contenus sont éphémères, n'étant disponibles qu'en direct, et n'ont pas donc le même potentiel de « partage » et de « viralité » que les contenus disponibles de manière durable. 2) Notre approche de la lutte contre la désinformation cible les créateurs de contenus propageant de la désinformation de manière répétée. Ainsi, une déclaration isolée ou l'expression ponctuelle d'une croyance en une théorie complotiste ne constituent pas en tant que telles des motifs de bannissement sur Twitch. Cette approche est cohérente avec le format longue durée du contenu de Twitch : nous apprécions l'ensemble du contenu d'une chaîne plutôt que des déclarations spécifiques et isolées.

L'audience moyenne correspond au nombre moyen de spectateurs mondiaux sur la période pendant laquelle le créateur streamer sur Twitch ; cela inclut donc les spectateurs en dehors de France. La majorité de ces vues datent d'avant l'entrée

en vigueur de nos règles relatives aux agents de désinformation préjudiciable (mars 2022). Un compte en particulier a donc pu se constituer une audience plus importante avant d'être suspendu définitivement.

<b>Données chiffrées et indicateurs (en France, en 2022)</b>	
Nombre de contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	3 comptes
Parmi ces contenus identifiés, nombre détectés par des outils automatisés	0
Parmi les contenus identifiés, nombre détectés à la suite de signalements utilisateurs	2 comptes
Nombre de contenus comportant une fausse information (ou notion voisine adoptée) qui ont fait l'objet de mesures (recontextualisation, apposition d'un libellé, déréférencement, suppression, etc.), en déclinant par types de mesures	3 comptes suspendus définitivement
Audience moyenne (impressions ou nombre de vues) des contenus identifiés comme comportant une fausse information (ou notion voisine adoptée)	Environ 1 700 spectateurs

Parmi les contenus détectés comme contenant une fausse information (ou notion voisine adoptée), fournir la liste des 20 ayant suscité le plus d'interactions de la part des utilisateurs en France en 2022 sur le service : (i) 20 premiers contenus relevant d'une communication commerciale (publicitaires, sponsorisés et en partenariat), et (ii) 20 premiers contenus « organiques » (hors communication commerciale)

Le premier compte en termes d'audience mondiale a été suspendu lors de l'entrée en vigueur de nos règles relatives aux agents de désinformation préjudiciable pour avoir partagé de la désinformation préjudiciable en matière de santé, ainsi que de la désinformation issue de réseaux complotistes liés à des actes de violence et/ou incitant à la violence. Les deux autres comptes ont été suspendus pour avoir partagé de la désinformation préjudiciable en matière de santé.

## Respect de la liberté d'expression

En se fondant sur des cas de fausses informations considérées comme telles par ses équipes en 2022, l'opérateur est invité à exposer **les éventuelles difficultés d'articulation entre sa politique de modération et le respect de la liberté d'expression et de communication sur son ou ses services.**

Twitch s'efforce de trouver l'équilibre entre protection des utilisateurs d'une part et liberté d'expression d'autre part, en ciblant son action de lutte contre la désinformation sur les utilisateurs qui partagent de manière persistante de la désinformation préjudiciable. Une déclaration isolée ne constitue pas une infraction en tant que telle. Ainsi, nous préservons le dialogue ouvert entre les streamers et leur communauté, tout en ayant une action efficace contre la principale source de désinformation dommageable et largement démentie sur Twitch.

Existence d'un traitement différencié pour certains comptes en matière de modération (ex. : personnalités publiques)

NON

Le cas échéant : (i) critères d'identification de ces comptes, et (b) expliquer en quoi le traitement diffère.

S/O

Préciser si les efforts de lutte contre la manipulation de l'information dans le contexte de la crise sanitaire ont eu un impact sur l'exercice de leur liberté d'expression par les utilisateurs, en explicitant les raisons le cas échéant

(modération automatisée, multiplication de fausses informations sur son service, montée des attaques coordonnées, etc.)

Non, les efforts déployés pour lutter contre les fausses informations dans le contexte de la crise sanitaire n'ont pas eu d'impact sur la liberté d'expression des utilisateurs de Twitch pour les raisons exposées ci-dessus. Qui plus est, Twitch privilégie la révision humaine pour toutes les procédures de sécurité à l'échelle du site afin de garder une longueur d'avance sur les utilisateurs malveillants et de s'assurer que nos processus de confiance et de sécurité sont adaptés à notre communauté d'utilisateurs. Notre équipe d'enquête spécialisée procède à un examen approfondi de chaque cas potentiel de désinformation, y compris en examinant le contenu sur Twitch et en dehors du service.

## Transparence des mesures

- a) Moyens pour fournir aux utilisateurs des **explications claires et accessibles** sur les mesures mises en œuvre face aux risques de manipulation de l'information :
- b) Si ces informations sont **personnalisées et/ou contextuelles, fournir des illustrations.**
- c) Actions pour **davantage de transparence** envers l'ensemble des publics (utilisateurs, médias, régulateur, citoyens, etc.) notamment hors service (rapports de transparence, études internes, communiqués, billets de blog, etc.) :

Twitch détaille les règles en vigueur contre les agents de désinformation préjudiciable dans les [Lignes de conduite de la communauté](#). Lorsqu'un utilisateur enfreint ces règles (ou d'autres règles et conditions d'utilisation du service), il reçoit une notification détaillée par e-mail. Celle-ci indique la sanction prise, son motif, le caractère permanent ou temporaire d'une éventuelle suspension, des exemples de contenus en infraction, un lien vers les Lignes de conduite de la communauté et un lien vers la procédure permettant de faire appel de la décision. Exemple :



Bonjour [REDACTED]

Twitch souhaite faire en sorte que notre communauté reste accueillante pour tous. Pour cela, nous demandons à tous les utilisateurs que tout ce qu'ils partagent sur leur compte respecte les Conditions d'utilisation et les Lignes de conduite de la communauté de Twitch.

Nous avons émis un avertissement pour infraction aux Lignes de conduite de la communauté à votre compte suite à l'examen d'activité ou de contenu. En raison de la nature grave de cette infraction ou du fait que vous avez commis plusieurs infractions, votre accès aux services Twitch est restreint de manière indéfinie.

#### RAISON

### Agent de désinformation

La répétition de désinformation qui menace la sécurité de notre communauté, ce qui inclut la désinformation concernant la santé, le cadre civique, ou encore la désinformation qui fait la promotion des réseaux violents et conspirationnistes, tels que QAnon.

Voici quelques exemples de comportements délictuels, sans toutefois s'y limiter :

- faire la promotion de traitements dangereux pour la santé, tels que boire de l'eau de javel pour soigner la COVID-19.
- Partager et faire la promotion de doctrines de réseaux violents et conspirationnistes, tels que QAnon.
- Partager continuellement de la désinformation sur les vaccins, dont la nature mensongère a été démontrée, telles que : les vaccins pour la COVID-19 contiennent des puces électroniques.

Lisez la section [Utilisateurs propageant de la désinformation préjudiciable des Lignes de conduite de la communauté de Twitch](#) pour en savoir plus.

#### LIEU OÙ L'INFRACTION A EU LIEU

### Dans le profil de votre chaîne

Le contenu lié à cette infraction a peut-être été supprimé.

#### DURÉE DE LA SUSPENSION

### Indéfinie

Votre suspension ne sera pas automatiquement levée. L'accès à nos services ne peut être rétabli qu'après approbation de votre recours.

REMARQUE : comme votre suspension dure 30 jours ou plus, les abonnements que vous pouvez avoir ne seront pas renouvelés.

Si vous souhaitez faire appel de cette suspension, veuillez vous connecter à [appeals.twitch.tv](https://appeals.twitch.tv) pour soumettre un appel. Pour plus d'informations sur la procédure d'appel, veuillez consulter la page [Appels](#) de notre Centre de sécurité. Veuillez noter que le fait d'abuser de notre procédure de recours ou de tenter de contourner les restrictions imposées à votre compte (ou vos comptes) peut entraîner des sanctions supplémentaires, pouvant aller jusqu'à la suspension permanente.

Cordialement,  
l'équipe Twitch

Nous informons également nos utilisateurs via notre [Centre de sécurité, dont les contenus sont régulièrement mis à jour](#). Nous y proposons des articles et ressources pédagogiques explicitant et illustrant par des exemples concrets les règles et conditions d'utilisation du service.

En outre, Twitch publie deux fois par an un [Rapport de transparence](#) détaillant l'évolution des règles et conditions d'utilisation du service, les améliorations apportées au service et notre démarche de conformité vis-à-vis des lois et réglementations applicables. Nous publions dans ce cadre le nombre total de sanctions, signalements et requêtes auprès des autorités. Ces chiffres sont ventilés par section de nos politiques et par millier d'heures visionnées, puis comparés aux chiffres du semestre précédent. Par ailleurs, nous publions des rapports séparés pour la Global Alliance for Responsible Media (ou « GARM », une initiative intersectorielle établie par la World Federation of Advertisers dans le but de relever le défi des contenus préjudiciables sur les plateformes de médias numériques et de leur monétisation par la publicité), ainsi que dans le cadre de la loi allemande NetzDG (qui vise à lutter contre les discours haineux et la désinformation en ligne) : ces deux rapports sont publics.

## 1. Dispositif de signalement des fausses informations

1.1 Existence **d'un dispositif de signalement** permettant de signaler de fausses informations susceptibles de troubler l'ordre public ou d'altérer la sincérité du scrutin

OUI

1.2 **Libellé du ou des motifs adéquats** : **Raisons** qui sous-tendent le choix de ce ou ces libellés

La « désinformation » est l'expression la plus courante pour désigner ce type de contenu.

### Description du dispositif de signalement

1.3 Moyens permettant d'assurer la **visibilité, l'accessibilité et l'ergonomie** de ce dispositif :

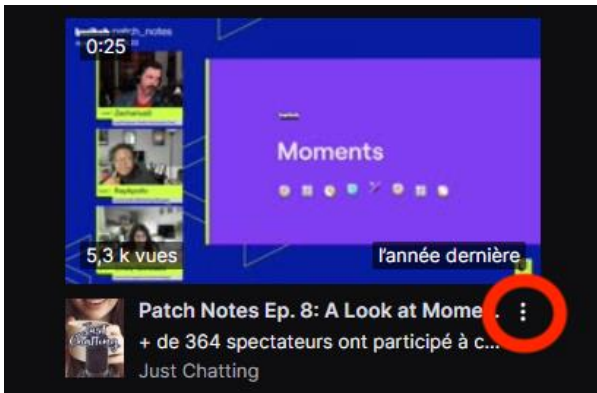
- présentation visuelle :
- **affichage** (visibilité par défaut ou non) :
- **proximité** du contenu :
- nombre d'actions nécessaires pour accéder au dispositif de signalement :
- nombre d'actions et d'hyperliens nécessaires à l'envoi du signalement d'une fausse information depuis la page du contenu :
- accessibilité aux personnes **en situation de handicap** :
- accessibilité à un **internaute non authentifié** sur la plateforme (et les éventuelles différences avec le dispositif de signalement de ceux qui le sont) : Fournir des éléments visuels pour illustrer la réponse.

Les [règles relatives aux agents de désinformation préjudiciable](#) indiquent la procédure à suivre pour envoyer un signalement, et contiennent un lien qui ouvre automatiquement un nouvel e-mail dans l'application de messagerie par défaut de l'utilisateur, afin de contacter notre équipe d'enquête spécialisée, qui examine les contenus à la fois sur Twitch et hors de Twitch.

Les utilisateurs de Twitch connectés peuvent envoyer un signalement concernant un contenu potentiellement problématique en cliquant sur l'icône à trois points qui s'affiche en bas à droite du lecteur vidéo sur la page de la chaîne, d'un clip, d'un temps fort ou d'un stream passé ; ou lorsque vous cliquez sur un identifiant. L'option « Signaler [identifiant] » apparaît et permet à l'utilisateur d'accéder à notre interface de signalement. L'utilisateur est alors invité à sélectionner la catégorie la plus pertinente pour l'infraction. Pour afficher plus d'options, dont la catégorie « désinformation », il peut cliquer sur « Rechercher ». Notre approche de la lutte contre la désinformation prenant en compte les contenus diffusés sur Twitch et hors de Twitch, ces signalements sont traités par notre équipe d'enquête spécialisée. Les utilisateurs ne peuvent donc pas signaler un compte via l'outil de signalement standard. Après avoir sélectionné « Désinformation », ils accèdent à un lien vers la page de nos règles relatives à la désinformation, où est indiquée la procédure à suivre pour envoyer un signalement. Voir les captures d'écran du processus de signalement ci-dessous.

*Début du processus de signalement depuis une page de chaîne : icône à trois points en bas à droite du lecteur vidéo qui affiche l'option « Signaler un live »*





### Catégories d'infraction principales

Pour quelle(s) raison(s) soumettez-vous ce rapport ?

- Violence ou gore
- Intimidation ou harcèlement
- Comportement haineux
- Automutilation
- Nudité ou contenu à caractère sexuel explicite
- Spam, escroquerie ou bots
- Terrorisme
- Rechercher 🔍

Retour Suivant

### Catégories supplémentaires sous « Rechercher », dont « Désinformations »



Signaler : Identifiant

Rechercher une raison

- Comportement sexuel impliquant une personne mineure
- Contenu à caractère sexuel explicite
- Cyberattaque
- Désinformation
- Escroquerie
- Harcèlement organisé
- Incitation au harcèlement
- Incitation à adopter un comportement haineux

Retour Suivant

Pop-up invitant l'utilisateur à consulter nos règles relatives à la désinformation, et contenant un lien pour envoyer un e-mail à l'équipe d'enquête

Signaler : Identifiant

Merci pour ce signalement.

La désinformation constante enfreint les conditions d'utilisation de Twitch. Pour en savoir plus sur notre politique et envoyer un rapport pour ce type d'infraction, rendez-vous sur cette page :

[Consulter notre politique concernant la désinformation](#)

Fermer

Depuis la page des règles contre la désinformation, il suffit de deux clics pour envoyer un signalement : un clic sur le lien qui ouvre un nouvel e-mail, et un second pour appuyer sur « Envoyer ». Pour accéder à cette page depuis le contenu problématique (et afficher l'adresse e-mail de notre équipe d'enquête spécialisée), il faut sept clics supplémentaires.

Les utilisateurs déconnectés peuvent envoyer un signalement à l'adresse e-mail de l'équipe d'enquête spécialisée.

Pour garantir à tous une accessibilité optimale au contenu de Twitch, (a) nous formons nos concepteurs, développeurs et gestionnaires de contenu aux enjeux d'accessibilité ; (b) nous intégrons une documentation dédiée dans nos ressources de base ; et (c) nous effectuons régulièrement des contrôles automatisés, semi-automatisés et manuels de notre contenu. Pour en savoir plus, consultez la [Politique d'accessibilité de Twitch](#).

1.4 Le dispositif décrit est-il **commun** : à toutes les **déclinaisons** du service (site internet, application) ? OUI / NON (préciser)

OUI

1.4 Le dispositif décrit est-il **commun** : à tous les types de contenu (texte, vidéo, image, commentaires, contenus sponsorisés, compte d'utilisateur, contenus en direct, contenus éphémères) accessibles sur le service ? OUI / NON (préciser) Fournir des éléments visuels pour illustrer la réponse (en cas de différences).

OUI

1.5 Les utilisateurs **sont-ils consultés** sur l'accessibilité, la visibilité et l'ergonomie du dispositif de signalement ? OUI / NON  
Le cas échéant, comment et quels sont leurs retours ?

NON

Les utilisateurs ne sont pas directement consultés. Toutefois, nous testons régulièrement de nouvelles fonctionnalités de l'interface utilisateur et évaluons l'impact sur le comportement des utilisateurs afin de continuer à améliorer le processus de signalement, en rendant celui-ci plus intuitif et plus efficace.

## La procédure de traitement des signalements

1.6 Décrire la **procédure de traitement du signalement d'une fausse information** en fournissant toute illustration permettant de favoriser la compréhension de la procédure.

- **Conditions** de mise en instruction d'un signalement :
- **Étapes** de l'instruction :
- **Issues possibles** de l'instruction (type d'actions prises) :
- **Information apportée aux utilisateurs** (lesquels sont informés ? quels stades de la procédure donnent lieu à une information ? sous quelle forme ?) :
- **Procédure spécifique** en période électorale ou en cas de crise : NON / OUI (préciser)
- **Recours à des signaleurs de confiance (trusted flaggers)** : NON / OUI (préciser lesquels et à quelles fins)

Tous les signalements de désinformation sont reçus et examinés par une équipe d'enquête interne spécialisée au sein de notre équipe de sécurité. L'équipe d'enquête procède à un examen approfondi du contenu de l'utilisateur signalé, à la fois sur Twitch et hors de Twitch, et l'évalue selon les critères énoncés dans nos règles relatives aux agents de désinformation préjudiciable. Si nécessaire, l'équipe fait appel à l'expertise complémentaire du Global Disinformation Index (GDI).

Si l'équipe d'enquête détermine qu'un utilisateur correspond à la définition d'agent de désinformation préjudiciable, le compte de cet utilisateur est suspendu définitivement, et tous les contenus associés sont supprimés. Tout compte connexe appartenant au même utilisateur est également suspendu définitivement. L'utilisateur reçoit alors un e-mail l'informant de l'infraction, et indiquant l'endroit où elle s'est produite, la durée et le motif de la suspension, des exemples de contenus concernés et un lien vers les Lignes de conduite de la communauté pour en savoir plus sur la politique. (Pour voir un exemple d'e-mail, reportez-vous à la section « Transparence des mesures ».)

Cette procédure d'examen des signalements individuels est la même en période électorale ou en cas de crise. Dans ces circonstances particulières, Twitch met en place une équipe transverse pour renforcer le dispositif de contrôle et adapter les processus le cas échéant.

Nous n'avons pas recours à des signaleurs de confiance à l'heure actuelle.

1.7 **Voies de recours** internes à l'issue d'une décision :

- **décisions de la plateforme** susceptibles d'être contestées :
- critères de recevabilité des appels :
- **présentation à l'utilisateur** de ses voies de recours possibles contre une décision :
- **équipes** en charge de l'instruction des recours :
- **modalités d'instruction** de l'appel (notamment : possibilité pour l'utilisateur de faire valoir ses explications lors de l'instruction) :
- **exceptions** à cette procédure : NON / OUI (préciser).

Dans la grande majorité des cas de suspension, les utilisateurs peuvent faire appel. Un lien vers le portail de la procédure d'appel est inclus dans l'e-mail envoyé à tous les utilisateurs lors de leur première suspension. Les utilisateurs peuvent également accéder à cette procédure en se connectant à leur compte. Chaque cas de recours en appel est examiné par l'équipe d'enquête interne de Twitch, systématiquement par une personne différente de celle qui a pris la décision initiale.

Les sanctions sont prises au vu des critères stricts selon lesquels nous déterminons qu'un utilisateur est un agent de désinformation préjudiciable, combinés à une enquête approfondie sur le comportement de l'utilisateur sur Twitch et en dehors de Twitch – c'est pourquoi les procédures d'appel se concluent en grande majorité par la confirmation de la sanction initiale.

## Les moyens alloués au traitement des signalements

### 1.8 Moyens humains déployés, et notamment :

- **statut** des modérateurs (collaborateurs ou prestataires de la société opératrice) :
- **formation** (initiale et continue) :
- **localisation** géographique :
- **langue(s)** de travail (notamment, langues régionales françaises prises en charge) :
- modération **à tout moment** (24h/24 et 7j/7) : OUI / NON (préciser)
- **sources** dont disposent les modérateurs (juridiques, journalistiques, académiques) :
- mesures pour **réduire les risques professionnels** pour les modérateurs :

Notre équipe en charge des opérations de sécurité travaille 24 heures sur 24, 7 jours sur 7 et 365 jours par an pour répondre rapidement aux signalements d'utilisateurs et prévenir tout préjudice supplémentaire. L'équipe s'appuie sur des professionnels dans le monde entier maîtrisant 25 langues au total, dont le français. L'équipe spécialisée qui enquête sur les potentiels agents de désinformation préjudiciable est constituée exclusivement de collaborateurs de l'entreprise employés à temps plein. Cette équipe procède à un examen approfondi du contenu des utilisateurs signalés, à la fois sur Twitch et en dehors du service, et l'évalue selon les critères énoncés dans nos règles relatives aux agents de désinformation préjudiciable. Si nécessaire, l'équipe fait appel à l'expertise du Global Disinformation Index (GDI).

Tous les professionnels de la modération de contenu chez Twitch suivent une formation rigoureuse et approfondie avant de rejoindre l'équipe, et bénéficient par la suite d'une formation continue. Nous sommes soucieux de préserver le bien-être de ces professionnels compte tenu de leur exposition prolongée à des contenus potentiellement négatifs et perturbants. Nous accordons autant d'importance à leur santé et à leur sécurité qu'à celles de la communauté Twitch, et prenons donc des mesures afin de les protéger. Dès le début du processus de recrutement, les candidats sont soumis à des tests psychométriques afin de s'assurer de leur capacité de résilience psychologique. Par la suite, nous prenons des mesures pour optimiser l'ergonomie des postes de travail et structurer les temps de pause dans l'emploi du temps des analystes. Nous avons également investi dans des outils visant à limiter l'impact des contenus préjudiciables pour les équipes qui y sont exposées, notamment grâce à un système automatique d'affichage des vidéos en noir et blanc et/ou sans son. Nos spécialistes de modération des contenus disposent d'un accès prioritaire et direct à des conseillers dédiés, avec un ratio minimum de conseiller par professionnel dans tous nos centres de modération. Les spécialistes peuvent aussi avoir recours à des coachs par visioconférence et bénéficient de sessions de formation continue en groupe. Nous réalisons enfin des entretiens de départ lorsque nos professionnels de la modération de contenu quittent l'entreprise, ce qui nous permet de tirer des enseignements et d'apporter des améliorations à nos processus de manière continue.

### 1.9 Existence de moyens humains et/ou automatisés de **détection proactive et dédiée** aux contenus diffusant de fausses informations : OUI/NON

NON

### 1.9 Processus d'instruction des contenus détectés identique à celui des contenus signalés par les utilisateurs : OUI / NON (préciser)

NON. Toutefois, si nous suspendons le compte d'un utilisateur pour non-respect des règles relatives aux agents de désinformation préjudiciable, nous suspendons également les autres comptes associés à la même adresse e-mail ou au même numéro de téléphone.

1.10 **Commenter la préconisation de l'ARCOM aux opérateurs**, dans son précédent bilan sur la lutte contre la manipulation de l'information, de maintenir une intervention humaine dans le processus de décision d'une action à l'égard d'un contenu ou d'un compte.

Twitch privilégie la révision humaine pour toutes ses procédures de sécurité afin de garder une longueur d'avance sur les utilisateurs malveillants et de s'assurer que les processus de confiance et de sécurité restent adaptés à la communauté d'utilisateurs du service. En ce qui concerne la désinformation, en particulier, notre équipe d'enquête spécialisée procède à un examen approfondi de chaque cas suspect.

### Questions complémentaires

1.11 Remplir le tableau suivant

Comme mentionné précédemment, les sanctions prévues par Twitch contre les agents de désinformation préjudiciable visent à suspendre définitivement tous les agents de désinformation et leurs comptes, ainsi qu'à supprimer tous les contenus associés. Les chiffres ci-dessous indiquent donc les comptes concernés (et non les contenus concernés dont les comptes étaient basés en France). En outre, le temps de traitement de ces signalements (soit le temps total écoulé entre la réception du signalement et la notification à l'utilisateur) est plus long, car l'examen des contenus sur Twitch et hors de Twitch est plus approfondi.

Données chiffrées et indicateurs (en France, en 2022)	
Nombre de signalements pour fausses informations (ou notion voisine adoptée) par les utilisateurs	3
Nombre de procédures de traitement mises en œuvre	3
Délai moyen de traitement d'un signalement	Environ 7 jours
Part des signalements traités : (i) en moins de 24h, (ii) en moins d'une semaine	(i) 0 (ii) 100 %
Nombre de contenus considérés comme étant de fausses informations (ou notion voisine adoptée) à l'issue de la procédure de traitement	3 comptes
Nombre d'appels reçus par l'opérateur à l'issue de la procédure de traitement	3 appels
Délai moyen du traitement d'un appel	Moins de 24 heures
Part des appels traités : (i) en moins de 24h, (ii) en moins d'une semaine	(i) 100 % (ii) 0
Budget alloué au traitement des signalements et des appels des utilisateurs (€)	Non disponible
Nombre de personnes en charge du traitement des signalements susceptibles d'agir sur des contenus du service en France	Non disponible

1.12 **Impact de la crise sanitaire sur le dispositif de signalement** de fausses informations (ou notion voisine adoptée) et mesures spécifiques prises dans ce contexte en 2022 :

Les règles de Twitch relatives aux agents de désinformation préjudiciable sont entrées en vigueur en mars 2022 dans le but de réduire les risques pour notre communauté, notamment ceux liés à la désinformation en matière de santé. Hormis ces nouvelles règles, la crise sanitaire n'a pas entraîné de modifications spécifiques de notre système de signalement, ni de notre processus de traitement des signalements.

### 1.13 Actions futures et nouvelles fonctionnalités envisagées à l'avenir :

Aucune pour le moment

## 2. Transparence des algorithmes

2.1 Commenter les définitions de la **loyauté**, de l'**explicabilité** et de l'**équité** indiquées en annexe du questionnaire : L'opérateur indiquera s'il s'y réfère (ou à des notions équivalentes) notamment dans le cadre des mesures qu'il met en œuvre afin de lutter contre la manipulation de l'information, et les éventuelles limites identifiées à leur application. Il est invité à formuler d'éventuelles propositions afin d'affiner ces définitions.

Nous utilisons des définitions de « loyauté », « explicabilité » et « équité » similaires à celles qui figurent dans l'annexe du document. Nous n'avons pas identifié de limites à leur application et ne suggérons pas de modifier les définitions.

### Systemes de modération de contenus

2.2 **Décrire chaque système de modération de contenus** utilisés sur le service pour lutter contre la manipulation de l'information en détaillant :

- ses **fonctionnalités** ;
- son **objectif principal** (tâche ou classe de problèmes à résoudre) ;
- le fait qu'il repose ou non sur un ou plusieurs algorithmes d'**apprentissage automatique** et, le cas échéant, si ceux-ci sont utilisés avec d'autres types d'algorithmes (ex. algorithmes déterministes) ;
- lorsqu'il repose sur un ou plusieurs algorithmes d'apprentissage automatique, la nature de(s) **jeu(x) de données d'apprentissage** utilisé(s) ;
- pour la reconnaissance de texte, les **langues utilisées** (notamment régionales) ;
- la liste la plus exhaustive possible des **données prises en entrée** de ce système ;
- la liste la plus exhaustive possible des **données de sortie** de ce système ;
- les conditions mises en œuvre par l'opérateur qui permettent à ses services et leurs algorithmes de respecter les **principes de loyauté et d'équité**, et d'être explicables ;
- le cas échéant, la manière dont ce système est **personnalisé** en fonction de l'utilisateur ;
- le cas échéant, les modalités de recours à une **intervention humaine** sur le système à partir de sa mise en œuvre, et si recours est le fait de personnes en interne, de personnes employées par des prestataires, ou bien des deux.

Les opérateurs pourront recourir à des **représentations graphiques** afin de répondre aux questions<sup>1</sup>, faisant figurer clairement l'ensemble des éléments demandés.

Notre système de modération de contenu contre la désinformation repose à la fois sur le signalement par les utilisateurs et sur le partage d'informations avec des experts, des gouvernements et d'autres services. Twitch travaille en partenariat étroit avec le Global Disinformation Index (GDI) pour mettre à jour en permanence notre approche de la désinformation et identifier les potentiels agents de désinformation sur notre service. En tant que signataire du [Code de bonnes pratiques de l'UE en matière de désinformation](#) et membre de la [Global Alliance for Responsible Media \(GARM\)](#), Twitch collabore également avec des agences gouvernementales, d'autres services de réseaux sociaux et acteurs du secteur de la publicité afin de partager l'information et d'identifier les agents de désinformation. Nos équipes de Confiance et sécurité et de Communication surveillent également les réseaux sociaux pour repérer d'éventuels agents de désinformation annonçant un passage à Twitch depuis un autre service.

2.3 Décrire la **politique de lutte contre les biais** de l'opérateur pouvant exister dans ses algorithmes d'apprentissage automatique : Indiquer **si des biais ont pu être observés dans la modération** des contenus sur le service. Le cas échéant, détailler les procédures qui ont permis l'identification de ces biais et les modifications apportées pour les supprimer.

<sup>1</sup> La [norme ISO 5807](#) pourra, par exemple, être utilisée pour établir des logigrammes.

Aucun algorithme d'apprentissage automatique n'est utilisé pour lutter contre la désinformation : notre système de modération de contenu contre la désinformation repose sur les signalements des utilisateurs ainsi que sur l'échange d'informations avec des experts.

#### 2.4 Décrire les **modalités d'évaluation** des algorithmes de modération utilisés sur le service.

Comme indiqué ci-dessus, Twitch n'utilise pas d'algorithmes pour la modération de la désinformation. Pour la détection des autres types de contenu, nous évaluons les algorithmes utilisés en examinant le taux de faux négatifs des modèles, mesuré par les suspensions et l'échantillonnage aléatoire. Pour les faux positifs, nous examinons le nombre de demandes envoyées à l'équipe en charge de l'Expérience client et liées à un algorithme de modération donné.

Pour identifier les comportements problématiques dans les textes, nous procédons à un échantillonnage continu afin d'évaluer les performances de nos classificateurs actuels dans chacun des vingt contextes comportementaux (racisme, misogynie, etc.) et dans les dix-neuf langues dans lesquelles le service est disponible. Les faux positifs sont détectés par comparaison avec un ensemble de tests de référence fréquemment mis à jour. Pour les faux négatifs, un système observe les changements de langage dans les contextes comportementaux que nous couvrons, et les soumet à un examen humain pour une éventuelle mise à jour du modèle.

Enfin, nous surveillons le comportement des modérateurs humains via nos systèmes d'auto-modération afin de déterminer si d'éventuels faux positifs ou négatifs sont à l'origine de performances anormales. Par exemple, les modérateurs humains peuvent choisir d'outrepasser le système d'auto-modération ; ces décisions sont alors prises en compte dans un signal agrégé indiquant qu'un classificateur est susceptible de nécessiter une révision.

#### 2.5 Si des **modifications des systèmes de modération** sont survenues en 2022, en préciser les raisons, détailler les modifications apportées et indiquer si elles ont eu un caractère temporaire :

En 2022, nous avons ajouté le turc à la classification comportementale linguistique complète de nos systèmes d'auto-modération, et avons augmenté les capacités de modération en italien. Nous avons également ajouté un niveau de prise en charge de base pour la classification comportementale linguistique du chat en arabe, tchèque, finnois et vietnamien. Qui plus est, nous avons sensiblement amélioré notre capacité à détecter et à distinguer les langues, en particulier dans les contextes où plusieurs langues sont présentes ou lorsqu'un utilisateur tente d'utiliser une langue ou des caractères dans le but de dissimuler un message toxique.

En ce qui concerne la toxicité masquée, nous avons amélioré notre capacité à détecter l'utilisation de nouveaux types de remplacements, d'insertions, de suppressions et de duplications de caractères pour dissimuler le sens voulu d'un mot ou d'une expression.

Nous ajoutons en permanence de nouveaux contextes comportementaux à nos modèles linguistiques en fonction des tendances que nos data scientists observent dans nos données de chat. Par exemple, en 2022, nous avons apporté plusieurs modifications à nos classificateurs de haine identitaire (qui couvrent le racisme, la misogynie, la haine à l'égard des LGBTQ+, etc.) pour tenir compte de l'évolution linguistique. Par ailleurs, nous ajoutons continuellement du contexte linguistique à nos modèles afin de déterminer la meilleure interprétation comportementale d'un mot ou d'un message dans un contexte textuel particulier.

#### 2.6 Indiquer si **des outils automatisés de modération** pour lutter contre **des pratiques de manipulation de l'information en particulier** (ex : outils de lutte contre les *deepfakes*, contre la détection de bots, etc.) ont été développés.

Pour détecter les bots (c'est-à-dire les comptes qui ont été créés et agissent de manière automatisée), Twitch procède ainsi :

- Recherche d'un comportement de type « bot » : nous utilisons Kasada (un outil tiers de défense contre les attaques de bots avancées et persistantes) pour vérifier si une demande donnée provient d'un bot. En complément, un modèle distinct examine les actions passées d'un utilisateur sur le site et détermine si ces actions sont susceptibles de provenir d'un bot.

- Caractéristiques de l'inscription : nous utilisons un modèle pour déterminer si une inscription donnée présente des caractéristiques similaires à celles d'utilisateurs qui ont déjà été suspendus pour cause d'activité de bot.
- Contenu de type « bot » : nous utilisons également des modèles pour détecter le contenu couramment envoyé par les bots, notamment les spams dans les chuchotements et les chats. Si des liens de désinformation étaient massivement partagés, ces modèles apprendraient à les intercepter.
- Détection des bots spectateurs : nous disposons de nombreux systèmes qui collaborent pour filtrer algorithmiquement les vues que les modèles considèrent comme frauduleuses.

## Systemes de recommandation de contenus

2.7 **Décrire** les systèmes de recommandation **de contenus** utilisés sur le service, en détaillant les mêmes points que ceux énoncés à la question 2.2.

Les recommandations du menu de gauche sur Twitch utilisent un système de classement en deux étapes. La première étape consiste à recommander des chaînes ou des catégories aux utilisateurs en fonction de ce que ces utilisateurs, ou d'autres comme eux, ont regardé par le passé (sur la base du nombre de minutes visionnées). La deuxième étape utilise un algorithme « d'apprentissage du classement » qui observe les comportements des utilisateurs avec les chaînes recommandées, et notamment les chaînes avec lesquelles ils ont interagi positivement (en cliquant dessus ou en les regardant). Le classement des recommandations est principalement influencé par la première étape. Si une chaîne n'est pas sélectionnée lors de la première étape, elle ne sera pas recommandée lors de la deuxième.

La liste la plus exhaustive possible des **données de sortie** de ce système : Le classement et le score (généré par le modèle) des 15 premières chaînes pour chaque utilisateur.

La liste la plus exhaustive possible des **données prises en entrée** de ce système\* :

- **Durée depuis la dernière apparition** de la chaîne dans le menu de gauche (en nombre de jours)
- **Durée de la présence** de la chaîne dans le menu de gauche (en nombre de jours)
- Nombre de minutes visionnées par l'utilisateur sur chaque chaîne au cours des 42 derniers jours
- Identifiant de la chaîne
- Nombre de minutes visionnées dans les 2 dernières heures par catégorie
- Surface Twitch concernée pour l'utilisateur
- Plateforme Twitch concernée pour l'utilisateur

\* AUCUNE information textuelle ou linguistique n'est utilisée actuellement.

2.8 Si des **modifications de ces systèmes de recommandation** sont survenues en 2022, en préciser les raisons, détailler les modifications apportées et indiquer si elles ont eu un caractère temporaire.

En 2022, les changements suivants ont été apportés aux recommandations :

1. [Mai 2022] Mise à jour du modèle de l'étape 2 pour le menu de gauche Twitch : passage de XGBoost à un modèle de classement par réseau de neurones profonds.
2. [Septembre 2022] Lancement de la bibliothèque « Tendances mondiales » sur la page d'accueil de Twitch dans 12 pays. Cette ressource regroupe les chaînes les plus populaires dans chaque région du monde.
3. [Octobre 2022] Lancement de la deuxième étape du système de classement pour toutes les recommandations Twitch (page d'accueil, page « Parcourir », menu de gauche, etc.) sur iOS, Android, TV...
4. [Novembre 2022] Déploiement du classement basé sur les recommandations dans la section des chaînes suivies du menu de gauche.
5. [Décembre 2022] Ajout d'une alerte au niveau de la surface et de la plateforme de l'utilisateur à la deuxième étape du système de classement.

2.9 **Actions mises en place pour lutter, via les systèmes de recommandation de contenus sur le service, contre la manipulation de l'information** (décrire chaque type d'actions et préciser dans quelles conditions elles sont appliquées (critères de déclenchement))



Lorsque des créateurs de contenu sont identifiés et suspendus en application de nos règles contre la désinformation préjudiciable, leurs comptes ne sont plus visibles et ne sont donc pas inclus dans les recommandations.

## Information des utilisateurs

2.10 Manière dont les **systèmes de recommandation et de modération** utilisés dans le cadre de la lutte contre la manipulation de l'information sont **décrits à l'utilisateur**.

Nous publions des informations concernant nos systèmes de recommandation et de modération dans la section d'aide de notre site web. Pour consulter des exemples, cliquez [ici](#) et [ici](#).

2.11 Pour chacun de ces systèmes, préciser **de quelle manière les utilisateurs** : - sont **informés des mises à jour** de ces systèmes :

- peuvent demander des **informations supplémentaires** sur ces systèmes
- bénéficient d'une **explication des résultats des systèmes** qui les affectent

Comme indiqué au point 2.10 ci-dessus, nous publions des informations concernant nos systèmes de recommandation et de modération dans la section d'aide de notre site web. Les utilisateurs sont informés des principales modifications apportées aux fonctionnalités de nos produits sur notre site web, sur Twitter et par e-mail. La section d'aide détaille également comment chaque changement affecte les utilisateurs finaux.

2.12 **Modifications qu'un utilisateur peut apporter** aux **paramètres** relatifs aux algorithmes de recommandation sur le service.

- **liste** des systèmes de recommandation paramétrables par l'utilisateur :
- comment ces réglages (2.11) sont rendus facilement **accessibles et compréhensibles** :

Les utilisateurs peuvent indiquer les chaînes et les catégories qu'ils souhaitent retirer de leurs recommandations, ce qui affecte le contenu recommandé sur la page d'accueil, sur la page « Parcourir » (lorsque vous triez par « Recommandé pour vous ») et dans le menu de gauche. Les utilisateurs peuvent demander à retirer des contenus de leurs recommandations via l'icône à trois points associée à chaque élément. Ils peuvent ensuite voir la liste complète de ces contenus dans leurs paramètres, et peuvent supprimer les éléments qu'ils ont ajoutés par erreur ou qu'ils veulent réintégrer à leurs recommandations. Twitch a également publié un [article d'aide](#) décrivant comment utiliser ces paramètres et comment ceux-ci affectent les recommandations.

2.13 **Commenter la préconisation de l'ARCOM aux opérateurs**, dans son précédent bilan sur la lutte contre la manipulation de l'information, **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**.

Twitch est en conformité avec les recommandations de l'ARCOM pour les opérateurs. Nous publions des informations concernant nos systèmes de recommandation et de modération dans la section d'aide de notre site web. Pour consulter des exemples, cliquez [ici](#) et [ici](#). Nous estimons que cela répond à l'objectif de transparence fixé par l'ARCOM à l'égard des utilisateurs.

## Questions complémentaires

2.14 Remplir le tableau suivant

Les deux premières questions ne s'appliquent pas, car notre moteur de recommandation n'utilise pas de sources d'information externes, et s'appuie sur les statistiques des visites du site, ainsi que sur les comportements et préférences des utilisateurs.

**Données chiffrées et indicateurs (2022)**

Parts des accès à de fausses informations (ou notion voisine adoptée) imputables : (i) à l'algorithme du moteur de recherche interne du service (ii) aux autres algorithmes de recommandation, classement ou référencement du service.	S/O
Nombre de contenus de fausse information (ou notion voisine adoptée) ayant fait l'objet d'une recommandation.	S/O
Part des utilisateurs du service ayant recours à la personnalisation des paramètres relatifs aux algorithmes de recommandation des contenus (organiques et commerciaux) sur le service	En 2022, plus de 6 millions d'utilisateurs dans le monde ont donné leur avis sur au moins une recommandation

### 2.15 Impact de la crise sanitaire sur le fonctionnement des systèmes de recommandation ou de modération pour lutter contre la manipulation de l'information, et mesures spécifiques prises dans ce contexte en 2022 :

Nos systèmes de recommandation et de modération sont basés sur le machine learning (apprentissage automatique) avec une supervision humaine. La crise sanitaire n'a pas eu d'impact sur les systèmes de machine learning. Elle a eu un impact humain, quoique minime, en termes de coordination des équipes à distance pendant les périodes de confinement. Les systèmes de recommandation et de modération de Twitch se sont maintenant bien adaptés à une combinaison d'équipes sur site, en télétravail et hybrides. Nous n'avons pas rencontré de difficultés supplémentaires en 2022 dues à la crise sanitaire dans ces deux domaines, et nous n'en prévoyons pas d'autres majeures en 2023.

### 2.16 Actions futures et nouvelles fonctionnalités envisagées à l'avenir.

En 2023, nous continuerons d'améliorer nos algorithmes de machine learning pour les recommandations et la modération afin d'optimiser divers aspects liés aux performances, à l'équité, à la loyauté et à l'explicabilité. Nous ne pouvons pas encore fournir de liste exhaustive, car la feuille de route est en cours de développement. Qui plus est, dévoiler toutes les nouvelles fonctionnalités risquerait de compromettre nos secrets commerciaux.

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

### Identification des sources des contenus et labellisation

#### 3.1 Identification des sources des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle : OUI / NON

NON

#### 3.6 Décrire les éventuels freins à la mise en œuvre des mesures d'identification et de labellisation :

Après avoir consulté des chercheurs et des experts lors de l'élaboration de nos conditions d'utilisation du service et de nos processus en matière de désinformation, il a été établi que la labellisation n'avait pas de sens pour un service de streaming en direct comme Twitch, étant donné le temps de latence qu'implique une labellisation précise.

### Dispositifs de vérification des faits (« fact-checking »)

#### 3.7 Recours à un dispositif de fact-checking (confié à des tiers ou interne) : OUI / NON

NON

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

Twitch concentre son dispositif anti-désinformation sur les utilisateurs qui partagent de manière persistante de la désinformation préjudiciable. Compte tenu du mode de diffusion sur Twitch, nous apprécions l'ensemble du contenu des streamers plutôt que des propos isolés. Notre approche se situant au niveau de l'utilisateur, nous ne procédons pas à un **fact-checking** pour chaque contenu. Par ailleurs, d'après nos échanges avec des experts, le **fact-checking** n'est pas recommandé pour un service de streaming en direct comme Twitch, en raison du temps de latence que cela impliquerait.

## Questions complémentaires

3.13 Impact de la **crise sanitaire sur la promotion des contenus issus d'entreprises et d'agences de presse et de services de communication audiovisuelle** et mesures spécifiques prises dans ce contexte en 2022 :

Aucun

3.14 **Actions futures et nouvelles fonctionnalités envisagées à l'avenir.**

Aucune

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

### Définitions et pratiques de diffusion massive de fausses informations

4.1 **Appréhension par l'opérateur** de la notion de comptes propageant massivement de fausses informations, telle que mentionnée à l'article 11 de la loi du 22 décembre 2018.

Notre service ne permet pas la propagation persistante de fausses informations sur Twitch ou hors de Twitch par des agents de désinformation. Sont considérés comme des « agents de désinformation préjudiciable » : les individus dont la présence en ligne est dédiée au (1) partage continu (2) de sujets de désinformation préjudiciable (3) largement discrédités et partagés en masse.

4.2 Décliner les **différentes catégories de pratiques** de diffusion massive de fausses informations identifiées sur le service, en en donnant une définition et en les caractérisant, au regard notamment de leur **origine** (types d'agent), de leurs moyens de diffusion (exclusivement algorithmiques, exclusivement humains ou hybrides) et de la **nature de l'objectif poursuivi par leurs émetteurs** (exclusivement économique et/ou politique, ou autre)

Tous les agents de désinformation préjudiciable identifiés et suspendus de notre service ont partagé de fausses informations dans le cadre de leurs streams Twitch. Leur moyen de diffusion était donc exclusivement humain. La quasi-totalité des comptes suspendus en vertu de nos conditions d'utilisation ne disposaient pas d'un nombre important de followers ou d'abonnés sur Twitch. Cela ne signifie pas que leur audience était similaire sur les autres services, mais leur présence sur Twitch était faible. Par conséquent, il est extrêmement difficile de déterminer les objectifs de ces individus.

4.3 Pratiques détectées sur le service consistant à **générer un trafic ou un engagement artificiel massif** pour propager fortement un contenu en augmentant sa viralité (achat/vente de faux abonnés, fraude aux clics...) :  
**Mesures prises à leur encontre :**

À notre connaissance, aucun agent de désinformation préjudiciable n'a réussi à générer un trafic massif ou un engagement artificiel. Comme indiqué ci-dessus, la quasi-totalité des comptes suspendus en vertu des règles contre la désinformation préjudiciable de Twitch ne disposaient pas d'un nombre important de followers ou d'abonnés.

En parallèle de nos règles relatives aux agents de désinformation préjudiciable, Twitch interdit les bots malveillants, le spam, les escroqueries, l'usurpation d'identité et le contournement de suspension, ainsi que tout contenu ou toute activité qui perturbe, entrave ou compromet de quelque manière que ce soit l'intégrité des services Twitch. Cela inclut tout effort visant à augmenter artificiellement le nombre de followers. Pour identifier ces comportements, nous disposons de modèles qui détectent les activités de type « bot » sur la plateforme (voir la réponse à la question 4.8), ainsi que de modèles permettant de vérifier si une tentative de « follow » donnée est susceptible de provenir d'un bot. Chaque semaine, notre équipe Opérations de sécurité suspend les comptes soupçonnés d'être des bots followers, en se basant sur les résultats de ces modèles.

4.4 Fournir la liste ou des exemples :

- **d'opérations de propagation de fausses informations soutenues par des entités étatiques ou institutionnelles** identifiées et mesures prises à leur égard :
- **de coopération** avec les autorités et/ou d'autres opérateurs dans le but de mieux détecter et de prévenir ces phénomènes :
- **de pratiques individuelles ou à l'échelle d'un groupe d'utilisateurs** de comportements trompeurs et abusifs massifs (ex. : multidiffusion, publication à répétition) identifiées et les moyens mis en œuvre pour lutter contre :

Nous n'avons pas identifié sur Twitch d'opérations de désinformation soutenues par des entités étatiques ou institutionnelles. De même, nous n'avons pas constaté de coordination entre les agents de désinformation sur le site. Ceux-ci n'ont pas non plus démontré de pratique particulière visant à élargir la portée de leur contenu, leur comportement sur Twitch consistant plutôt à partager de la désinformation dans le cadre de leurs streams en direct.

Nous continuerons à collaborer avec des experts, des agences gouvernementales et d'autres services afin de surveiller les tendances et de détecter les risques affectant notre secteur.

## Moyens de lutte contre la diffusion massive de fausses informations et informations aux utilisateurs

4.5 Pour chacune des catégories identifiées en réponse à la question 4.2, indiquer : les **moyens** (humains et technologiques) pour les **détecter** et y **faire obstacle**, et les **mesures** prises et critères déclenchant l'application de chacune d'entre elles

Voir les réponses aux questions 1.6 et 2.2 ci-dessus.

4.6 **Information des utilisateurs** sur les pratiques liées à l'utilisation de comptes propageant massivement de fausses informations susceptibles d'entraîner une intervention et les mesures susceptibles d'être prises : NON / OUI (préciser)

OUI. Les détails des règles applicables aux agents de désinformation préjudiciable sont disponibles dans les [Lignes de conduite de la communauté](#).

4.7 **Commenter la préconisation de l'ARCOM aux opérateurs**, dans son précédent bilan sur la lutte contre la manipulation de l'information, **d'informer davantage les utilisateurs sur les pratiques coordonnées d'influence et les risques qui en découlent**.

Comme indiqué ci-dessus, nous n'avons constaté aucune pratique d'influence coordonnée sur Twitch. Nous n'avons donc pas jugé prioritaire de partager ces informations avec les utilisateurs.

4.8 Décrire les mesures et moyens mis en œuvre pour **la détection de bots et les outils développés** pour lutter contre ceux utilisés à des fins de manipulation de l'information. L'opérateur précisera notamment :

- s'il possède une **définition** du terme « bot » et quels en sont les usages qu'il autorise et proscrit sur son service :
  - les estimations **du trafic généré** par ces derniers sur son service :
- s'il permet à des **applications tierces** d'utiliser certaines fonctionnalités des comptes des utilisateurs (ex. : publication et partage de contenus, demande de connexion entre utilisateurs) et comment il s'assure que ces pratiques ne soient pas utilisées à des fins de manipulation de l'information.

Un bot est un compte qui, selon Twitch, a été créé et agit de manière automatisée. Les bots légitimes sont largement utilisés sur Twitch, par exemple pour partager les profils du streamer sur d'autres réseaux sociaux. Au quatrième trimestre 2022, les [bots malveillants](#) représentaient 0,009 % des messages de chat, 8,9 % des nouveaux followers et 1,4 % des chuchotements (messages privés). Les bots sont principalement utilisés pour attirer du trafic hors Twitch ou mener des attaques de type DDOS dans le chat ou sur les nouveaux followers.

Pour détecter les bots, nous procédons ainsi :

- Recherche d'un comportement de type « bot » : nous utilisons Kasada pour vérifier si une demande donnée provient d'un bot. En complément, un modèle distinct examine les actions passées de l'utilisateur sur le site et détermine si ces actions sont susceptibles de provenir d'un bot.
- Caractéristiques de l'inscription : nous utilisons un modèle pour déterminer si une inscription donnée présente des caractéristiques similaires à celles d'utilisateurs qui ont déjà été suspendus pour cause d'activité de bot.
- Contenu de type « bot » : nous utilisons des modèles pour détecter le contenu couramment envoyé par les bots, notamment les spams dans les chuchotements et les chats. Si des liens de désinformation étaient massivement partagés, ces modèles apprendraient à les intercepter.
- Détection des bots spectateurs : nous disposons de nombreux systèmes qui collaborent pour filtrer algorithmiquement les vues que les modèles considèrent comme frauduleuses.

L'API de Twitch comporte des points de terminaison qui permettent à un tiers, selon son rôle, d'agir au nom d'un utilisateur Twitch (ex. : pour créer un sondage ou une prédiction). Certains rôles utilisateur permettent d'effectuer des actions sur la chaîne d'un autre utilisateur. Les modérateurs, par exemple, peuvent réaliser des actions de modération sur la chaîne d'un streamer.

Pour pouvoir utiliser une API au nom d'un utilisateur Twitch, le tiers doit obtenir la permission de l'utilisateur en question. Cette permission est accordée sous la forme d'un jeton d'accès utilisateur associé à des niveaux d'autorisation spécifiques. Les API qui exigent la permission de l'utilisateur avant de donner l'accès à certaines ressources utilisent des jetons OAuth 2.0. Lorsqu'une application tierce demande un jeton d'accès à l'utilisateur, elle spécifie un ou plusieurs niveaux d'autorisation. La liste de ces niveaux d'autorisation est disponible à l'adresse suivante : <https://dev.twitch.tv/docs/authentication/scopes/#twitch-api-scopes>. Ces niveaux sont extrêmement précis et permettent à une application de ne demander que la fonctionnalité spécifique à laquelle elle doit accéder (par exemple, si votre application ne doit gérer que les termes bloqués, il existe un niveau d'autorisation spécifique pour cela).

Lorsque l'application demande cette autorisation à l'utilisateur, celui-ci est redirigé vers la page de destination OAuth de Twitch, qui répertorie le nom de l'application, les permissions demandées et leur description. L'utilisateur peut choisir d'accorder ou de refuser les autorisations demandées. Il est ensuite redirigé vers l'application tierce.

Lorsqu'un utilisateur accorde une autorisation demandée par une application tierce, celle-ci apparaît dans ses paramètres Twitch, dans l'onglet « Connexions ». Cet onglet affiche toutes les applications tierces auxquelles son compte est connecté. Pour déconnecter une application tierce de son compte, il suffit de cliquer sur le bouton « Se déconnecter » pour l'application en question. L'autorisation associée est alors révoquée (ainsi que tous les jetons d'accès que l'application a obtenus de l'utilisateur au cours de la procédure d'autorisation décrite ci-dessus).

#### 4.9 S'agissant des hypertrucages (« **deepfakes** ») :

- politique du service (si autorisation ou interdiction sous conditions, lesquelles ?) :
- mesures prises en cas d'infraction :
- moyens d'identification développés :

Nous n'avons pas actuellement de politique générale ou de mesures contre les « deepfakes ». Toutefois, nous interdisons les deepfakes lorsqu'ils sont utilisés dans le contexte d'images non consensuelles explicites (IENC) de synthèse ou lorsqu'ils enfreignent nos règles en matière de prévention du harcèlement.

## Questions complémentaires

#### 4.10 Remplir le tableau suivant :

L'approche de Twitch étant axée sur les agents de désinformation préjudiciable, les informations demandées ici sont les mêmes que celles fournies au point 1.11 ci-dessus.

#### 4.11 Impact de la **crise sanitaire la lutte contre les comptes propageant massivement de fausses informations** (ou notion voisine adoptée) (notamment, types de pratiques accrues dans ce contexte) et mesures spécifiques prises dans ce contexte en 2022 :

Les règles relatives aux agents de désinformation préjudiciable de Twitch sont entrées en vigueur en mars 2022 dans le but de réduire les risques pour notre communauté, notamment ceux liés à la désinformation en matière de santé.

#### 4.12 **Actions futures et nouvelles fonctionnalités envisagées à l'avenir**

Nous faisons constamment évoluer notre approche en fonction des conseils des experts et des tendances de notre communauté, et nous continuerons à nouer des partenariats avec le monde universitaire et la société civile. De même, en tant que membre du Code de bonnes pratiques de l'UE en matière de désinformation, nous continuerons à collaborer avec l'industrie.

## 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 majorité des contenus de marque sur notre service (publicités pré-roll, mid-roll et sponsoring) sont vendus par Twitch à des clients annonceurs. Toutes les publicités vendues par Twitch doivent être conformes aux Instructions et politiques d'acceptation des publicités pour Twitch. Les marques et les annonceurs doivent donc veiller à ce que leurs publicités soient claires, précises et vérifiables, et tout contenu trompeur ou mensonger est spécifiquement interdit. Les règles d'utilisation de Twitch en matière de publicité interdisent également :

- Les contenus menaçants, abusifs ou de harcèlement, et les contenus qui discriminent ou militent contre un groupe protégé (sur la base de l'origine ethnique, de la nationalité, de la religion, du handicap, du sexe, du genre, de l'orientation sexuelle, de l'âge, etc.)
- Les publicités politiques ou portant sur des sujets controversés pour des candidats, des partis ou des enjeux de débat public

Considérant les règles strictes mises en place en matière de publicité, la diffusion de désinformation par communication commerciale ne présente qu'un risque très faible sur Twitch. Par ailleurs, rien n'indique à ce jour une présence importante de désinformation dans les communications commerciales sur Twitch.

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

### Actions d'éducation aux médias et à l'information (EMI)

#### 6.1 **Actions de sensibilisation des utilisateurs à une utilisation maîtrisée du service** (ex : encart d'information au moment de la publication d'un message, rappel des règles de savoir-vivre en ligne, etc.), menées par l'opérateur seul ou en partenariat :

Twitch sensibilise les utilisateurs via les [Lignes de conduite de la communauté](#) qui détaillent précisément, exemples à l'appui, ce qui constitue un agent de désinformation préjudiciable. Lorsqu'un utilisateur enfreint les Lignes de conduite de la communauté, il reçoit une notification détaillée qui lui rappelle les règles applicables et lui donne de nombreux exemples supplémentaires de contenus non conformes. Nous publions également dans notre [Centre de sécurité](#), mis à



jour régulièrement, des articles de blog, des lignes de conduite, des articles éducatifs et d'autres ressources explicitant comment Twitch développe, actualise et applique ses conditions d'utilisation.

**6.2 Ressources et fonctionnalités développées** permettant aux utilisateurs d'identifier la fiabilité des informations et sources d'informations qui sont proposées sur le service :

Voir la réponse à la question 6.3 ci-dessous.

**6.3 Liste des campagnes menées, accompagnées ou promues en matière d'EMI** dans le domaine de la lutte contre la **manipulation de l'information en ligne**.

À noter que cette question nécessite de remplir le tableau en annexe du questionnaire qui consiste à indiquer pour chaque campagne les éléments suivants : (i) description et objectifs, (ii) partenariat, (iii) nature de l'implication de l'opérateur, (iv) public(s) visé(s), (v) dates de début et de fin, (vi) budget alloué, (vii) mise en valeur auprès du public, (viii) données sur les audiences touchées (si aucune, préciser les difficultés rencontrées), (ix) données sur l'impact de la campagne sur le comportement des publics visés et méthodes pour les mesurer (si aucune, préciser les difficultés rencontrées).

Twitch collabore actuellement avec les experts en éducation aux médias de MediaWise pour développer un ensemble de ressources éducatives visant à aider les streamers et les spectateurs sur Twitch à mieux identifier les fausses informations et la désinformation en ligne, et à éviter d'en partager. Dans le cadre de cette collaboration avec MediaWise, Twitch a organisé au 4e trimestre 2022 un stream en direct réunissant des participants de MediaWise, de Twitch et de la communauté Twitch, afin de partager des trucs et astuces pour repérer les fake news et la désinformation à l'approche des élections de mi-mandat aux États-Unis, et de façon plus générale, d'améliorer l'éducation aux médias. Un enregistrement de ce stream est disponible, ainsi que d'autres contenus originaux de MediaWise sur l'habileté numérique, dans le [Centre de sécurité de Twitch](#).

Le budget de cet événement organisé avec MediaWise était de 50 000 USD. Le stream en direct et l'enregistrement ont totalisé 2 800 vues (entre l'événement et la fin de l'année 2022). Nous n'avons pas recueilli de données relatives à l'impact de la campagne sur le comportement des utilisateurs.

**6.4 Évaluation par l'opérateur de ses campagnes d'EMI : OUI / NON**

- **critères d'évaluation** (notamment ceux permettant d'évaluer les conséquences réelles de la campagne sur l'utilisation du service) :
- **actions recommandées** à la suite de ces évaluations :

NON

**6.5 Commenter les préconisations de l'ARCOM aux opérateurs**, dans son précédent bilan sur la lutte contre la manipulation de l'information, **d'évaluer les conséquences sur le comportement des utilisateurs des actions en matière d'EMI et d'en rendre compte au public et au régulateur**.

Twitch continue d'explorer les opportunités éducatives liées à la désinformation et aux médias, et examinera la possibilité d'évaluer l'impact des futures campagnes sur le comportement des utilisateurs. La principale difficulté réside dans le fait que toute évaluation nécessitera probablement des enquêtes directes auprès des utilisateurs pouvant entraîner des coûts importants, dans la mesure où la désinformation n'est pas répandue sur Twitch.

**6.6 Impact de la crise sanitaire sur la politique de l'opérateur en matière d'EMI** et mesures spécifiques prises dans ce contexte en 2022 :

Pas d'impact spécifique

**6.7 Actions futures et nouvelles fonctionnalités envisagées à l'avenir** dans la mise en place de ses campagnes d'EMI :

Twitch continuera d'évaluer les opportunités éducatives potentielles.



## Études et relations avec la recherche

### 6.8 Études publiées par l'opérateur sur le fonctionnement du service (liste avec URL) :

Seering, Joseph et Sanjay Kairam : « [Who Moderates on Twitch and What Do They Do? Quantifying Practices in Community Moderation on Twitch](#) » (Qui sont les modérateurs sur Twitch et que font-ils ? Quantifier les pratiques de modération de communauté sur Twitch). Proceedings of the ACM on Human-Computer Interaction, Volume 7, Issue GROUP, Article 18, pp 1–18

### 6.9 Liste et description des initiatives indépendantes (journalistes, entreprises, chercheurs, etc.) que l'opérateur a soutenues ou auxquelles il a contribué, notamment avec des acteurs français :

- visant à mieux **cerner certains comportements et phénomènes liés à la manipulation de l'information en ligne**
- visant à vérifier la qualité des **systèmes de recommandation et de modération** du service en termes d'explicabilité, de loyauté et d'équité et, le cas échéant, proposer des pistes d'amélioration.

Twitch ne participe actuellement à aucune initiative indépendante avec des acteurs français.

### 6.10 Relations avec la recherche dans le domaine de la lutte contre la manipulation de l'information

- **moyens mis à disposition** des chercheurs (ex. : concernant la propagation de fausses informations, l'opérateur donne-il accès à des données – notamment via une API –, et si oui, sous quelles conditions cet accès est-il possible ?) :
- **types de données** communiquées :

En 2023, nous testerons un programme pilote visant à aider les chercheurs universitaires à accéder aux utilisateurs de Twitch qui souhaiteraient participer à des enquêtes ou à des groupes de discussion sur la désinformation et les fausses informations en général.

### 6.11 Politique de partage de données en *open data* : NON / OUI (préciser les modalités)

- indiquer les limites à cet accès aux données et **les difficultés rencontrées** :

Certains types de données sont disponibles via l'API publique de Twitch. Pour obtenir plus de détails, consultez la [documentation](#).

### 6.12 Actions futures pour développer les relations avec le monde de la recherche

Aucune autre action pour le moment.