



**2024**

# **TRANSPARENCY REPORT**



# TABLE OF CONTENTS

<u>02</u>	<b>EXECUTIVE SUMMARY</b>
<u>03</u>	<b>KEY TAKEAWAYS</b>
<u>04</u>	<b>CONTENT MODERATION AT XGS</b>
<u>06</u>	<b>OUR VISION</b>
<u>09</u>	<b>OUR APPROACH</b>
<u>11</u>	Community Standards
<u>12</u>	Player Choice via Settings
<u>13</u>	Parental Controls
<u>14</u>	Enforcement
<u>14</u>	Microsoft Digital Safety Content Report
<u>15</u>	Help When Players Need It
<u>16</u>	Appeals / Case Reviews
<u>17</u>	Training and Assistance Program
<u>18</u>	<b>SHARING OUR SAFETY DATA</b>
<u>19</u>	Methodology
<u>20</u>	Proactive Detection: Scanning
<u>21</u>	Proactive Detection: Restrictions Imposed
<u>22</u>	Reactive Moderation/Report a Concern
<u>23</u>	Report a Concern – Restrictions Imposed
<u>24</u>	Automated Means
<u>25</u>	Microsoft Digital Safety Content Report Data
<u>25</u>	Crisis Text Line Data
<u>26</u>	Appeals Data
<u>27</u>	<b>POLICIES AND PRACTICES</b>
<u>29</u>	<b>GLOSSARY OF TERMS</b>
<u>31</u>	<b>APPENDIX</b>
<u>32</u>	Player Journey Infographic
<u>33</u>	Player Image Upload Infographic
<u>34</u>	Enforcement Strike System Infographic

# EXECUTIVE SUMMARY



## EXECUTIVE SUMMARY

At Xbox, our mission is to bring the joy and community of gaming to everyone on the planet.

Our players deserve the opportunity to experience a place free from fear and intimidation, safe within the boundaries they set.

Our games and services are used by millions of players worldwide to engage in a broad range of gaming experiences, explore digital worlds through play, and connect to the player communities they love. As our Gaming family of products and services grows, this report will grow to give our players the full picture of our commitment to safety. For the first time, this 2024 report covers the full year of content moderation activities of both the Xbox platform and Xbox Game Studios. While the core values of safety, inclusivity, and transparency remain consistent across our games and studios, the specific policies and practices may vary across studios to best suit the unique needs of their players. As always, we are deeply committed to working with industry partners, regulators, and our community to improve our multifaceted approach to safety. We will continue to learn, iterate, and be transparent about our approach in keeping the Xbox community safe and fun for everyone.



## EXECUTIVE SUMMARY

### Key takeaways from the report

#### 01 Combined Impact

Previous transparency reports reported on content moderation activities of the Xbox platform. This combined 2024 report includes the activities of both Xbox platform and our Xbox Game Studios. Although moderation policies and practices may vary to align to the unique communities of each game, we are unified in the values of our Community Standards.

These values are key to sustaining a vibrant and welcoming community.

In 2024, we also more closely aligned with our Microsoft Digital Safety harm definitions in a combined effort to keep our communities safe and welcoming.

#### 02 Preventing Unwanted Content

To protect players from being exposed to unwanted content, we proactively scan user-generated content, including text, usernames, videos, and images. This ensures that the content avoids harmful behavior, respects others' rights, and maintains privacy. From February 17 to December 31, 2024, **17.2 billion pieces of content - across games and platforms - were scanned globally**. Of these, 2.4% were identified as harmful or violating policy. The majority of violations were due to abuse of platform and services, which means the content breached our terms and conditions. Proactive moderation helps maintain a safe and respectful environment by preventing unwanted content from ever reaching players in the first place.

#### 03 Improved Harm Prevention with Xbox Automod

We prioritize player safety through advanced automation and AI-enabled tools that is supported by human expertise. In our previous transparency report, we introduced our Xbox Automod system that launched in 2024 and assists in the moderation of reported Spam and Advertising content. So far, Xbox Automod has handled **4.4M pieces of reported content** and enabled the team to **act on unwanted content 93% faster**. We are continuing to invest in the system.

## EXECUTIVE SUMMARY

### Content Moderation at Xbox

At Xbox, we ensure a safe and welcoming environment for players by detecting and restricting unwanted content, and by empowering players -the stewards of our gaming community- with useful reporting tools to escalate any issues they might encounter online.

The [Xbox Community Standards](#) provide a comprehensive framework for maintaining a positive and inclusive gaming environment. It sets clear expectations for player behavior, emphasizing the importance of respect, fairness, and safety. These standards include guidelines on keeping content appropriate, avoiding harmful behavior, respecting others' rights, and maintaining privacy. They also highlight the importance of contributing positively to the community and being a force for good.

We apply content and account restrictions in line with our Xbox Community Standards on text, usernames, images, and other user generated content. This includes the detection and removal of harmful content, temporary or permanent actions on user accounts, and the use of labels or warnings.

Our moderation teams apply proportional and appropriate enforcement actions **while providing players with opportunities to modify their behavior**. The approach helps prevent repeated unwanted behavior and informs users about the nature of the content. They do this with the help of automated tools to meet the demands of an ever-growing gaming and technology landscape.



## EXECUTIVE SUMMARY

### Fostering Safe Gaming Experiences for All

While the core values of safety, inclusivity, and transparency remain consistent across all our games and studios, the specific policies and practices may vary to best suit the unique needs of our diverse audiences. For mature audiences, that means the presence of more complex themes and content, and for younger audiences that means implementing stringent content controls to ensure gameplay is age appropriate.



# OUR VISION





## OUR VISION

The Xbox community is yours.

We all bring something unique, and that uniqueness is worth protecting.

Whether you are new to gaming or have been playing for decades, you are stewards of this place, protecting each other even as you compete.

**Because when everyone plays, we all win.**



## OUR VISION

Our [Xbox Community Standards](#) outline the conduct and content that are acceptable within our community. We acknowledge that negative activity can and has taken place. This conduct is not okay and goes against the community we strive to create – a place that is vibrant, safe, and welcoming.

We want you to feel confident that we are listening and acting upon your feedback – we use that feedback to test and implement new features and better understand the activity and conduct of our players. One way to help us deliver the best gaming experience possible is to [provide feedback](#) and by taking part in our [Xbox Insider Program](#).



# OUR APPROACH



## OUR APPROACH

### Our multifaceted approach

Xbox Platform and Xbox Game Studios have a multi-faceted approach:

- Working to create a strong community of gamers who are thoughtful about their conduct and guided by comprehensive [Community Standards](#)
- Using proactive technology and tools to detect and remove problematic content before it is seen and to reduce conduct that runs counter to our Community Standards
- Enabling useful [reporting tools](#) for our players to identify issues
- An [Appeals](#) process to educate our users about the Community Standards
- Continued learning and investment in our safety measures

In addition, Xbox Platform:

- Gives players controls to customize their settings across the entire Xbox ecosystem from console to PC to Xbox Cloud Gaming (Beta), including comprehensive [parental controls](#) so children can engage in safer experiences that are appropriate for them
- Helps players better understand their enforcements and prevent them from repeating with the [enforcement strike system](#).

---

⇒ [Learn about our shared commitment to safer gaming](#)

Protecting our community requires constant work and diligence. Our foundational approach to safety-by-design and a dedicated team ensures safety is, and will always be, a priority for everyone.



## OUR APPROACH



## Community Standards

The [Microsoft Services Agreement's](#) Code of Conduct section applies to Xbox and its players. Our [Xbox Community Standards](#) offers an additional level of explanation, providing specifics on our expectations for player conduct on our network. They also reflect the policies we have in place to moderate conduct and, when necessary, impose consequences for players that violate our policies. Each game's Code of Conduct or Community Standards hold players to the Xbox Community Standards.

---

⇒ [Learn about the Xbox Community Standards](#)



## OUR APPROACH

### Enforcement

When a player's conduct or content has been found to violate our policies, the content moderation agents or systems will take action - we call this an enforcement. Most often this comes in the form of removing the offending content from the service and issuing the associated account a suspension.

The length of suspension is primarily based on the type of offending conduct or content while taking into consideration the account's previous history. Repeated violations of the policies result in lengthier suspensions and can culminate in a 12-month suspension of social features. Particularly egregious violations can result in permanent account suspensions or device bans.

In 2023, we introduced a [new strike system](#) to our Enforcement approach designed to better educate players about enforcements and to further empower players to engage positively and appropriately on Xbox and with the community.

---

⇒ [Learn about types of enforcements](#)

⇒ [Enforcement strike system FAQ](#)

⇒ [Enforcement action FAQ](#)

### Microsoft Digital Safety Content Report

For several years, Microsoft has published a bi-annual [Digital Safety Content Report \(DSCR\)](#), which covers actions Microsoft has taken against terrorist and violent extremist content ([TVEC](#)), non-consensual intimate imagery ([NCII](#)), child sexual exploitation and abuse imagery ([CSEAI](#)), and grooming of children for sexual purposes across its consumer services, including Xbox.

At Xbox, violations of our CSEAI, grooming of children for sexual purposes, or TVEC policies will result in removal of the content and a permanent suspension to the account, even if it is a first offense. These types of cases, along with threats to life (self, others, public) and other imminent harms are immediately investigated and escalated to law enforcement, as necessary.

---

⇒ [Learn about the Digital Safety Content Report \(DSCR\)](#)



## OUR APPROACH

### Appeals / Case Reviews

Our [appeals](#) process enables a player to get more information about any enforcements they have received including account suspensions or content removals. A player can launch an appeal, otherwise known as a case review, to provide us with more information if they disagree with our determination that a policy was violated. Based on the appeal, the original decision may be confirmed, modified, or overturned and the account reinstated.

---

⇒ [How to file a case review](#)

⇒ [Learn about types of enforcements](#)

⇒ [Enforcement action FAQ](#)

⇒ [Enforcement Strike System FAQ](#)

## OUR APPROACH

### Help When Players Need It

We also look to help our players when they need it. If a player's communications are flagged as concerning (including content associated with suicidal ideation or self-harm), either by our system or by other players, we may provide Crisis Text Line, or regional equivalent, information to the player so they can reach out to resources who can help.

Crisis Text Line is a US-based nonprofit organization that [Xbox has been partnering with since 2018](#), which provides free, text-based 24/7 support. Texting Crisis Text Line from anywhere in the US, Canada, UK, and Ireland allows people to connect with trained and confidential crisis counselors who are dedicated to helping others stay safe and healthy.

Outside of these countries, we partner with a number of trusted [mental health resources for the Xbox community](#) around the globe.





## OUR APPROACH

### Player Choice via Settings

We know that when it comes to preferences on content and experiences, it is not one-size-fits-all. Content or language that is fine for one player may not be suitable for others.

We offer our players choices about the types of content they want to see and experience on our network, which include:

- [Automated text, media and web link filtration](#) so you can decide what text-based messages you would be comfortable receiving
- [Filter flexibility](#), allowing players to configure safety settings along a spectrum from most filtered to least so you can choose what is best for you
- Customizable [parental controls](#), including a convenient [Xbox Family Settings App](#) on mobile devices
- [Mute and block](#) other players and their messages
- [Real name sharing](#) if players want to share their real name with friends
- [Customizable privacy and notification settings](#) that let players manage their social experience.

Players have the opportunity to adjust and select their privacy and safety settings at any time, with those settings being effective across all the ways players access the Xbox platform.

---

⇒ [Learn about safety settings for Xbox messages](#)

⇒ [Learn about managing Xbox safety and privacy settings](#)



## OUR APPROACH

### Parental Controls

Xbox offers a robust set of [parental controls](#) on Xbox Series X|S, Xbox One, Windows, and Xbox mobile apps that help children on our platform have safer experiences on our services. A convenient [Xbox Family Settings App](#) for mobile devices helps parent manage those settings across those devices. Child accounts on Xbox come with default settings that prevent children from viewing or playing games that have mature ratings and require parental permission for other actions such as playing multiplayer games, chatting with other players, and making purchases. Parents can also receive [weekly activity reports](#) about their children's time on Xbox, including games played, time spent on each game, and purchases made.

We care deeply about what our Xbox Community wants. That is why we've continued to add to our capabilities since the debut of our Xbox Family Settings App. Because of direct feedback from parents of gamers, we've added more options to [prevent unauthorized purchases](#) and the ability for caregivers to [set good screen time habits](#). These options also help spark conversation between parents and children to help younger players build stronger digital skills and safely navigate their online presence.

---

⇒ [Download the Xbox Family Settings app](#)

⇒ [Learn more about Parental Controls](#)

⇒ [Learn more about the Xbox Family Settings App](#)

## OUR APPROACH

### Training and Assistance Program

Xbox has implemented comprehensive training programs for content moderators to ensure they are well-equipped to handle complex and nuanced issues. This includes training on the use of moderation tools and deep understanding of the Xbox Community Standards.

There are ongoing efforts to provide support and resources to moderators to help them manage the emotional and psychological challenges of their work as well. This includes regular workshops, mental health resources, and access to professional support services.



# **SHARING OUR SAFETY DATA**



# Methodology

The data we'll be sharing in this section:

- Covers the period between **Feb 17–Dec 31, 2024**.
- Is from the global perspective across Xbox platform and Xbox Game Studios unless otherwise noted.
- In addition to the Xbox platform, participating Xbox Game Studios include Age of Empires, Gears of War, Forza, Ara: History Untold, Minecraft, Microsoft Flight Simulator, Sea of Thieves, Halo, and Grounded.
- In this edition, we have adopted the Microsoft Digital Safety harm definitions.

Data was collected in accordance with Microsoft's commitment to privacy.

SHARING OUR SAFETY DATA

Proactive Detection: Scanning

To reduce the risk of harm and prevent our players from being exposed to inappropriate content, we proactively scan content to identify and remove unwanted content before it impacts players.

We apply these methods to scan text, usernames, images, and other user generated content to ensure it avoids harmful behavior, respects others' rights, and maintains privacy.

For example, proactive moderation allows us to find and remove inauthentic accounts and other cases of abuse of our platform and services.

How much content was moderated?

17.2B

How much was harmful?

2.4%  
(409M)

17.2B pieces of content globally across text, usernames, images, and other user generated content were proactively scanned from February 17 to December 31, 2024. 2.4% of this content was identified as containing harmful or policy-violating content.

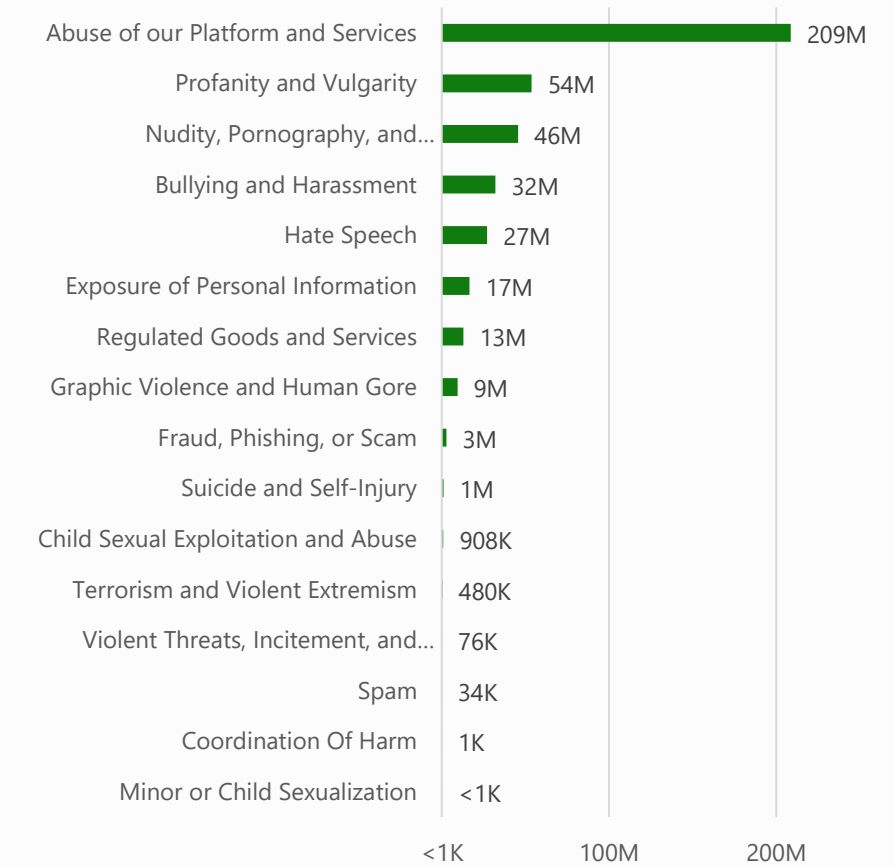
SHARING OUR SAFETY DATA

Proactive Detection: Restrictions Imposed

The figure shows the number of restrictions imposed on content or accounts by our proactive moderation efforts due to the content being harmful or in violation of policy.

The majority were identified as being an *Abuse of our Platform and Services* which means the content was in violation of our terms and conditions. *Profanity and Vulgar*ity, *Nudity Pornography and Inappropriate Content*, *Bullying and Harassment*, and *Hate Speech* also represented a sizable number of restrictions.

What harm types and policy violations lead to restrictions imposed on content or accounts?



Globally, 397M restrictions were imposed on content and 15M restrictions were imposed on accounts as a result of proactive detection during the period from February 17 to December 31, 2024.



## SHARING OUR SAFETY DATA

### Reactive Moderation/Report a Concern

Proactive blocking and filtering are only one part of the process in reducing toxicity on our service. Xbox offers robust reporting features, in addition to [privacy and safety controls](#) and the ability to [mute and block](#) other players; however, inappropriate content can make it through the systems and to a player.

Reactive moderation is any moderation and review of content that a [player reports to Xbox](#) via the web or reporting tools on the device, as applicable. When a player reports another player, a message, or other content on the service, the report is logged and sent to our moderation platform for review by content moderation technologies and human agents. These reactive reports are reviewed and acted upon according to the relevant policies that apply.

As stewards of our community, player use of our reporting features continues to be a crucial way in which communities are kept fun, safe, and welcoming.

How many reports did we receive?

**53.2M**

How many lead to action?

**9.2%**  
(5M)

---

53.2M reports were received from players globally during the period from February 17 to December 31, 2024. 9.2% of those reports lead to restrictions imposed on content and accounts.



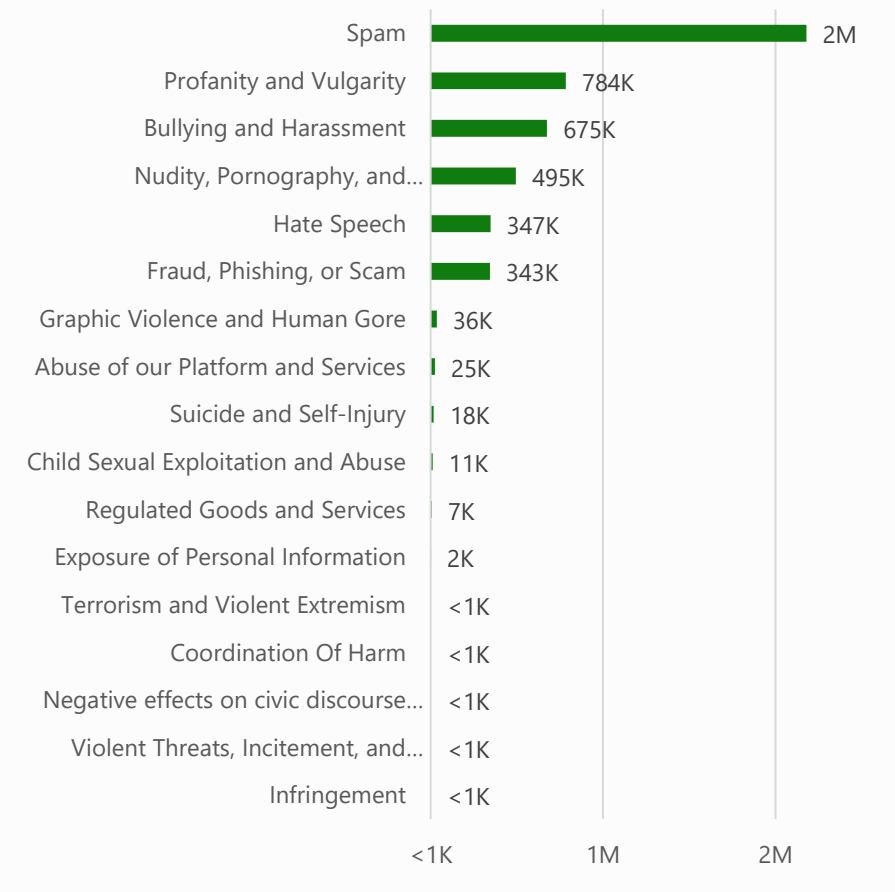
SHARING OUR SAFETY DATA

Report a Concern – Restrictions Imposed

The figure shows restrictions imposed on content and accounts that were reported by players.

The majority of these actions were identified as being *Spam*. *Profanity and Vulgarities*, *Bullying and Harassment*, *Nudity Pornography and Inappropriate Content*, *Hate Speech*, and *Fraud, Phishing, or Scam* also represented a sizable number of restrictions imposed.

What harm types and policy violations lead to restrictions imposed on content or accounts?



Globally, 5M restrictions were imposed on content and 5M restrictions were imposed on accounts as a result of reports during the period from February 17 to December 31, 2024.

SHARING OUR SAFETY DATA

Automation Assisting Humans

With the gaming and technology landscape evolving and our community of players growing, not just on Xbox consoles but across platforms, our solutions to keep players safe need to continue to evolve to ensure safety across all the ways players engage with each other.

At Xbox, we believe automations and the use of AI-enabled solutions such as Community Sift, combined with human expertise, play crucial and complementary roles in effectively identifying, reporting, and preventing harms at scale, especially as these online harms become more technologically sophisticated. **They not only prevent unwanted content from reaching players, they also reduce human exposure to sensitive content and help focus human moderation efforts on more nuanced and complex issues.**

We have a number of safeguards in place to monitor automations. This includes a robust quality assurance program, moderator reviews, a suite of player reporting tools, and an appeals process.

Our foundation is a safety system that relies on both players and the expertise of human moderators to ensure the consistent and fair application of Xbox Community Standards.

How much content was...

	Global
Reported by users and processed manually?	1M
Reported by users and auto-processed?	4M
Proactively scanned and processed manually?	12M
Proactively scanned and auto-processed?	400M

79% of reports and 97% of proactive content were processed using automated systems. This data is representing the period from February 17 to December 31, 2024.

## SHARING OUR SAFETY DATA

### Microsoft Digital Safety Content Report Data

As a US-based company, Microsoft reports all apparent Child Sexual Exploitation or Abuse Imagery ([CSEAI](#)) or grooming of children for sexual purposes to the National Center for Missing and Exploited Children ([NCMEC](#)) via the [CyberTipline](#), as required by US law.

In the period covered by this report, **1,254** of Microsoft's reports were from Xbox.

More information on Microsoft's efforts regarding CSEAI, grooming of children for sexual purposes, and terrorist and violent extremist content ([TVEC](#)) can be found in the [Digital Safety Content Report](#).

### Crisis Text Line Data

The most common real-world concerns that we see on the platform have to do with threats of self-harm, which are handled with a referral to counseling services via the [Crisis Text Line](#), or international equivalent partner.

In the period covered by this report, we sent **12,495** referral messages to players.

SHARING OUR SAFETY DATA

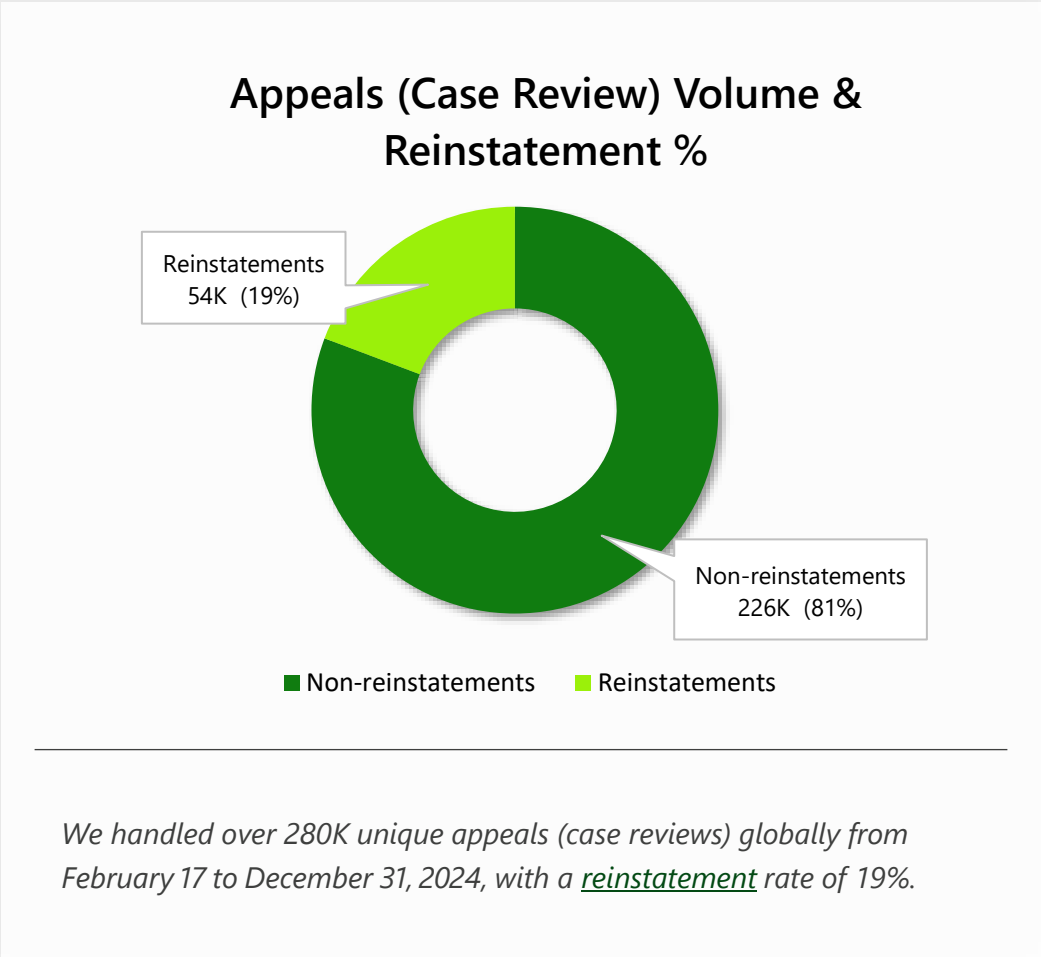
Appeals (Case Review) Data

When a player receives an enforcement beyond a certain length of time, they can dispute or ask for clarification through an appeal, otherwise known as a case review.

When filing a case review, the player can explain their actions and a moderation agent will review the case to see if an error was made or if special reconsideration is warranted.

Reinstatements are issued when an error is uncovered or if the player deserves reconsideration specific to their enforcement. A non-reinstatement is when the original enforcement action was found to be warranted and upheld after review.

Here we look at the volume of appeals handled and the associated percentage of accounts that were reinstated:



# **POLICIES AND PRACTICES**



## POLICIES AND PRACTICES

Here is some supplemental information that may help you better understand the content of this report:



### Policy & Standards

- [Xbox Community Standards](#)
- [Microsoft Services Agreement](#)



### Reporting Process

- [How to report a player](#)



### Appeals Process (Case Review)

- [How to submit a case review](#)



### Glossary of Definitions

- [Definitions](#)



### Additional Resources

- [Family & Online Safety](#)
- [Privacy & Online Safety](#)
- [Parental Controls](#)
- [Family Hub](#)
- [Responsible Gaming for All](#)
- [Learn about the Xbox Family Settings app](#)
- [Learn about safety settings for Xbox messaging](#)
- [Xbox Family Settings app](#)
- [Xbox Insiders Program](#)
- [Privacy dashboard](#)
- [Enforcement Strike System FAQ](#)
- [Reactive Voice Reporting](#)

# **GLOSSARY OF TERMS**



## GLOSSARY OF TERMS

**Appeals (Case Review)** – A mechanism through which a player that received an enforcement can find out more information as to the circumstances and appeal to have the enforcement removed or shortened

**Case Review** – See Appeals

**CSEAI** – Child Sexual Exploitation or Abuse Imagery

**CyberTipline** – The centralized reporting system for the online exploitation of children

**DSCR (Digital Safety Content Report)** – A half yearly report published by Microsoft that covers digital safety concerns. Found [here](#)

**Enforcement** – Action taken against a player, usually in the form of a temporary suspension which prevents the player from using certain features of the Xbox service

**Inauthentic accounts** – Throwaway accounts that are commonly used for purposes such as spam, fraud, cheating, or other actions that ultimately create an unlevel playing field for our players or detract from their experiences

**NCII** – Non-consensual intimate imagery

**NCMEC** – National Center for Missing & Exploited Children

**Non-reinstatement** – When a player appeals an enforcement action on their account and the original enforcement was found to be warranted

**Player Report** – When a player files a complaint or brings a policy violation to the attention of the Safety Team

**Proactive Enforcement** – When we action on inappropriate content or conduct before a player brings it to our attention

**Reactive Enforcement** – When we action on inappropriate content or conduct via a player bringing it to our attention

**Reinstatement** – When a player appeals a received enforcement, and their account is reinstated (enforcement is removed). This usually occurs due to an error, extenuating circumstances, or when compassion is shown

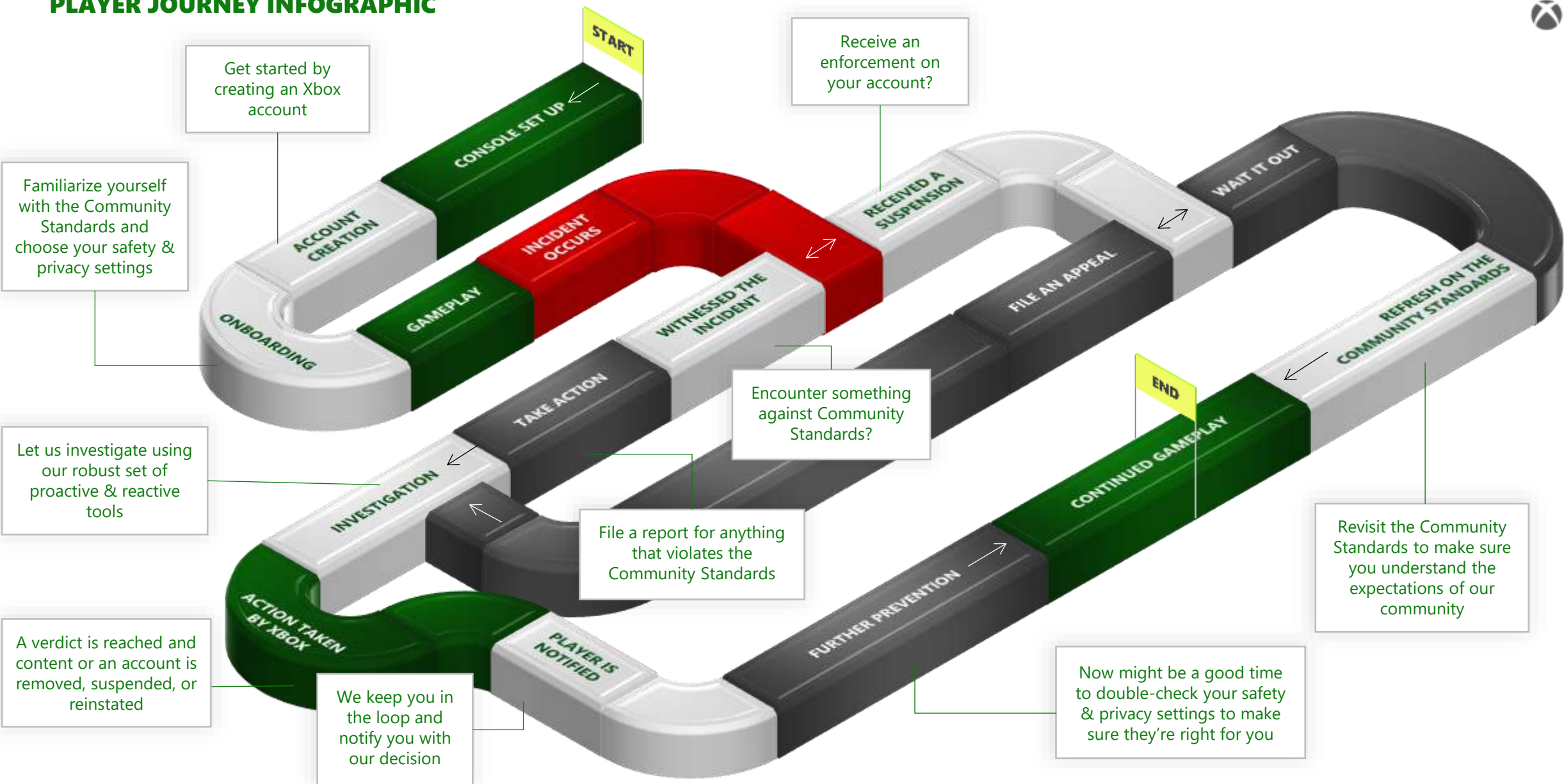
**TVEC** – Terrorist and Violent Extremist Content



# APPENDIX

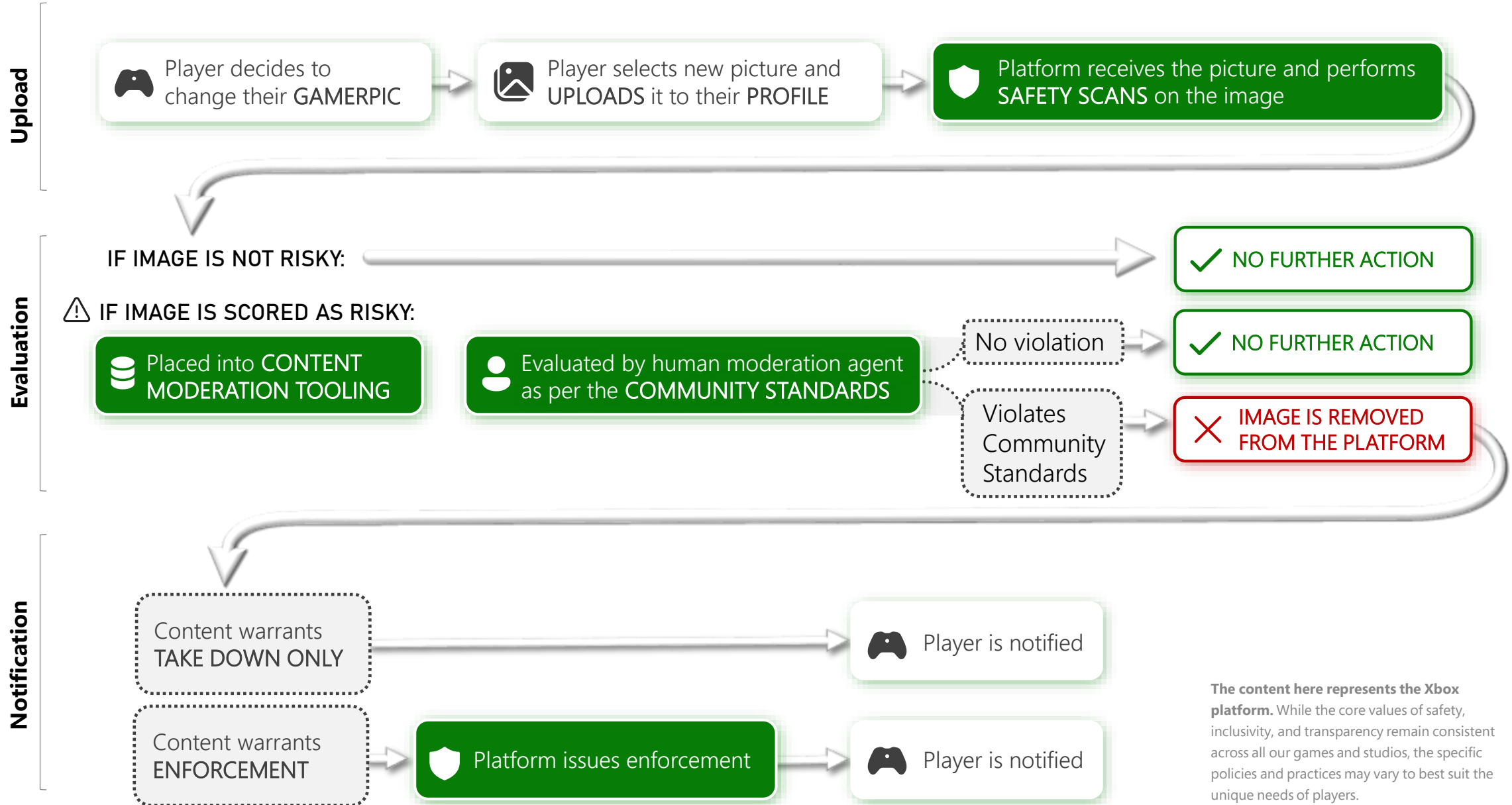


# PLAYER JOURNEY INFOGRAPHIC



The content here represents the Xbox platform. While the core values of safety, inclusivity, and transparency remain consistent across all our games and studios, the specific policies and practices may vary to best suit the unique needs of players.

## PLAYER IMAGE UPLOAD INFOGRAPHIC



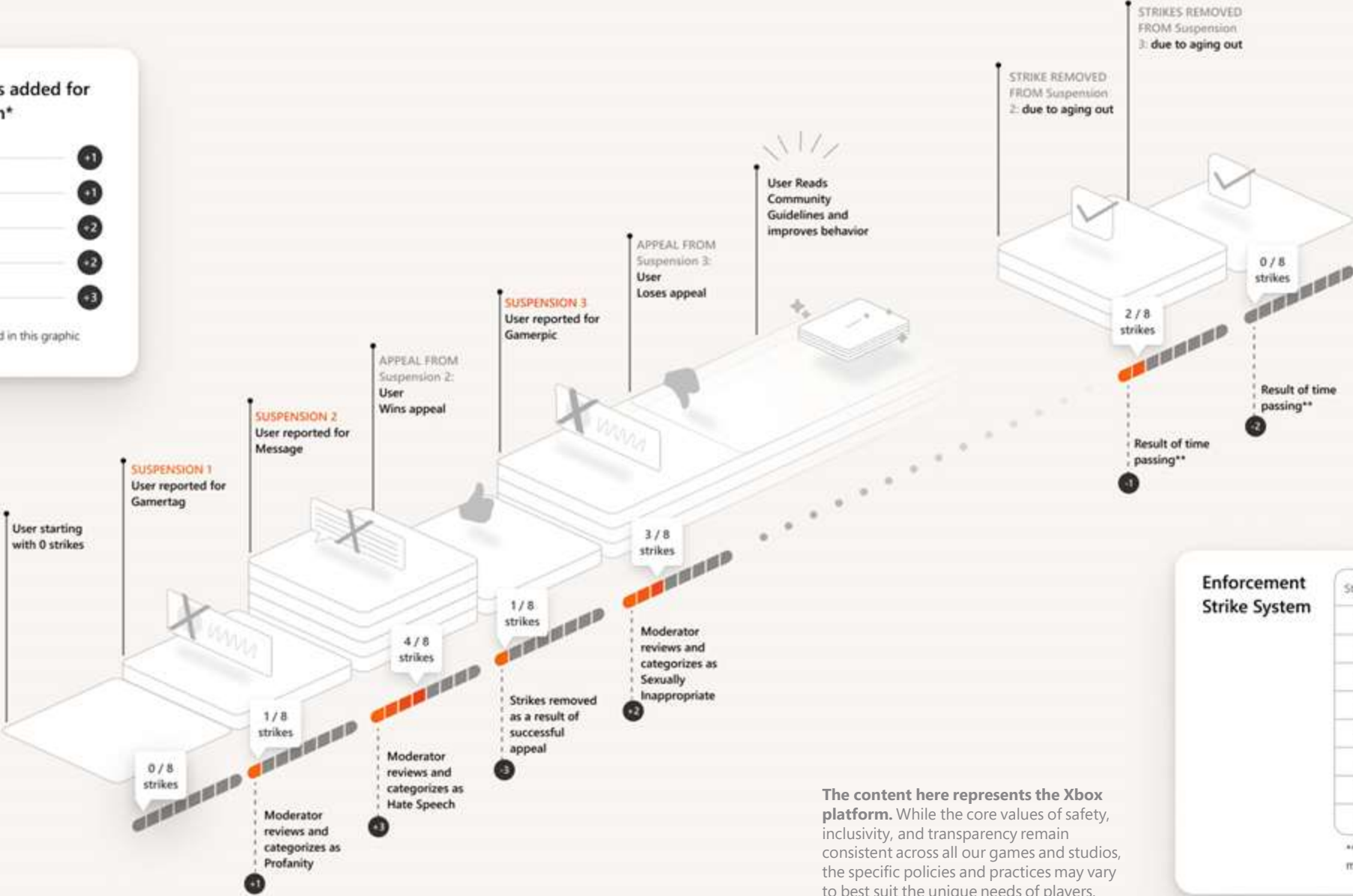
# ENFORCEMENT STRIKE SYSTEM | USER JOURNEY INFOGRAPHIC



## 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



The content here represents the Xbox platform. While the core values of safety, inclusivity, and transparency remain consistent across all our games and studios, the specific policies and practices may vary to best suit the unique needs of players.

## 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