



1er semestre 2023

# RAPPORT DE TRANSPARENCE



# TABLEAU DE CONTENU

<a href="#">03</a>	RÉSUMÉ EXÉCUTIF
<a href="#">05</a>	UTILISER L'INNOVATION POUR CRÉER DES EXPÉRIENCES PLUS SÉCURISÉES
<a href="#">06</a>	POINTS CLÉS À RETENIR
<a href="#">07</a>	RÉSUMÉ DES DONNÉES DE HAUT NIVEAU
<a href="#">08</a>	NOTRE VISION
<a href="#">11</a>	NOTRE APPROCHE
<a href="#">12</a>	normes communautaires
<a href="#">13</a>	Choix du joueur via les paramètres
<a href="#">14</a>	Contrôle parental
<a href="#">15</a>	Modération proactive
<a href="#">15</a>	Modération réactive
<a href="#">16</a>	Mise en vigueur
<a href="#">16</a>	Rapport sur le contenu de sécurité numérique de Microsoft
<a href="#">17</a>	Utiliser l'IA/ML pour créer des expériences plus sûres
<a href="#">18</a>	Aide lorsque les joueurs en ont besoin
<a href="#">19</a>	Appels/examens de cas
<a href="#">20</a>	PARTAGE DE NOS DONNÉES DE SÉCURITÉ
<a href="#">21</a>	Données de modération proactive
<a href="#">24</a>	Données de modération réactive (joueur signalé)
<a href="#">26</a>	Données d'application
<a href="#">28</a>	Données du rapport Microsoft sur le contenu de sécurité numérique
<a href="#">28</a>	Données de la ligne de texte de crise
<a href="#">29</a>	Données d'appel
<a href="#">30</a>	Toxicité évitée
<a href="#">31</a>	POLITIQUES ET PRATIQUES
<a href="#">32</a>	Information supplémentaire
<a href="#">33</a>	Infographie du système de frappe d'application
<a href="#">34</a>	Questions et réponses sur le système de frappe d'application
<a href="#">35</a>	GLOSSAIRE DES TERMES
<a href="#">36</a>	Glossaire
<a href="#">37</a>	ANNEXE
<a href="#">38</a>	Infographie du parcours du joueur
<a href="#">39</a>	Infographie de téléchargement d'image de joueur

# RÉSUMÉ EXÉCUTIF



## RÉSUMÉ EXÉCUTIF

Chez Xbox, notre mission est d'apporter la joie et la communauté du jeu à tous les habitants de la planète.

Lorsque vous venez jouer, vous méritez l'opportunité de découvrir un endroit libre de peur et d'intimidation, en sécurité dans les limites que vous avez fixées.



La sécurité de la communauté chez Xbox est un domaine unique et stimulant qui nécessite une innovation constante. Notre équipe continue de se concentrer sur la façon dont nous protégeons les joueurs contre les contenus inappropriés, construisons des expériences accueillantes et positives et aidons les joueurs à en apprendre davantage sur ce que nous pouvons faire ensemble pour respecter nos standards communautaires.

Les nouvelles avancées en matière d'IA, telles que les modèles permettant de détecter les images toxiques et d'améliorer la compréhension contextuelle de l'utilisation du langage, évoluent rapidement. Nous voyons une incroyable opportunité de développer nos outils existants basés sur l'IA et la connaissance humaine afin d'améliorer la vitesse, l'échelle et la précision avec lesquelles nous opérons. Nous continuons à faire évoluer ces systèmes pour accélérer nos progrès vers la création d'expériences de jeu plus inclusives, accessibles et plus sûres pour tous les joueurs.

Nous travaillons en étroite collaboration avec d'autres équipes Microsoft et continuons de travailler en tandem avec des partenaires industriels, des associations, des régulateurs et des membres de la communauté pour améliorer notre stratégie de sécurité à multiples facettes. À mesure que nous mettons en ligne de nouvelles technologies et fonctionnalités, nous aborderons leur impact dans les prochaines éditions de ce rapport. Nous restons déterminés à apprendre, à répéter et à être transparents quant à notre approche.

## UTILISER L'INNOVATION POUR CRÉER DES EXPÉRIENCES PLUS SÉCURISÉES

### Faire progresser la modération du contenu et la sécurité des plateformes grâce à l'IA

À mesure que l'industrie du jeu continue de croître, tous les systèmes de sécurité nécessitent encore plus de profondeur, de vitesse et d'agilité pour protéger les joueurs d'une toxicité potentielle. L'IA devient de plus en plus importante dans l'accélération de la modération du contenu dans le monde, et ici sur Xbox. Nos efforts visent à combiner l'importance de la surveillance humaine avec les capacités évolutives de l'IA pour s'appuyer sur les fondements de notre travail en matière de sécurité à ce jour. Nous innovons dans l'application de l'IA dans le domaine de la sécurité dans le but de créer des expériences de jeu plus sûres qui peuvent être utilisées dans l'ensemble de l'industrie.

Dans notre suite de sécurité actuelle, nous utilisons une gamme de modèles d'IA existants pour détecter les contenus toxiques tels que :

- Community Sift, une plateforme de modération de contenu basée sur l'IA et la connaissance humaine qui classe et filtre des milliards d'interactions humaines par an, alimentant bon nombre de nos systèmes de sécurité sur Xbox. L'entreprise a également effectué plus de 36 millions d'évaluations sur les rapports des joueurs Xbox au cours de la dernière année afin d'augmenter et d'aider à étendre le travail de nos modérateurs de contenu humain. La technologie de Community Sift est développée par des locuteurs natifs et prend en charge plus de 22 langues numériques dans le monde, y compris les emojis, pour capturer l'évolution des nuances culturelles en temps réel.
- L'utilisation de Turing Bletchley v3, un modèle fondamental de langage de vision multilingue qui analyse les images générées par l'utilisateur pour garantir que seul le contenu approprié est montré. Ce modèle a contribué au total de 4,7 millions de contenus bloqués au cours de cette dernière période.

Notre équipe développe activement des moyens par lesquels nos systèmes de sécurité peuvent être encore améliorés par l'IA et le Community Sift afin de classer et d'agir avec une meilleure compréhension du contexte des interactions. Ces solutions nous permettent d'atteindre une plus grande échelle, d'élever les capacités de nos modérateurs humains et de réduire l'exposition aux contenus sensibles.

Nous ne prenons pas cette technologie ou cette responsabilité à la légère. Nous nous engageons à créer une IA responsable dès la conception, et notre travail est guidé par la norme IA responsable de Microsoft. L'IA doit être associée à une surveillance humaine pour garantir l'exactitude, la nuance et le contexte culturel du contenu évalué.

## RÉSUMÉ EXÉCUTIF

### Points clés du rapport

01

#### De nouvelles informations sur les volumes de contenu bloqués

Prévenir la toxicité avant qu'elle n'atteigne nos joueurs est un élément crucial de nos efforts proactifs visant à offrir une expérience accueillante et inclusive pour tous. Notre équipe combine une IA responsable avec une supervision humaine pour examiner et empêcher la publication de contenus préjudiciables sur notre plateforme. Pour mieux mesurer notre succès, nous incluons désormais une nouvelle mesure couvrant notre travail dans cet espace appelée Toxicité évitée. Au cours de cette dernière période, plus de 4,7 millions de contenus ont été bloqués avant d'atteindre les joueurs, dont une augmentation de 135 000 (+39 % par rapport à la dernière période) des images grâce aux investissements dans l'utilisation du nouveau Turing Bletchley v3. \_\_\_\_\_  
modèle de fondation.

02

#### Accent accru sur la lutte contre le harcèlement

Nous nous engageons à créer un environnement sûr et inclusif pour tous les joueurs. Nous travaillons activement à identifier et à lutter contre tout comportement abusif, y compris les discours de haine, l'intimidation et le harcèlement. Avec cet objectif à l'esprit, nous avons apporté des améliorations à nos processus internes pour accroître nos efforts proactifs d'application de la loi au cours de la période écoulée en émettant 84 000 plaintes pour harcèlement/intimidation. \_\_\_\_\_  
des mesures proactives (+95 % par rapport à la dernière période). Nous avons également lancé notre nouvelle fonctionnalité de rapport vocal pour capturer et signaler le harcèlement vocal dans le jeu. Notre équipe de sécurité continue de prendre des mesures proactives pour garantir que tous les joueurs sont conscients que tout comportement abusif de toute nature est inacceptable sur notre plateforme, et nous prenons ce comportement au sérieux.

03

#### Comprendre le comportement des joueurs après l'application

Nous profitons toujours de l'occasion pour en savoir plus sur la façon dont nous pouvons conduire une meilleure compréhension des standards de la communauté avec nos joueurs. À cette fin, nous avons analysé le comportement des joueurs après une application. Les premières informations indiquent que la majorité des joueurs ne violent pas les standards de la communauté après avoir reçu une mise en application et s'engagent positivement avec la communauté. Pour aider davantage les joueurs à comprendre ce qui constitue ou non un comportement acceptable, nous avons récemment lancé notre système de frappe d'application, conçu pour mieux aider les joueurs à mieux comprendre la gravité de l'application, l'effet cumulé de plusieurs mises en application et l'impact total sur leur dossier. Des FAQ supplémentaires sont disponibles. \_\_\_\_\_

Résumé des données de sécurité de haut niveau du premier semestre 2023 (janvier – juillet 2023)

## Rapports des joueurs

27,31 millions

13,71 M (50 %) Communications

10,54 M (39 %) Conduites

3,07 M (11 %) Contenu généré par les utilisateurs

## Exécutions émises

19,56 millions

17,09 millions (87 %) Proactifs<sup>1</sup>

2,47 millions (13 %) Réactifs<sup>2</sup> (à partir des rapports des joueurs)

## Rapports NCMEC<sup>3</sup>

766

Références de lignes de texte de crise

2 225

## Appels (examen de cas)

280,15k

268,55k (96%) Non-réintégrations 11,60k

(4%) Réintégrations

<sup>1</sup>Application proactive – Lorsque nous agissons sur un contenu ou un comportement inapproprié avant qu'un joueur ne le porte à notre attention.

<sup>2</sup>Application réactive – Lorsque nous agissons sur un contenu ou un comportement inapproprié via un joueur qui le porte à notre attention.

<sup>3</sup>NCMEC – Centre national pour les enfants disparus et exploités

# NOTRE VISION





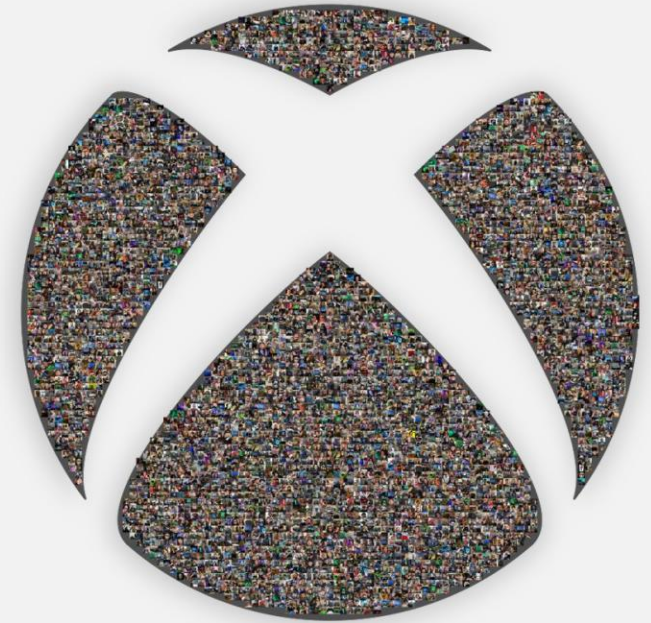
## NOTRE VISION

La communauté Xbox est la vôtre.

Nous apportons tous quelque chose d'unique, et ce caractère unique mérite d'être protégé.

Que vous soyez nouveau dans le jeu vidéo ou que vous jouiez depuis des décennies, vous êtes les intendants de cet endroit, vous protégeant les uns les autres même lorsque vous êtes en compétition.

Parce que quand tout le monde joue, nous gagnons tous.



## NOTRE VISION

Nos normes de la communauté Xbox décrivent la conduite et le contenu qui sont acceptables au sein de notre communauté. Nous reconnaissons qu'une activité négative peut avoir lieu et a eu lieu. Cette conduite n'est pas acceptable et va à l'encontre de la communauté que nous nous efforçons de créer – un lieu dynamique, sûr et accueillant.

Nous voulons que vous ayez l'assurance que nous écoutons et agissons en fonction de vos commentaires. Nous utilisons ces commentaires pour tester et implémenter de nouvelles fonctionnalités, et mieux comprendre l'activité et le comportement de nos joueurs. Une façon de nous aider à offrir la meilleure expérience de jeu possible est de nous faire part de vos commentaires, et en participant à notre programme Xbox Insider.



# NOTRE APPROCHE



## NOTRE APPROCHE

### Notre approche multifacette

- Travailler à créer une communauté forte de joueurs réfléchis à leur conduite et guidés par des [normes communautaires](#)
- Donner aux joueurs des contrôles pour personnaliser leurs paramètres sur l'ensemble Écosystème Xbox de la console au PC en passant par Xbox Cloud Gaming (bêta), y compris un [contrôle parental complet afin que les enfants](#) puissent vivre des expériences plus sûres et qui leur conviennent
- Utiliser une technologie et des outils proactifs pour détecter et supprimer contenu problématique avant qu'il ne soit vu et pour réduire les comportements qui vont à l'encontre de nos normes communautaires
- Activation d'[outils de reporting utiles](#) pour que nos joueurs identifient les problèmes
- Un [processus d'appel](#) pour informer nos utilisateurs sur la communauté Normes
- Un nouveau [système de frappe d'exécution](#) pour aider les joueurs à mieux comprendre leurs exécutions et les empêcher de répéter
- Apprentissage continu et investissement dans nos mesures de sécurité

---

↔ [Découvrez notre engagement commun pour un jeu plus sûr](#)

Protéger notre communauté nécessite un travail et une diligence constants. Notre approche fondamentale de la sécurité dès la conception et notre équipe dévouée garantissent que la sécurité est et sera toujours une priorité pour chacun.



## NOTRE APPROCHE



## normes communautaires

Le [contrat de services Microsoft](#) La section [Code de conduite](#) s'applique à Xbox et à ses joueurs. Nos [normes de la communauté Xbox offre un niveau d'explication](#) supplémentaire, fournissant des détails sur nos attentes en matière de comportement des joueurs sur notre réseau. Elles reflètent également les politiques que nous avons mises en place pour modérer les comportements et, si nécessaire, imposer des conséquences aux joueurs qui enfreignent nos politiques.

---

↔ [En savoir plus sur les normes de la communauté Xbox](#)



## NOTRE APPROCHE

### Choix du joueur via les paramètres

Nous savons qu'en matière de préférences en matière de contenu et d'expériences, il n'y a pas de solution universelle. Un contenu ou une langue qui convient à un joueur peut ne pas convenir à d'autres.

Nous offrons à nos joueurs le choix des types de contenu qu'ils souhaitent voir et expérimenter sur notre réseau, notamment :

- Filtrage automatisé des textes, des médias et des liens Web afin que vous puissiez décider quels messages texte vous seriez à l'aise de recevoir
- Flexibilité du filtre, permettant aux joueurs de configurer les paramètres de sécurité selon un spectre allant du plus filtré au moins afin que vous puissiez choisir ce qui vous convient le mieux
- Contrôles parentaux personnalisables, y compris une famille Xbox pratique [Application Paramètres](#) sur les appareils mobiles
- Muet et blocage les autres joueurs et leurs messages
- Partage du vrai nom si les joueurs souhaitent partager leur vrai nom avec amis

Chaque joueur a la possibilité d'ajuster et de sélectionner ses paramètres de confidentialité et de sécurité à tout moment, ces paramètres étant effectifs pour toutes les manières dont les joueurs accèdent à Xbox.

---

 [En savoir plus sur les paramètres de sécurité pour les messages Xbox](#)

 [En savoir plus sur la gestion des paramètres de sécurité et de confidentialité Xbox](#)

## NOTRE APPROCHE



## Contrôle parental

Xbox propose un ensemble robuste de contrôles parentaux qui aident les enfants sur notre plateforme à vivre des expériences plus sûres sur nos services, y compris une application pratique Paramètres Xbox Famille pour les appareils mobiles. Les comptes enfants sur Xbox sont dotés de paramètres par défaut qui empêchent les enfants de regarder ou de jouer à des jeux classés pour adultes et nécessitent une autorisation parentale pour d'autres actions telles que jouer à des jeux multijoueurs, discuter avec d'autres joueurs et effectuer des achats. Les parents peuvent également recevoir des rapports d'activité hebdomadaires sur le temps passé par leurs enfants sur Xbox, y compris les jeux joués, le temps passé sur chaque jeu et les achats effectués.

Nous nous soucions profondément de ce que veut notre communauté Xbox. C'est pourquoi nous avons continué à accroître nos capacités depuis le lancement de notre application Xbox Family Settings. Grâce aux commentaires directs des parents de joueurs, nous avons ajouté davantage d'options pour éviter achats non autorisés et la capacité des soignants à adopter de bonnes habitudes en matière de temps d'écran.

Ces options contribuent également à susciter la conversation entre parents et enfants pour aider les jeunes joueurs à développer de meilleures compétences numériques et à naviguer en toute sécurité dans leur présence en ligne.

---

↳ [Téléchargez l'application Paramètres de la famille Xbox](#)

↳ [En savoir plus sur le contrôle parental](#)

↳ [En savoir plus sur l'application Paramètres de la famille Xbox](#)

## NOTRE APPROCHE

### Mise en vigueur

Lorsqu'il s'avère que le comportement ou le contenu d'un joueur enfreint nos politiques, les agents ou les systèmes de modération du contenu prendront des mesures - nous appelons cela une application. Le plus souvent, cela prend la forme de la suppression du contenu incriminé du service et de la suspension du compte associé.

La durée de la suspension dépend principalement du type de comportement ou de contenu incriminé tout en tenant compte de l'historique du compte. Les violations répétées des politiques entraînent des suspensions plus longues et peuvent aboutir à une suspension de 12 mois des fonctionnalités sociales. Des violations particulièrement flagrantes peuvent entraîner des suspensions permanentes de compte ou des interdictions d'appareils.

Nous avons récemment introduit [un nouveau système de frappe](#) à notre approche d'application conçue pour mieux informer les joueurs sur les applications et pour leur permettre de s'engager de manière positive et appropriée sur Xbox et avec la communauté.

---

[↔ En savoir plus sur les types d'applications](#)

[↔ FAQ sur le système de frappe d'exécution](#)

[↔ FAQ sur les mesures coercitives](#)

### Rapport sur le contenu de sécurité numérique de Microsoft

Depuis plusieurs années, Microsoft publie un rapport semestriel [sur la sécurité numérique Content Report \(DSCR\)](#), qui couvre les mesures prises par Microsoft contre les contenus terroristes et extrémistes violents (TVEC), [les images intimes non consentuelles \(NCII\)](#), [les images d'exploitation et d'abus sexuels sur enfants \(CSEAI\)](#) et [le toilettage d'enfants à des fins sexuelles](#) dans l'ensemble de ses activités. services aux consommateurs, y compris Xbox.

Chez Xbox, les violations de nos politiques CSEAI, toilettage d'enfants à des fins sexuelles ou TVEC entraîneront la suppression du contenu et une suspension permanente du compte, même s'il s'agit d'une première infraction. Ces types de cas, ainsi que les menaces contre la vie (de soi-même, d'autrui, du public) et d'autres préjudices imminents, font immédiatement l'objet d'une enquête et sont transmis aux forces de l'ordre, si nécessaire.

---

[↔ En savoir plus sur le rapport sur le contenu de sécurité numérique \(DSCR\)](#)



## NOTRE APPROCHE

### Modération proactive

Pour réduire le risque de toxicité et empêcher nos joueurs d'être exposés à du contenu inapproprié, nous utilisons des mesures proactives qui identifient et arrêtent les contenus nuisibles avant qu'ils n'affectent les joueurs. Par exemple, [la modération proactive](#) nous permet de rechercher et de supprimer des comptes inauthentiques afin d'améliorer l'expérience des vrais joueurs.

Depuis des années chez Xbox, nous utilisons un ensemble de technologies de modération de contenu pour nous aider de manière proactive à traiter les textes, images et vidéos violant les règles partagés par les joueurs sur Xbox. Grâce à ces méthodes de modération courantes, nous avons pu automatiser certains de nos processus. Cette automatisation nous permet d'atteindre une plus grande échelle, d'élever les capacités de nos modérateurs humains et de réduire l'exposition au contenu sensible. Si un contenu qui enfreint nos politiques est détecté, il peut être bloqué ou supprimé de manière proactive.

### Modération réactive

Le blocage et le filtrage proactifs ne constituent qu'une partie du processus de réduction de la toxicité de notre service. Xbox offre des fonctionnalités de reporting robustes, en plus [des contrôles de confidentialité et de sécurité](#) et la [possibilité de couper le son et bloc](#) d'autres joueurs ; cependant, un contenu inapproprié peut traverser les systèmes et atteindre un joueur.

La modération réactive est toute modération et examen du contenu qu'un [joueur signale à Xbox](#). [Lorsqu'un joueur](#) signale un autre joueur, un message ou tout autre contenu sur le service, le rapport est enregistré et envoyé à notre plateforme de modération pour examen par les technologies de modération de contenu et les agents humains. Ces rapports réactifs sont examinés et traités conformément aux politiques pertinentes applicables. Nous considérons les joueurs comme des partenaires dans notre voyage et nous souhaitons travailler avec la communauté pour réaliser notre [vision](#).

## NOTRE APPROCHE

### Aide lorsque les joueurs en ont besoin

Nous cherchons également à aider nos joueurs lorsqu'ils en ont besoin. Si les communications d'un joueur sont signalées comme préoccupantes (y compris le contenu associé à des idées suicidaires ou à l'automutilation), soit par notre système, soit par d'autres joueurs, nous pouvons fournir des informations sur la ligne de texte de crise au joueur afin qu'il puisse contacter des ressources qui peuvent l'aider. .

Crisis Text Line est une organisation à but non lucratif basée aux États-Unis que Xbox a partenaire depuis 2018, qui fournit une assistance textuelle gratuite 24h/24 et 7j/7 .



## NOTRE APPROCHE



## Appels/examens de cas

Notre ~~processus~~ **processus d'appel** permet à un joueur d'obtenir plus d'informations sur les mesures d'application qu'il a reçues, y compris les suspensions de compte ou les suppressions de contenu. Un joueur peut lancer un appel, également appelé examen du cas, pour nous fournir plus d'informations s'il n'est pas d'accord avec notre détermination selon laquelle une politique a été violée. Sur la base de l'appel, la décision initiale peut être confirmée, modifiée ou annulée et le compte rétabli.

---

↔ [Comment déposer une révision de dossier](#)

↔ [En savoir plus sur les types d'applications](#)

↔ [FAQ sur les mesures coercitives](#)

↔ [FAQ sur le système de frappe d'exécution](#)

# PARTAGER NOTRE DONNÉES DE SÉCURITÉ



Les données que nous partagerons ci-dessous couvrent la période comprise entre le 1er janvier et le 30 juin 2023 et ont été collectées conformément à [l'engagement de Microsoft en matière de confidentialité](#).

## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

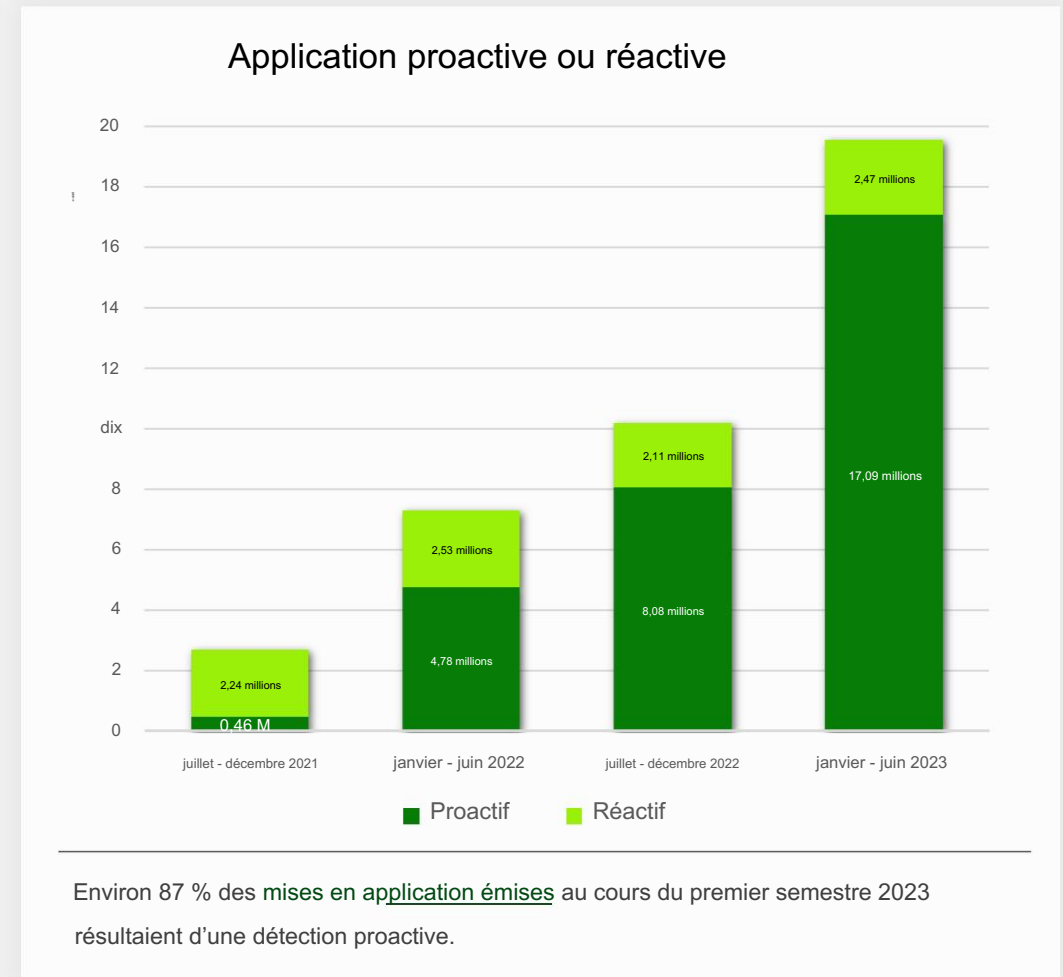
### Données de modération proactive

Les mesures proactives consistent à utiliser notre portefeuille de technologies et de processus de protection pour détecter et gérer un problème avant qu'il ne soit porté à notre attention par un joueur. Au cours de la période de référence, nous avons constaté une multiplication par deux des mesures proactives par rapport au rapport précédent.

En examinant notre travail proactif au cours de la dernière période, 16,3 millions de contrôles sur 17,09 millions (95 %) étaient centrés sur la détection de comptes falsifiés ou utilisés de manière non authentique. C'est le résultat de notre investissement continu dans la détection proactive et la suppression des comptes inauthentiques.

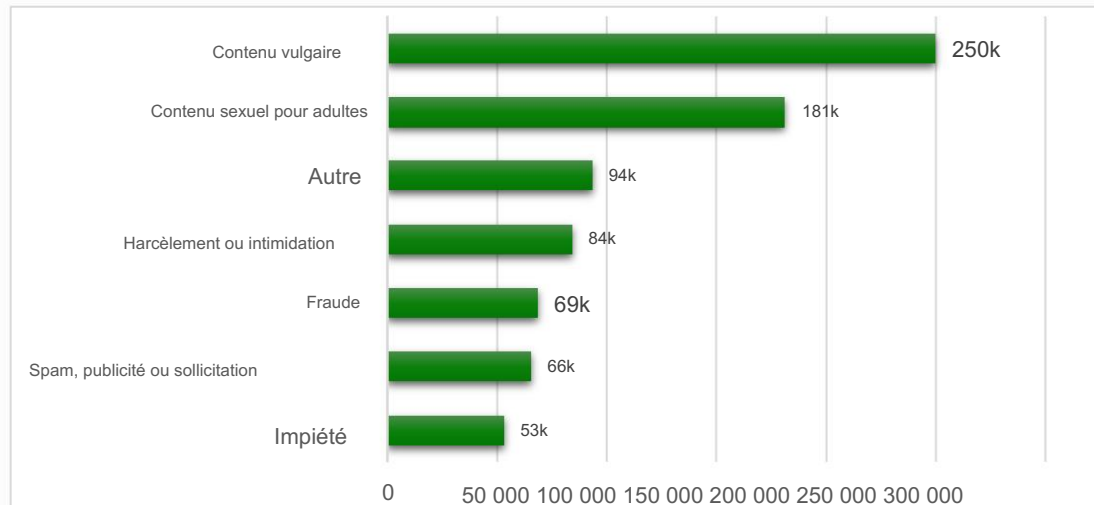
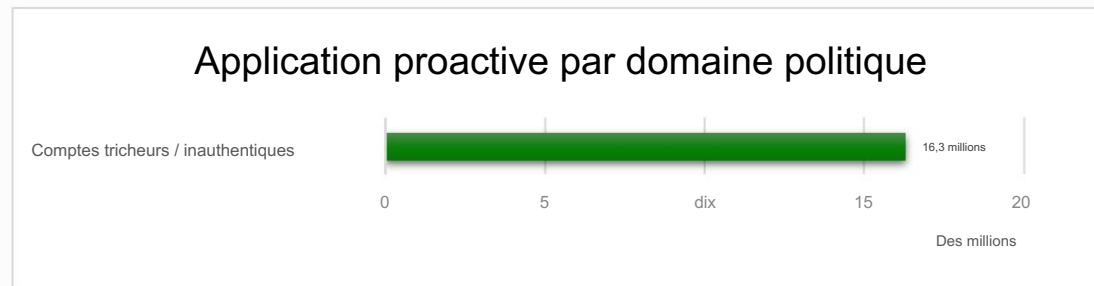
Ces comptes ont un impact sur les joueurs de multiples façons, notamment par la production de messages non sollicités (spam), la facilitation d'activités de triche qui perturbent le jeu, l'inflation inappropriée du nombre d'amis/suiveurs et d'autres actions qui, en fin de compte, créent des règles du jeu inéquitables pour nos joueurs ou nuisent à leur jeu. à partir de leurs expériences.

Vous trouverez ci-dessous une répartition des applications proactives et réactives au fil du temps :



## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

Nous pouvons diviser nos mesures proactives en domaines politiques pour la période de 6 mois précédente :



Au-delà de notre objectif de bloquer les comptes non authentiques dès leur création, les autres domaines qui connaissent un nombre élevé de mesures proactives incluent le contenu vulgaire, le contenu sexuel pour adultes, le harcèlement ou l'intimidation, la fraude, le spam et l'impunité. La catégorie Autres comprend des domaines de moindre volume tels que le piratage, falsification de comptes, drogues et discours de haine dans leur ensemble.

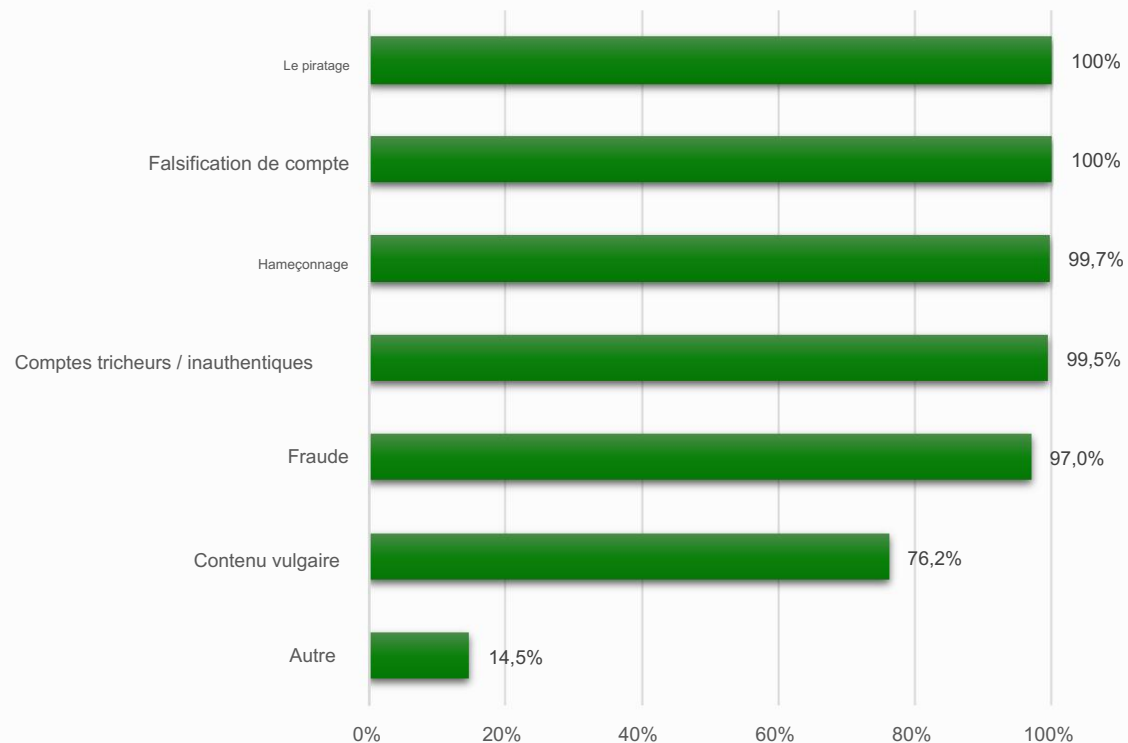
Nous avons constaté une augmentation de 95 % des mesures proactives de harcèlement ou d'intimidation grâce à l'amélioration des processus internes pour lutter contre ce type de comportement. En nous appuyant sur notre définition élargie du terme vulgaire de la dernière période, nous avons continuellement mis à jour nos méthodologies de détection et d'évaluation, ce qui a abouti à une augmentation de 184 % des mises en application proactives des contenus vulgaires.

Nous devons prendre en compte plusieurs facteurs lors de l'examen du nombre d'applications proactives par domaine politique, notamment la quantité de ce type de contenu sur la plate-forme, l'efficacité de nos technologies proactives pour détecter ce contenu et si nous offrons aux utilisateurs des contrôles personnalisés pour gérer eux-mêmes leur expériences dans ces domaines (ce qui produit des mises en application réactives).

## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

Nous pouvons continuer à examiner les mises en application en examinant le % qui ont été émis de manière proactive (avant qu'un acteur ne porte le problème à notre attention) par domaine politique au cours de la période de 6 mois précédente :

### % d'applications proactives par domaine politique



Traiter les comportements et contenus inappropriés avant qu'ils ne nous soient signalés par les joueurs est un élément important pour créer un environnement de jeu sain et compétitif.

En plus de notre objectif de bloquer les comptes non authentiques, les autres domaines dans lesquels des pourcentages élevés de mesures proactives sont appliquées incluent la falsification de compte, le piratage, le phishing et le contenu vulgaire. La catégorie Autre comprend des domaines tels que les drogues, les grossièretés, les discours de haine, le harcèlement ou l'intimidation, le spam, la publicité ou la sollicitation.

## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

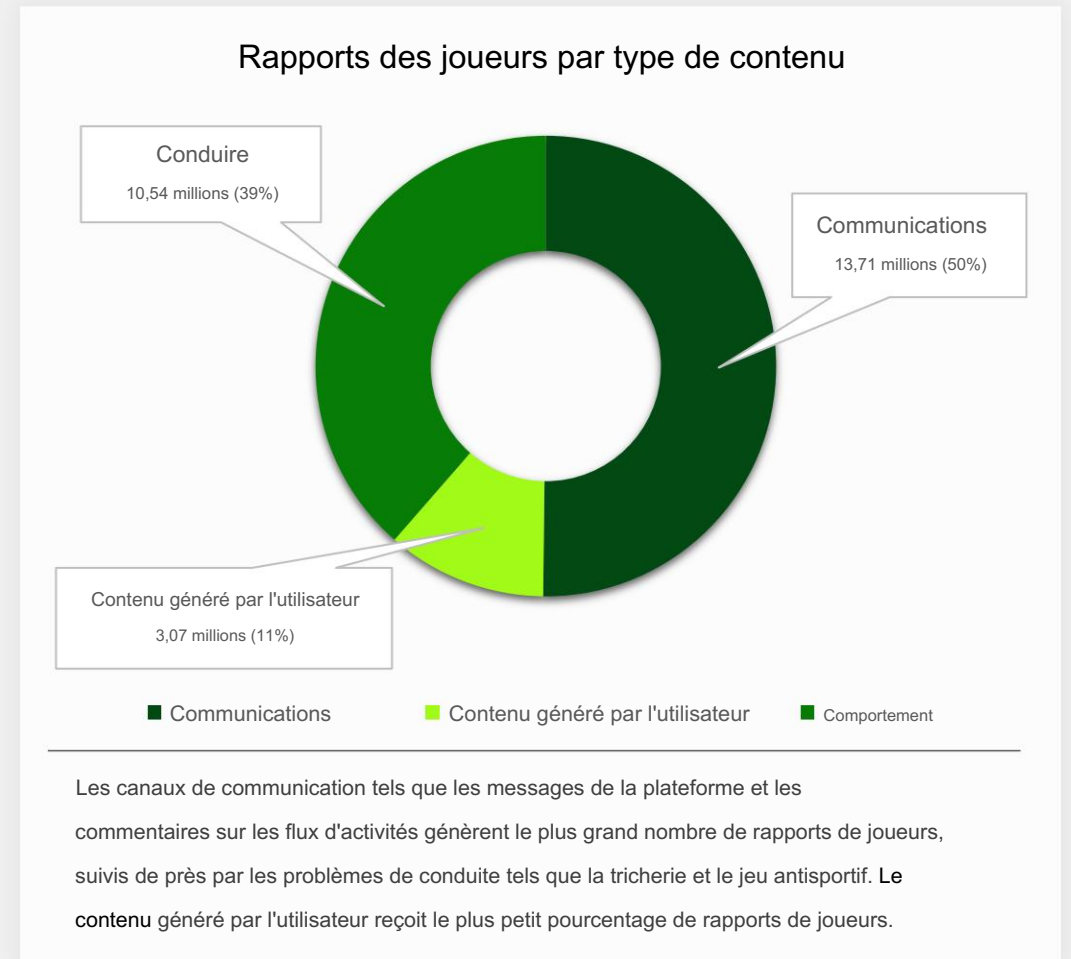
### Données de modération réactive (joueur signalé)

Lorsqu'un joueur porte quelque chose à notre attention au lieu d'être détecté par notre système, nous considérons ce rapport comme réactif.

Nous classons le contenu signalé par les joueurs en trois catégories principales :

- 01 Conduite – La manière dont un joueur agit sur Xbox, y compris la tricherie, les comportements antisportifs tels que le chagrin, le teamkilling, etc.
- 02 Contenu généré par l'utilisateur (UGC) – Tout contenu créé par un joueur qui n'est pas lié à la messagerie, comme un gamertag, le logo d'un club ou une capture d'écran ou un clip vidéo téléchargé.
- 03 Communications – Contenu lié à la communication avec d'autres joueurs, comme un message de plateforme ou un commentaire laissé sur une publication de fil d'activité.

Vous trouverez ci-dessous une vue des rapports des joueurs en fonction de la catégorie du rapport :





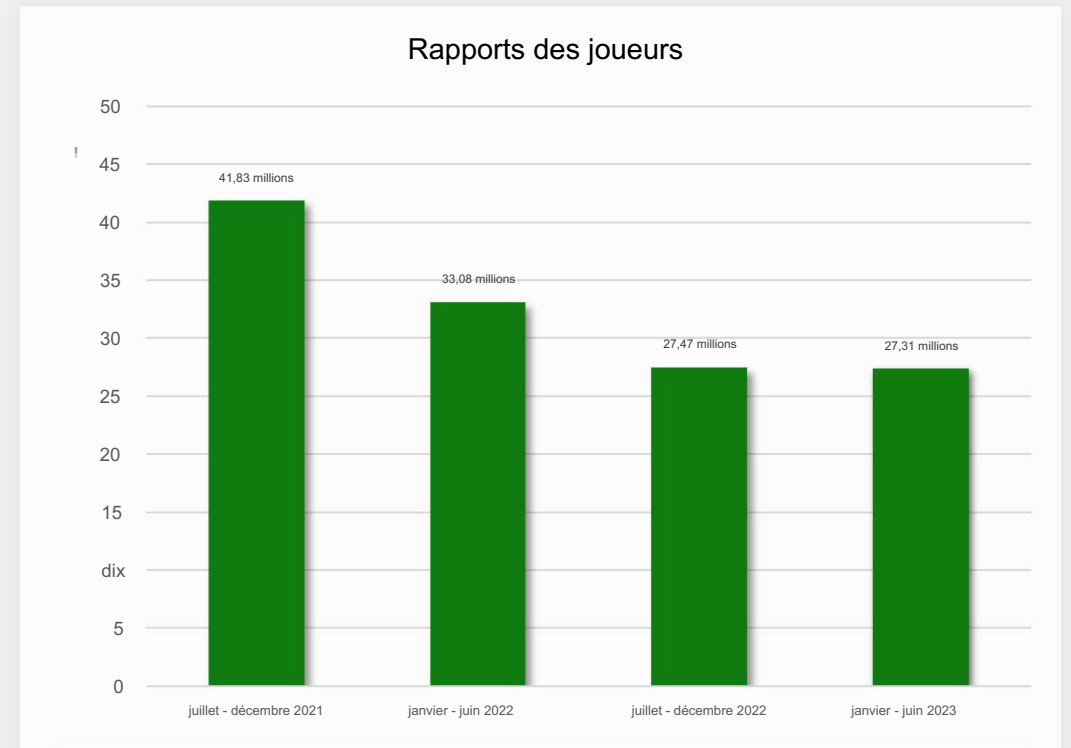
## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

### Rapports des joueurs

Lorsque les rapports des joueurs entrent dans le système, ils sont souvent d'abord évalués par les technologies de modération de contenu pour voir si une violation peut être déterminée, le reste étant examiné par des agents humains de modération de contenu pour prendre une décision.

Les agents de modération de contenu sont présents 24 heures sur 24, 7 jours sur 7, 365 jours par an pour s'assurer que le contenu et le comportement trouvés sur notre plateforme respectent nos normes communautaires.

Vous trouverez ci-dessous le nombre de rapports soumis par les joueurs :



La fréquence à laquelle les joueurs déposent des rapports sur la plateforme Xbox a diminué au cours des dernières années.

## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

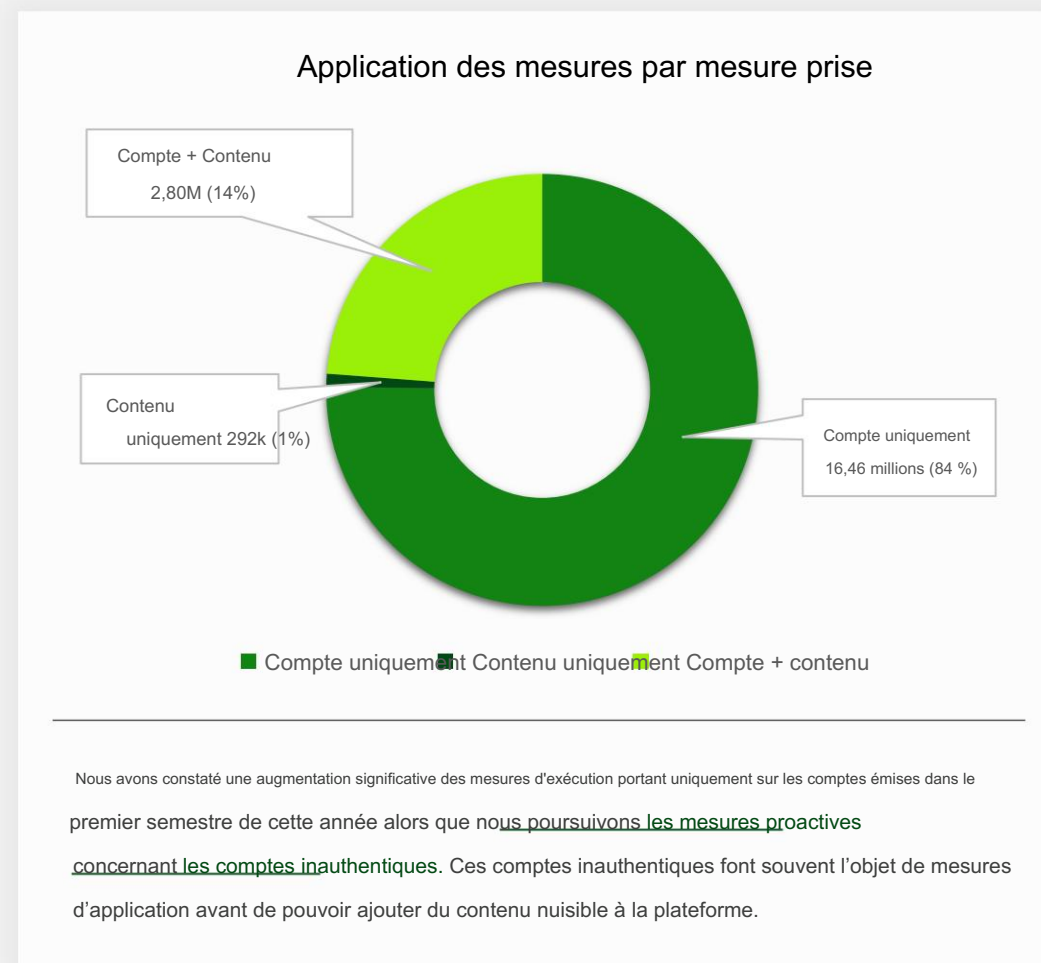
### Données d'application

Lorsqu'il est déterminé qu'une violation de nos standards communautaires a eu lieu, l'une des trois choses suivantes se produit :

- 01 Le contenu est supprimé (Application du contenu uniquement)
- 02 Le compte joueur est suspendu (Application du compte uniquement)
- 03 Une combinaison des deux se produit (Compte + Application du contenu)

Ces actions sont appelées une **exécution**. \_\_\_\_\_

Nous examinons ici les types de mesures coercitives prises au cours des six premiers mois de 2023 :

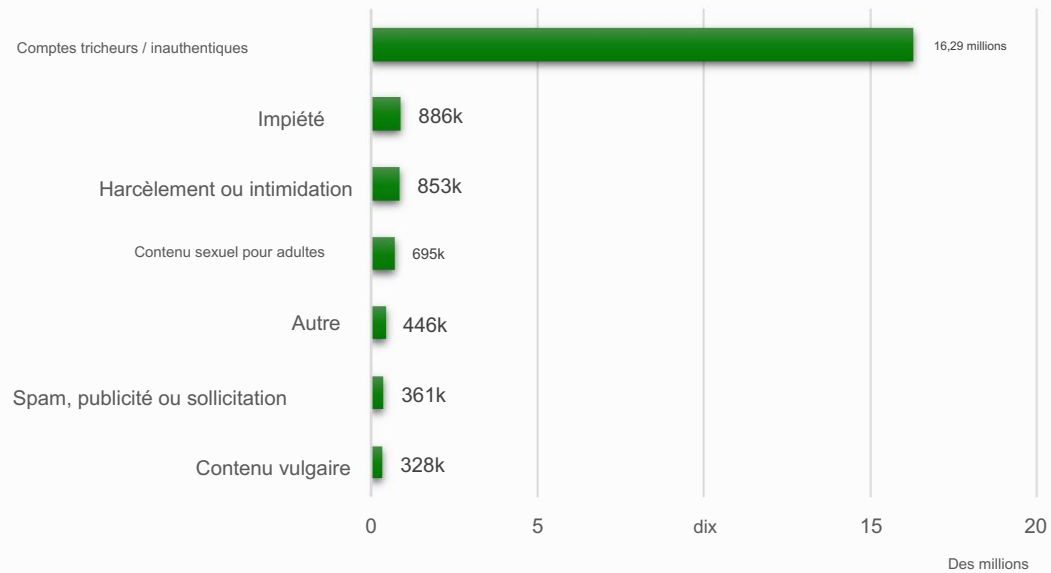


## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

La plupart des mesures d'application sont classées selon le domaine politique dans lequel la violation a eu lieu.

Une répartition des domaines de violation des politiques les plus courants (provenant de sources proactives et réactives) peut être consultée ci-dessous :

Total des mesures d'application par domaine politique



Les comptes frauduleux/inauthentiques sont notre domaine avec le plus grand nombre d'applications, les grossièretés, le harcèlement ou l'intimidation, le contenu sexuel pour adultes et le spam étant les autres types de politiques qui complètent notre top cinq. La catégorie Autres comprend des domaines de moindre volume tels que le piratage, phishing, falsification de compte, contenu vulgaire ou drogue.

## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

### Données du rapport Microsoft sur le contenu de sécurité numérique

En tant que société basée aux États-Unis, Microsoft signale toutes les images apparentes d'exploitation ou d'abus sexuels sur des enfants (CSEAI) ou le toilettage d'enfants à des fins sexuelles au Centre national pour les enfants disparus et exploités (NCMEC) via CyberTipline, comme l'exige la loi américaine.

Au cours de la période couverte par ce rapport, 766 des rapports de Microsoft provenaient de Xbox.

De plus amples informations sur les efforts de Microsoft concernant le CSEAI, le toilettage d'enfants à des fins sexuelles et les contenus terroristes et extrémistes violents (TVEC) sont disponibles dans le rapport sur le contenu de sécurité numérique.

### Données de la ligne de texte de crise

Les préoccupations réelles les plus courantes que nous constatons sur la plateforme concernent les menaces d'automutilation, qui sont traitées en référant à services de conseil via la Crisis Text Line.

Au cours de la période couverte par ce rapport, nous avons envoyé 2 225 messages Crisis Text Line aux joueurs.

## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

### Données sur les appels (examen des cas)

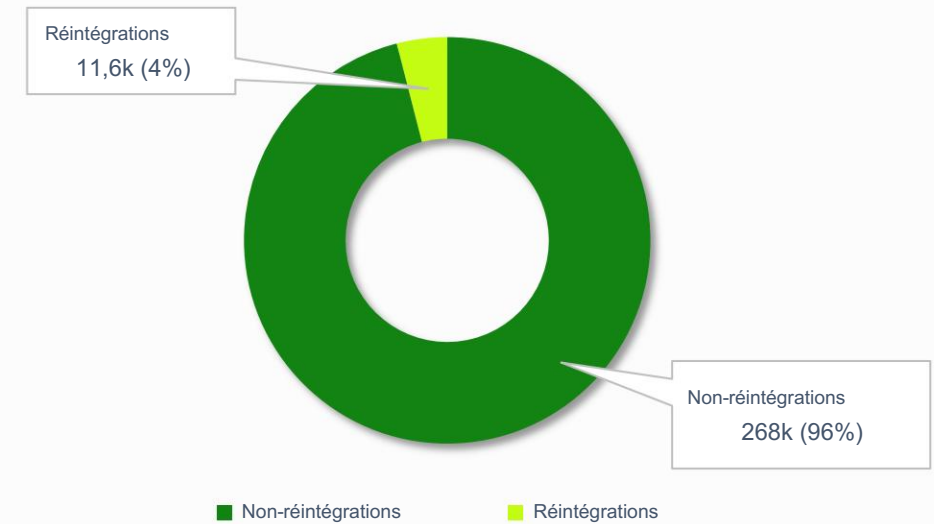
Lorsqu'un joueur reçoit une exécution au-delà d'un certain délai, il peut contester ou demander des éclaircissements par le biais d'un appel, également appelé examen du cas.

Lors du dépôt d'une révision de cas, le joueur peut expliquer ses actions et un agent de modération examinera le cas pour voir si une erreur a été commise ou si un réexamen spécial est justifié.

Au cours de la dernière période, nous avons traité plus de 280 000 dossiers d'appel, soit une hausse de 22 % par rapport à la période précédente.

Nous examinons ici le volume d'appels traités et le pourcentage associé de comptes rétablis :

### Appels (examen des cas) Volume et % de réintégration



Nous avons traité plus de 280 000 appels (examens de dossiers) au cours de cette dernière période, avec un taux de réintégration d'environ 4,1 %. Les réintégrations sont émises lorsqu'une erreur est découverte ou si le joueur mérite un réexamen spécifique à son application. Il y a un non-rétablissement lorsque la mesure coercitive initiale a été jugée justifiée et maintenue après examen.

## PARTAGE DE NOS DONNÉES DE SÉCURITÉ

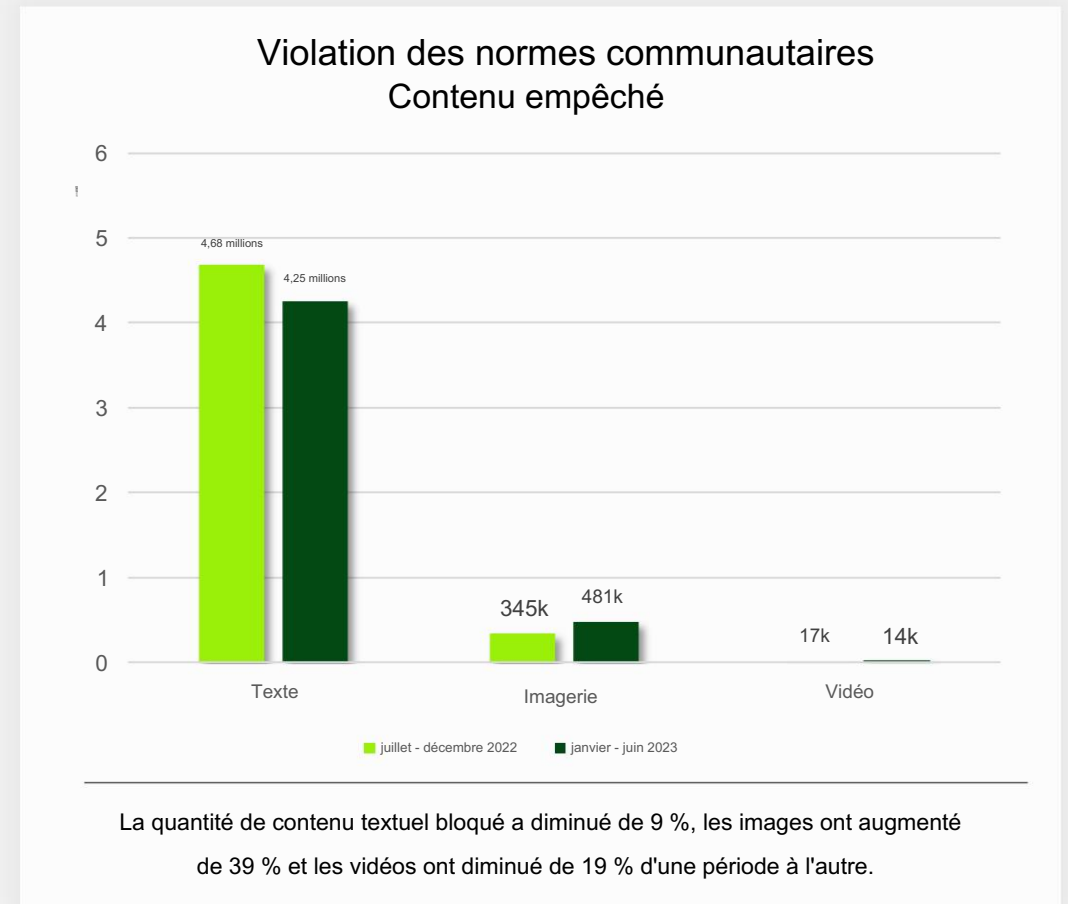
# Toxicité évitée

Alors que nous nous efforçons de créer des expériences plus sûres pour nos joueurs, mesurer l'impact des technologies de sécurité que nous mettons en œuvre devient une considération importante. Garder une trace de la quantité de contenu que nous empêchons d'entrer ou de proliférer sur notre plateforme est l'un des meilleurs indicateurs de la performance des technologies que nous utilisons dans le domaine de la sécurité.

- Texte – La quantité de texte, comme un commentaire inapproprié sur le fil d'activité de quelqu'un, que nous avons empêché
- Images – La quantité d'images, telles qu'une image de joueur inappropriée téléchargée, que nous avons empêchées.
- Vidéo – La quantité de vidéo, comme un GameDVR inapproprié clip, que nous avons empêché

L'augmentation de 39 % de la prévention des images inappropriées est le résultat d'investissements dans des technologies avancées telles que [Turing Bletchley v3](#), modèle de fondation.

Vous trouverez ci-dessous le nombre de contenus en infraction qui ont été empêchés d'entrer ou de proliférer sur notre plateforme, ventilés par type de contenu :



# STRATÉGIES ET PRATIQUES



## POLITIQUES ET PRATIQUES

Voici quelques informations supplémentaires qui peuvent vous aider à mieux comprendre le contenu de ce rapport :



### Politique et normes

- [Normes de la communauté Xbox](#)
- [Contrat de services Microsoft](#)



### Processus de déclaration

- [Comment signaler un joueur](#)



### Processus d'appel (examen des cas)

- [Comment soumettre une révision de cas](#)



### Glossaire des définitions

- [Définitions](#)



### Ressources additionnelles

- [Sécurité familiale et en ligne](#)
- [Confidentialité et sécurité en ligne](#)
- [Contrôle parental](#)
- [Centre familial](#)
- [Jeu responsable pour tous](#)
- [En savoir plus sur l'application Paramètres Xbox Family](#)
- [En savoir plus sur les paramètres de sécurité pour la messagerie Xbox](#)
- [Application Paramètres de la famille Xbox](#)
- [Programme Xbox Insiders](#)
- [Tableau de bord de confidentialité](#)
- [FAQ sur le système de frappe d'exécution](#)
- [Rapports vocaux réactifs](#)



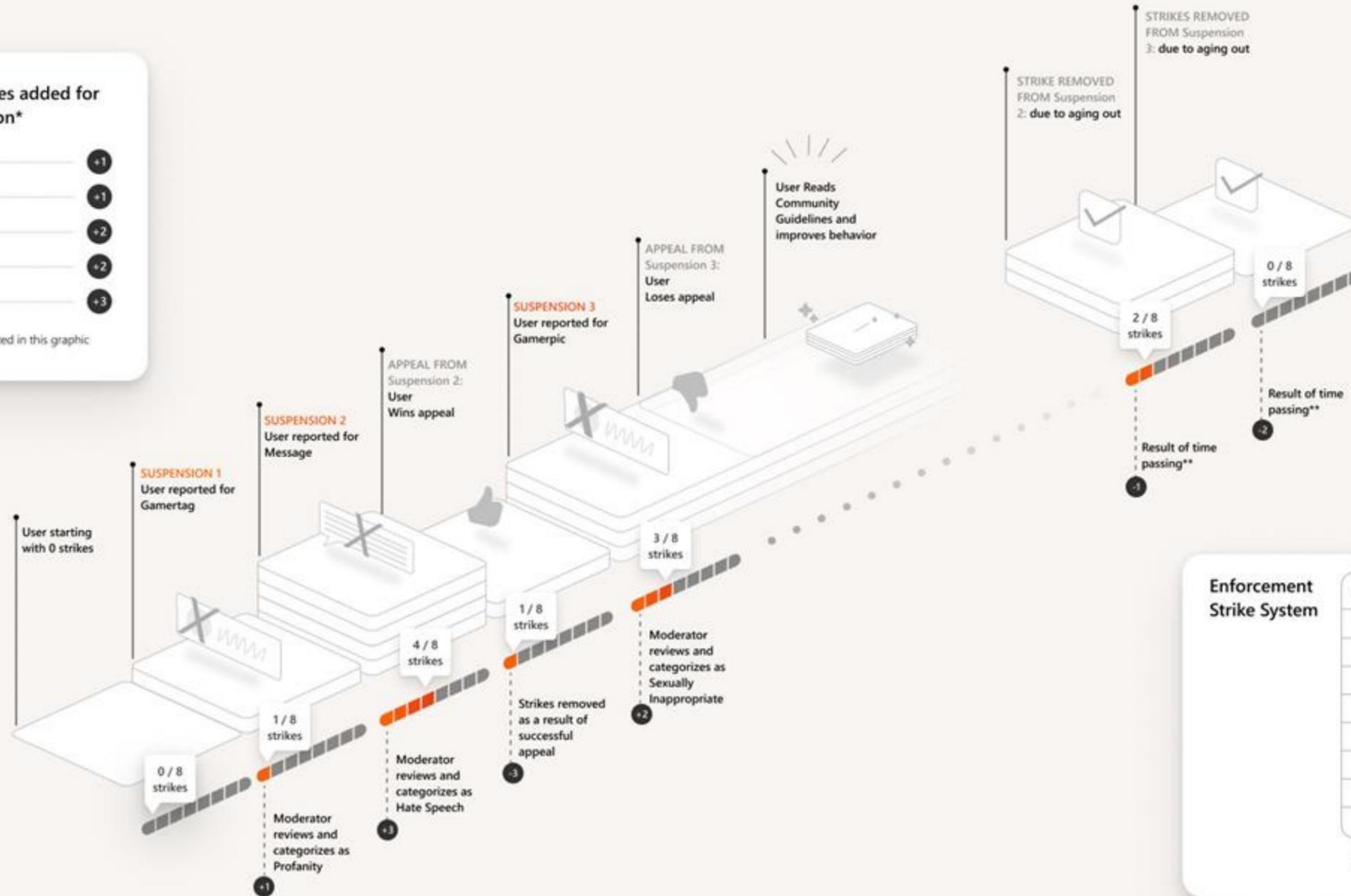
# SYSTÈME DE GRÈVE D'APPLICATION | INFOGRAPHIE DU PARCOURS UTILISATEUR



## Examples of strikes added for each type of action\*

Profanity	+1
Cheating	+1
Sexually Inappropriate	+2
Harassment or Bullying	+2
Hate Speech	+3

\*not all actions are represented in this graphic



## Enforcement Strike System

Strike	Suspension Length
1	1-day
2	1-day
3	3-days
4	7-days
5	14-days
6	21-days
7	60-days
8	365-days

\*\*All strikes stay on record for 6 months

## SYSTÈME DE GRÈVE D'APPLICATION | Questions et réponses

Pour améliorer nos mesures de sécurité et permettre aux joueurs d'interagir respectueusement les uns avec les autres, nous avons introduit un nouveau [système de frappe d'application](#). Le nouveau système applique un nombre variable de grèves à chaque application en fonction de la gravité de l'activité inappropriée. L'accumulation de grèves entraîne des délais d'exécution plus longs. Lorsque le nombre total de grèves d'un joueur atteint huit, il sera suspendu des fonctionnalités sociales de Xbox pendant un an à compter de la date d'application. Plus d'informations peuvent être trouvées dans la [FAQ sur le support Xbox](#).

Donc, vous donnez sept avertissements, puis une exécution ?

Certaines actions sont plus nocives

impact et ceux-ci reçoivent plus de frappes. En fonction de la gravité, un joueur pourrait très rapidement atteindre huit frappes en un plus petit nombre d'exécutions. Pour

Par exemple, si un joueur recevait une sanction pour discours de haine, cela entraînerait trois avertissements et une sanction de trois jours. Si le joueur choisit de ne pas améliorer son comportement et reçoit une deuxième sanction pour discours de haine, il recevra alors six avertissements et recevra une sanction pendant 21 jours.

Tous les rapports sont-ils simplement examinés par un processus automatisé d'IA ?

Notre système de modération de contenu

emploie à la fois des modérateurs automatisés et humains pour garantir que notre

les mesures coercitives sont exactes et équitable. Nos modérateurs humains jouent un rôle crucial dans les décisions d'application.

Nous travaillons constamment à améliorer la précision et l'efficacité de nos évaluations afin de fournir la meilleure expérience et les meilleurs résultats possibles à nos joueurs.

Si je reçois un avertissement, cela suivra-t-il mon compte pour toujours ?

Toutes les frappes reçues restent

dans le dossier d'un joueur pendant six mois.

Quelqu'un peut-il me faire suspendre simplement en demandant à d'autres joueurs de déposer un rapport contre moi ?

Non. Nous étudions les signalements des joueurs et prenons en compte chaque signalement ainsi que les informations associées.

preuve en considération lorsque prendre une décision d'exécution.

La fréquence ou le volume des rapports ne signifie pas que vous recevrez une sanction action.



[En savoir plus sur les types d'applications](#)



[FAQ sur le système de frappe d'exécution](#)



[FAQ sur les mesures coercitives](#)

# GLOSSAIRE DES TERMES



## GLOSSAIRE DES TERMES

Appels (examen de cas) – Un mécanisme par lequel un joueur qui a reçu une exécution peut trouver plus d'informations sur les circonstances et faire appel pour que l'exécution soit supprimée ou raccourcie.

Examen du cas – Voir appels

CSEAI – Images d'exploitation ou d'abus sexuels sur enfants

CyberTipline – Le système national de signalement centralisé de l'exploitation des enfants en ligne

DSCR (Digital Safety Content Report) – Un rapport semestriel publié par Microsoft qui couvre les problèmes de sécurité numérique. Trouvé ici \_\_\_\_\_

Application – Mesure prise à l'encontre d'un joueur, généralement sous la forme d'une suspension temporaire qui empêche le joueur d'utiliser certaines fonctionnalités du service Xbox.

Comptes non authentiques – Comptes jetables couramment utilisés à des fins telles que le spam, la fraude, la triche ou d'autres actions qui, en fin de compte, créent des règles du jeu inégales pour nos joueurs ou nuisent à leur expérience.

NCII – Images intimes non consensuelles

NCMEC – Centre national pour les enfants disparus et exploités

Non-réintégration – Lorsqu'un joueur fait appel d'une mesure d'exécution sur son compte et que l'exécution initiale s'est avérée justifiée.

Rapport de joueur – Lorsqu'un joueur dépose une plainte ou porte une violation de la politique à l'attention de l'équipe de sécurité.

Application proactive – Lorsque nous agissons sur un contenu ou un comportement inapproprié avant qu'un joueur ne le porte à notre attention.

Application réactive – Lorsque nous agissons sur un contenu ou un comportement inapproprié via un joueur qui le porte à notre attention.

Réintégration – Lorsqu'un joueur fait appel d'une application reçue et que son compte est rétabli (l'application est supprimée). Cela se produit généralement en raison d'une erreur, de circonstances atténuantes ou lorsque de la compassion est manifestée.

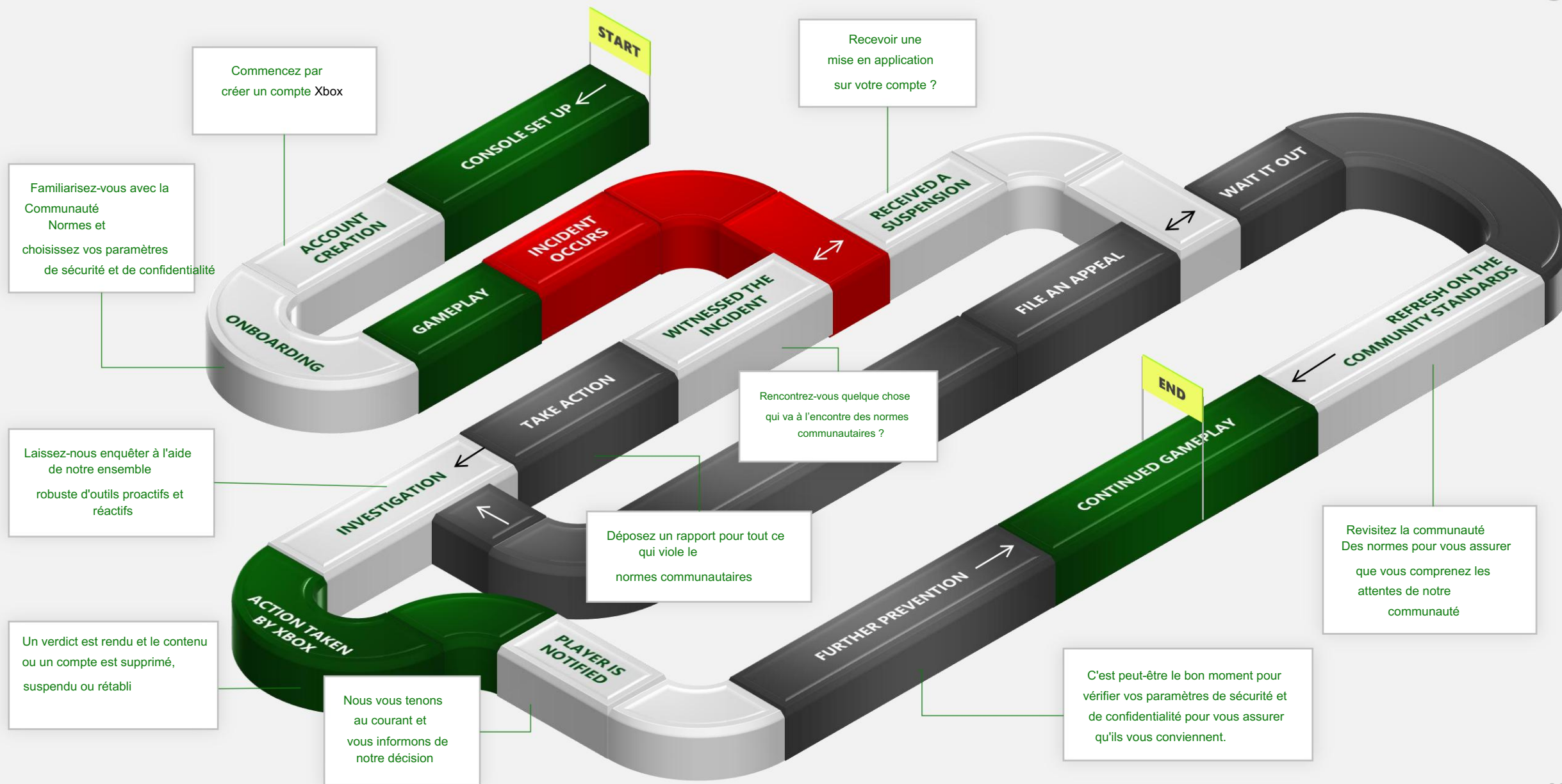
TVEC – Contenu terroriste et extrémiste violent

# ANNEXE





## Infographie du parcours du joueur



## INFOGRAPHIE DE TÉLÉCHARGEMENT D'IMAGE DU JOUEUR

