Sony Interactive Entertainment

PlayStation

Safety Report

February 2025

Introduction

At Sony Interactive Entertainment (SIE), the company behind PlayStation, we aim to provide innovative experiences to all players. With over 100 million monthly active users on PlayStation Network across the globe, our cross-functional safety team works tirelessly to deliver an environment that is welcoming and supportive to players from all backgrounds.

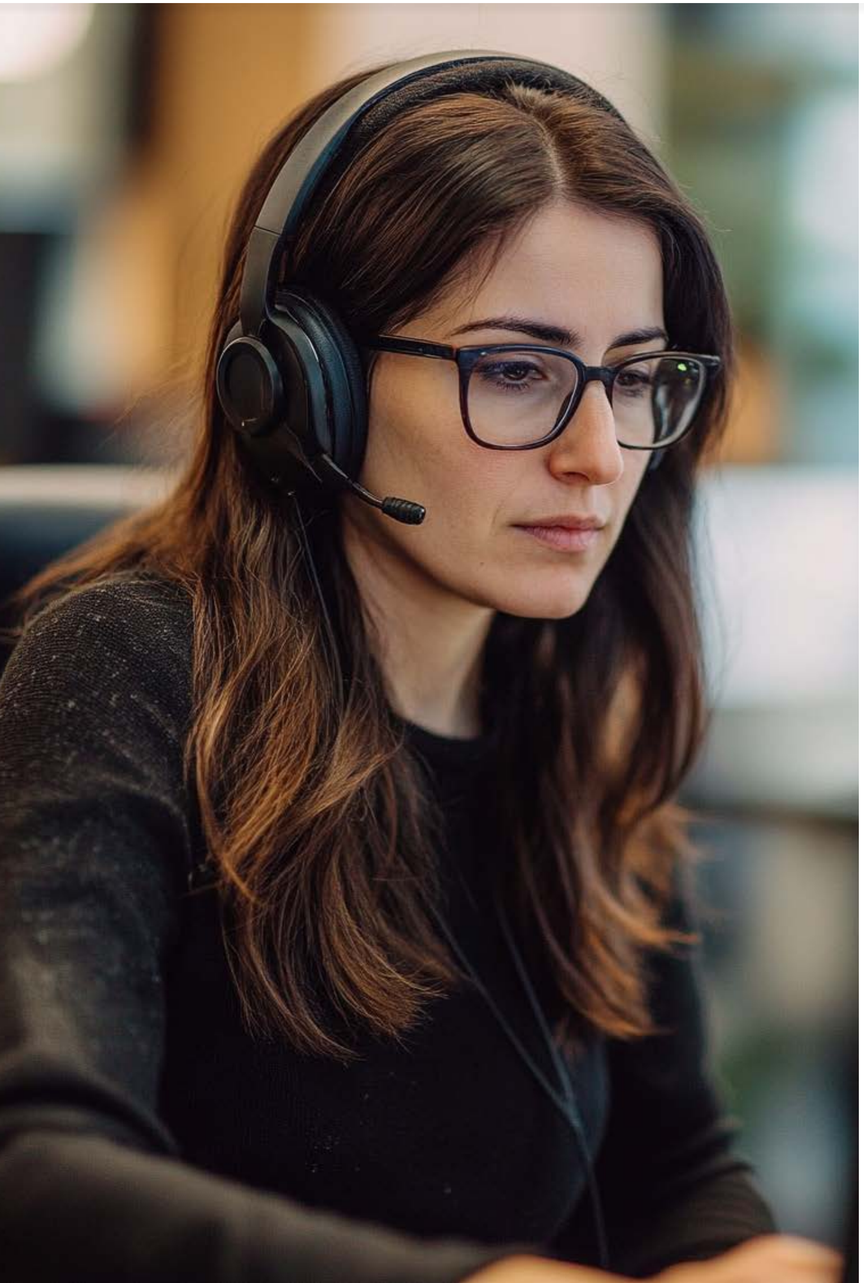
*The Sony Group Code of Conduct outlines our commitment to customer safety, as we strive to inspire a world filled with emotion (**Kando**) for our gaming community. Part of that commitment is to provide information that is accurate and easy to understand. This report introduces the three safety pillars that underpin our safety philosophy for keeping our players safe: Control Your Experience, Shield Our Players, and Enforce Our Standards.*

PlayStation is for everyone. We are grateful for the support of our players and their shared commitment to creating and maintaining a safe and inclusive online community for all.

Catherine Jensen
VP, Global Consumer Experience
Sony Interactive Entertainment

Learn more

Sony Group Code of Conduct



Our Safety Philosophy is Built on Three Pillars:

- **Control Your Experience**: We have built tools that empower players and parents of players, to tailor their online experience. These tools encompass everything from customizing communications and play time to ensuring privacy and account security. Our family management features ensure parents and guardians can make informed decisions for their families.

- **Shield Our Players**: We leverage advanced technology to proactively protect players from a wide range of online harms including violative URLs, offensive text and imagery, exploitative language, and serious threats such as child sexual exploitation and abuse (CSEA). Our proactive approach minimizes exposure to harmful content before it can impact the player experience.

- **Enforce Our Standards**: We use skilled human moderators to ensure players adhere to our Code of Conduct and that appropriate action is taken against those who violate it. We use proactive reporting from players and reporting by our automated tools coupled with human moderator review to identify content that violates the code. Our policies are designed carefully to balance freedom of expression with the need to protect players from harmful content, fostering fairness, clarity, and consistency in enforcement.

Control Your Experience

Empowering players and family managers to make the right decisions for themselves and their families.

Players can control their experiences on PlayStation Network through features such as ‘blocking,’ ‘muting,’ and ‘reporting’ when they experience negative interactions. We also listened to player feedback and implemented solutions to shield players further. For example, we enabled a ‘block by default’ function in early 2024 to prevent a negative experience from reoccurring when the ‘reporting’ function was used.

We are actively testing and exploring new technologies and policies to enhance safety on PlayStation Network, including piloting an age assurance process for players who register new accounts in select countries.

We will continue to innovate and advance our online safety solutions to promote a safe and friendly online community for all.

Learn more

Block Players on PlayStation Network

Learn more

PS5 Party Voice Chat

Learn more

Report Behavior on PlayStation Network

Shield Our Players

Features and technology that protect players.

We want players to experience a sense of joy and community whenever they log in to play. Every player deserves to feel safe and confident to express their authentic selves while gaming. At SIE, we employ a range of safety measures to help keep our player experiences positive and fun. One of our most visible measures is the 'Text Profanity Filter,' which blocks hateful or profane language in many different languages in areas visible to other players, such as Online ID and About Me bios, helping to prevent harmful content from reaching other players.

We also employ less visible but equally important protections, such as advanced spam-blocking. We use machine learning tools to detect when bad actors are using technology to send harmful messages in bulk to our players. To further safeguard our community, we use our own bots to detect and remove the bots of others wishing to cause our players disruption.

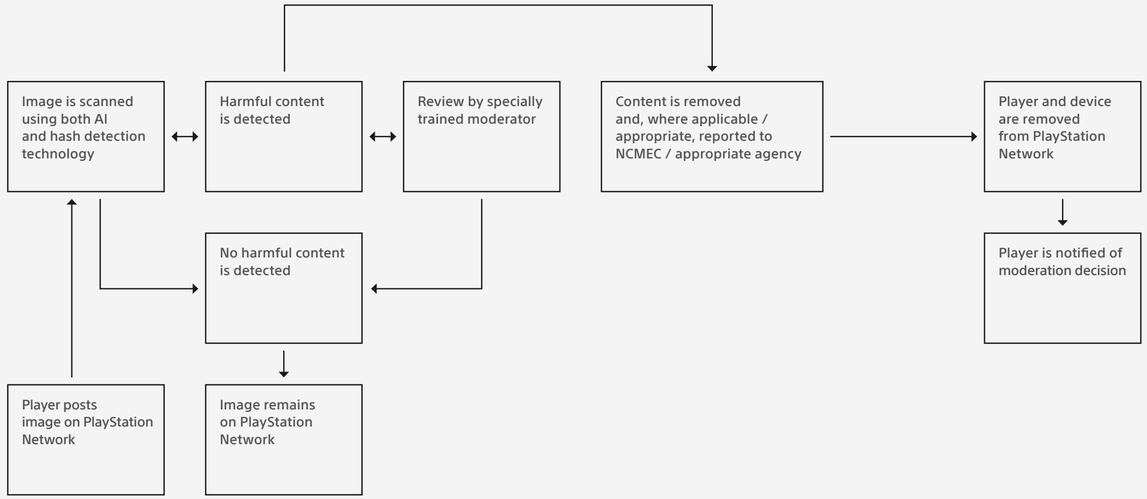
While we allow the sharing of links on PlayStation Network to encourage open communication and the sharing of gaming-related content, we remain vigilant. URLS that violate our Code of Conduct are actively identified and restricted, and help ensure platform remains safe and respectful for all. URLs that violate our Code of Conduct are added to a block list, and from that point forward, are no longer viewable on PlayStation Network. Sites that link hate, pornography, or extreme violence breach the Code of Conduct on PlayStation Network and are strictly prohibited.

Playing with, and sometimes against, others is a cornerstone of online gaming. We believe that players can experience tremendous satisfaction as their skill levels increase by overcoming difficult challenges within games. Players welcome healthy competition, but no player wants to be matched with someone using cheating software to give them an unfair advantage. Our systems proactively detect the use of cheating technologies, and we may suspend accounts found using such tools to help preserve a fair and balanced gaming environment.

All players are welcome on PlayStation to play individually or with friends. Player safety is a top priority, and we take additional measures to protect children on the PlayStation Network.

We collaborate with globally respected organizations, including the Family Online Safety Institute (FOSI), the National Center for Missing and Exploited Children (NCMEC), the Tech Coalition, and other established organizations to encourage player safety.

Image Detection Flow



This ranges from activities such as promoting youth online safety and good digital citizenship with FOSI to developing tools to detect and remove harmful content, including CSEA material with NCMEC and the Tech Coalition.

To safeguard our community, every image uploaded to PlayStation Network is scanned using hash-matching technology to flag potential matches of harmful images. Hash-matching tools assign a unique digital signature to content, called a 'hash' and then compare the hash against a database of known signatures. If a match is flagged, the content is escalated to a specialized moderation team for review. Confirmed cases of child sexual abuse material (CSAM) are reported directly to NCMEC. In addition to hash-matching, we employ machine learning technologies capable of identifying previously unreported CSAM. While the detection mechanisms differ, the moderation process remains the same.

Given the critical nature of CSAM detection, all flagged content undergoes thorough human review by trained specialists. When violations are confirmed, both the offending account and associated PlayStation consoles are permanently suspended from our network.

Enforce our Standards

Skilled human moderation, player communication, and appeals.

Moderation involves assessing user-reported content on PlayStation Network against our Code of Conduct and applying appropriate enforcement measures when violations are confirmed. Our human moderation teams, extensively trained in content evaluation, operate around the clock, every day of the year. We actively moderate reports related to messages, voice chat, profile content, Game Base group names, Online IDs, and in-game communications upon receipt.

Our moderation services cover over 30 languages, carefully considering cultural context, historical sensitivities, and local laws to ensure fair and respectful enforcement. We take all reports seriously and all actions taken on reported accounts are done only after the content has been verified by a human moderator.

We clearly communicate enforcement actions and proactively educate our players on the impact of negative behavior within PlayStation Network, including the potential consequences such as temporary or permanent suspensions. We also offer players an option to make an appeal against a permanent suspension in 40 countries where PlayStation Network is available with plans to extend the appeal process to temporary suspensions and additional regions in the future.

Learn more

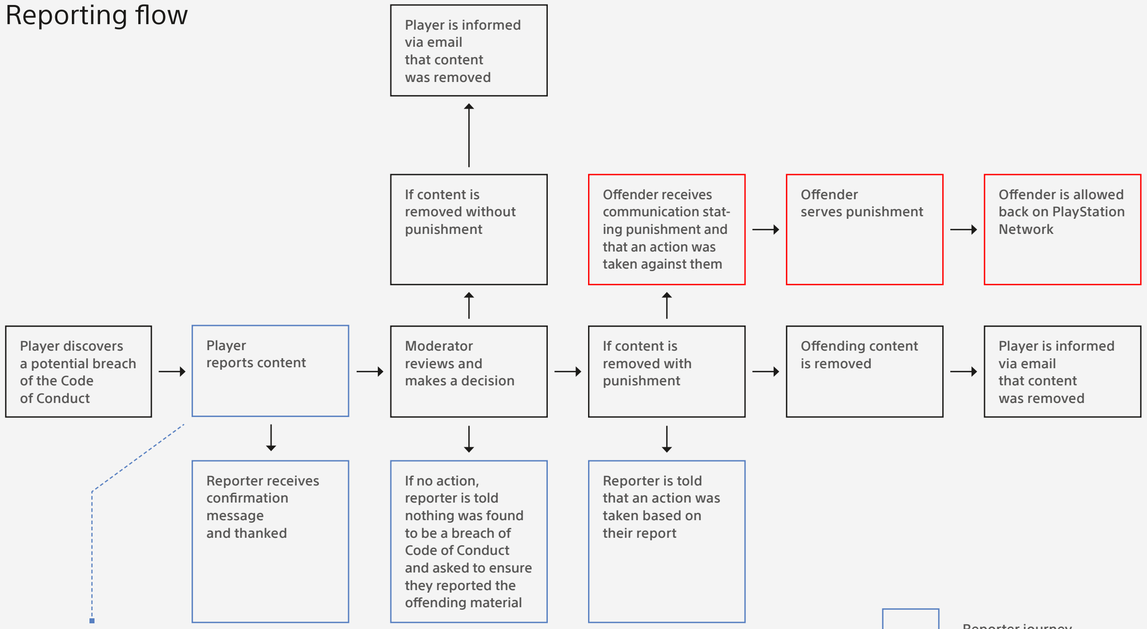
Suspensions on PlayStation Network

Moderator decisions are guided by a comprehensive educational framework, structuring consistent and fair responses to Code of Conduct violations. Suspension types are determined by the severity of the violation and the player's disciplinary history. For example, a first-time minor violation may result in a warning, while repeated or severe breaches could lead to escalating suspensions of 30-day, 60-day, or permanent account suspension. Generally, moderation decisions may result in the following actions: warning, 3-day, 7-day, 30-day, 60-day, or permanent suspension. Some suspensions are limited to player communications (PS5 only) while severe offenses can result in a console suspension, preventing the console from connecting to our network.

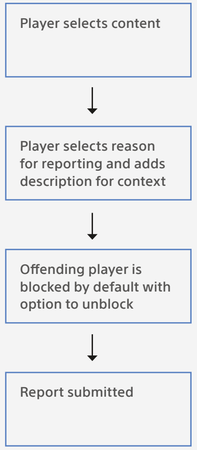
Players who submit a report receive confirmation of receipt and are notified once a moderation decision has been made, including whether the reported content was found to violate our Code of Conduct. If a violation is not confirmed, players receive educational resources on how to submit effective reports in the future.

Players found in violation of the Code of Conduct are notified via email.

Reporting flow



Reporting experience



How to Appeal a Permanent Suspension on PlayStation Network

For users whose accounts have been permanently suspended, we provide an opportunity to appeal via a dedicated webform. Once a player submits an appeal form, it is reviewed by trained moderators who review the player's offense status and history. The moderator will decide on the appeal, and the user will then be notified of the outcome of their appeal decision.

Learn more

Appeal Suspension on PlayStation Network

Our Commitment to Player Safety and Criminal Escalations

In addition to acting on an offender's account, our moderation teams are empowered to escalate reports for further review when a breach involves potential threats to life, player safety, or other unlawful activities. When necessary, this includes notifying law enforcement agencies, relevant government bodies, or other appropriate authorities.

We maintain a zero-tolerance approach toward CSEA. All identified CSEA cases are reported to NCMEC and law enforcement simultaneously.

Threats to people or property and intentions of self-harm undergo a multi-tier review process. Threats deemed to be credible are referred directly to law enforcement.

CSEA Identified, Removed and Reported by Year

Calendar Year	Number of CyberTips
2021	2,071
2022	4,102
2023	3,974
The increase in SIE's submissions from 2021 to 2022 and 2023 comes after we introduced technology to enhance detection of exploitive content.	

Easier to Understand Code of Conduct

Our Code of Conduct helps players maintain a positive and respectful gaming community on PlayStation Network. We've simplified and updated the guidelines so that they are easier to understand, consistent worldwide, and accessible to all players. Before playing online, all users must review and agree to the updated rules. We also encourage players to revisit the Code of Conduct and our policy against hate speech regularly to stay informed about acceptable gaming behavior.

Learn more

Code of Conduct

Learn more

Policy Against Hate Speech

Policy Against Hate Speech

We encourage both inclusivity and individuality on PlayStation, both in-game and online. PlayStation Network serves a community of players of all backgrounds to facilitate fun gaming experiences. Players can express themselves and their views to one another, but hate speech and hateful conduct will not be tolerated.

When players create an account, they agree to abide by the Code of Conduct, which outlines how we expect players to interact on PlayStation Network. Players can help our community by being welcoming and inclusive, behaving appropriately, and reporting misconduct to SIE. There are no exceptions for using hateful language or slurs in any form on PlayStation Network, even if the context is light-hearted, non-serious, or used as reappropriation. Even if some players may think it is appropriate to use profanity, among friends, or jokingly insulting each other, there are consequences for offensive content or using offensive language on PlayStation Network – the Code of Conduct applies.

Player and Moderator Support and Wellness

We are deeply committed to the wellbeing of our PlayStation community. That is why we continuously explore new opportunities to connect with our players and provide the support they might need.

If our team of human moderators receives a report that a player might be struggling, we will send the player a message providing action-oriented support resources.

We partner regionally with Crisis Text Line, Shout, Anata no Ibasho, and globally with ThroughLine to provide free 24/7 live mental health support and crisis intervention from trained volunteers.

Even without messages from our safety team, our player community can utilize our partnerships to connect with trained crisis counselors in times of need or distress.

Our US and UK-based consumers can share our keyword "CIRCLE" to enable others to connect directly to Crisis Text Line and Shout without the need for moderator intervention. In addition to being a PlayStation symbol, our keyword "CIRCLE" represents unity of support.

All our support partners operate with complete confidentiality and only share information with emergency services when absolutely necessary for player safety or the safety of others. For information about these services and how your information is used, please refer directly to the partner's applicable terms.

Moderator wellness is also a priority for SIE. In February 2024, SIE Online Safety won the Great British Workplace Wellbeing Award for Best Strategy for our commitment to moderator wellness and safeguarding.


Sony
Interactive
Entertainment
PlayStation