

# Australian Code of Practice on Disinformation and Misinformation Twitch Transparency Report - May 2026

## Summary

Twitch is an interactive live streaming service built around real-time communities, where creators engage in a wide range of activities, including video games, art, cooking, and other creative content.

This submission represents Twitch's second annual transparency report as a signatory to the Australian Code of Practice on Disinformation and Misinformation (ACPDM), following our first full report last year and the baseline report the year prior. Unless otherwise specified, this report covers the reporting period from 1 January 2025 through 31 December 2025. References to H1 2025 and H2 2025 refer to the first and second half of 2025, respectively.

At Twitch, we seek to create a space that supports and sustains streamers' ability to express themselves, and provides a welcoming and engaging environment for viewers, free of illegal and harmful interactions. This starts with Twitch's [Community Guidelines](#), which seek to balance user expression with community safety, and set the rules for the behaviour on the service. We develop the guidelines in consultation with the Twitch community and external safety, human rights, and policy experts, and continue to review and update them regularly to respond to evolving needs.

We identify potential violations of our Community Guidelines using a combination of automated detection, proactive human review, and user reporting. Our global Trust and Safety operations team works to quickly assess content and accounts flagged by users and our automated systems. The speed at which we can respond to user reports is critical given the live nature of Twitch, and in H2 2025, we responded to 94% of reports in under 1 hour and 99% of reports globally in under 24 hours. Twitch employs extensive human review to help ensure that enforcement actions remain accurate and fair.

We do not consider Twitch to be a high-risk service for misinformation. While misinformation is not currently prevalent on the service, we recognize the harm this type of content can cause. We maintain a dedicated policy to address misinformation: our [Harmful Misinformation Actor policy](#). This policy targets individuals who repeatedly and persistently share, widely debunked misinformation with the potential to cause real-world harm.

Harmful Misinformation Actors share three characteristics: their online presence—whether on or off Twitch—is dedicated to (1) persistently sharing (2) widely disproven and broadly shared (3) harmful misinformation topics, such as conspiracies that promote violence. We prohibit actors who meet all three of these criteria, since taken together they create the highest risk of harm, including inciting real-world harm.

Even where a user does not meet the Harmful Misinformation Actor threshold, Twitch prohibits and enforces against misinformation that targets specific communities under our [Hateful Conduct & Harassment policies](#). Twitch also takes action on content that encourages physically harmful behaviour under our [Self-Destructive Behaviour policy](#).

In addition to misinformation, Twitch takes steps to address inauthentic behaviour, including bots, spammers, impersonators, and other types of malicious actors. We use a combination of automated detection, proactive human review, and user reporting to identify and remove such activity, including attempts to evade suspension or enforcement.

We continue to evolve our approach to safety in accordance with expert guidance and emerging trends. We recognize that the prevalence of harmful misinformation can change and engage with industry, academia, and civil society to adapt our approach as needed.

Twitch participates in a variety of industry knowledge-sharing initiatives—including the [EU Code of Practice on Disinformation](#), the [New Zealand Code of Practice for Online Safety and Harms](#) (which also addresses disinformation), the [EU Hate Speech Code](#), the [EU Internet Forum](#), and the [Global Internet Forum to Counter Terrorism \(GIFCT\)](#). Twitch also recently stepped into an at-large Operating Board seat for GIFCT, supporting cross-industry coordination and information sharing on emerging online harms.

We are proud to contribute to the goals and commitments of the Australian Voluntary Code of Practice on Disinformation and Misinformation (ACPDPM).

## Commitments under the Code

Twitch has committed to the following six Objectives and related Outcomes.

Objective 1 - Provide safeguards against harms that may arise from disinformation and misinformation	
1a	Signatories contribute to reducing the risk of harms that may arise from the propagation of disinformation and misinformation on digital platforms by adopting a range of scalable measures.
1b	Users will be informed about the types of behaviours and types of content that will be prohibited and/or managed by Signatories under this Code.
1c	Users can report content or behaviours to Signatories that violate their policies under section 5.10 through publicly available and accessible reporting tools.
1d	Users will be able to access general information about Signatories' actions in response to reports made under 5.11.
1e	Users will be able to access general information about Signatories' use of

	recommender systems and have options relating to content suggested by recommender systems.
Objective 2 - Disrupt advertising and monetisation incentives for disinformation	
2	Advertising and/or monetisation incentives for disinformation and misinformation are reduced.
Objective 3 - Work to ensure the integrity and security of services and products delivered by digital platforms.	
3	The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.
Objective 4 - Empower consumers to make better informed choices of digital content.	
4	Users are enabled to make more informed choices about the source of news and factual content accessed via digital platforms and are better equipped to identify misinformation.
Objective 6 - Strengthen public understanding of disinformation and misinformation through support of strategic research	
6	Signatories support the efforts of independent researchers to improve public understanding of disinformation and misinformation.
Objective 7 - Signatories will publicise the measures they take to combat disinformation and misinformation.	
7	The public can access information about the measures Signatories have taken to combat disinformation and misinformation.

Twitch did not subscribe to Objective 5 (Improve public awareness of the source of political advertising carried on digital platforms) as we do not permit political ads.

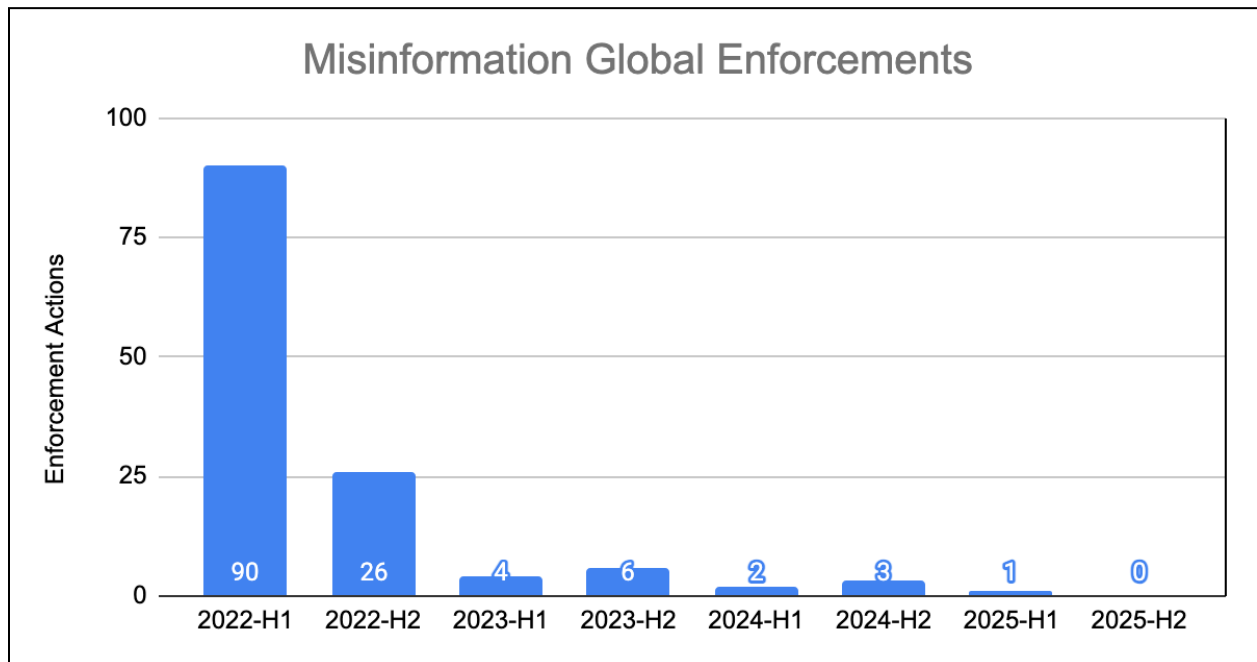
**Reporting against commitments**

**Outcome 1a: Reducing harm by adopting scalable measures**

To reduce harm to our community and the public, while preserving open dialogue between streamers and their audiences, Twitch prohibits [Harmful Misinformation Actors](#) who persistently share misinformation on or off of the service. This policy applies to users whose online presence is dedicated to (1) persistently sharing (2) widely disproven and broadly shared (3) harmful misinformation topics.

This policy is focused on Twitch users who persistently share harmful misinformation, including AI-generated misinformation, representing a targeted application of the ACPDM definition. Twitch’s policy will not be applied to users based upon individual statements or discussions that occur on the channel. We evaluate whether a user violates the policy by assessing both their on-service behaviour as well as their off-service behaviour. During 2025,

Twitch did not observe generative AI materially changing the prevalence of harmful misinformation on the service or requiring changes to existing enforcement approaches.



[\(Twitch Safety Center, H2 2025 Transparency Report\)](#)

In H2 2025, we did not issue any indefinite suspensions globally for violations of our Harmful Misinformation Actor Policy.

Our enforcement numbers remain low due to several structural and policy-specific factors. (i) Misinformation is less prevalent on Twitch relative to other online services, as the mechanics of Twitch are not conducive to the rapid spread of misinformation or the coordination of large-scale disinformation campaigns. It is difficult for a new streamer to quickly reach large audiences, as building viewership typically occurs over time. Most Twitch content is also long-form and ephemeral, meaning it is less likely to be shared or go viral in the same way as content on other user-generated or social media services. (ii) Our Harmful Misinformation Actor policy is targeted in scope and applies to accounts who persistently share harmful misinformation. Due to the long-form nature of Twitch content, enforcement decisions focus on a streamer's aggregated content over time rather than isolated statements within a single stream. (iii) Following the introduction of the policy in H1 2022, we took enforcement action against accounts that met its criteria. We consider these early enforcement actions to have had a deterrent effect, and have not observed sustained attempts by such actors to establish a presence on the service.

Twitch's Harmful Misinformation Actor Policy continues to cover a range of topics, including conspiracy networks tied to violence, public health misinformation, and civic process interference. A full list of covered topics is provided in **Appendix A**.

Even if someone is not a Harmful Misinformation Actor, Twitch takes action on content that targets specific communities under our Hateful Conduct & Harassment policies, and on content that encourages physically harmful behaviour under our Self-Destructive Behaviour policy. More information on enforcements under these policies can be found in [Twitch's Safety Transparency Report](#).

### **Outcome 1b: Inform users about what content is targeted**

Twitch's Harmful Misinformation Actor policy is outlined in our [Community Guidelines](#) ("CGs"). Our aim is for the CGs to be clear and easy to follow, while also thorough (with examples of prohibited behaviour) to help Twitch users understand the boundaries we have set so they confidently express themselves within those boundaries.

When we launched our Harmful Misinformation Actor policy in March 2022, we also published a [blog post](#) to flag the change for our community and provide more context on the policy.

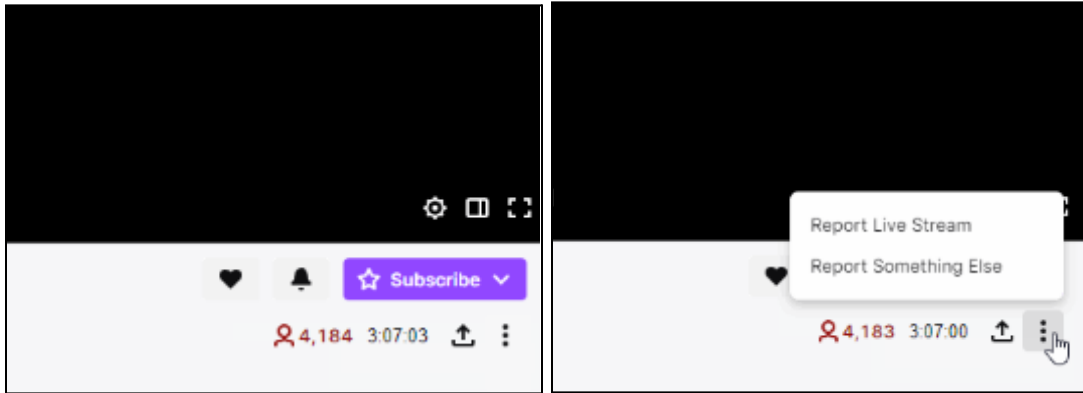
Additionally, when a user violates our Harmful Misinformation Actor Policy—or any of our policies—they receive an email notification. The notification includes the action taken, whether a suspension is permanent or temporary, and general information about the reason for the enforcement. The notification directs users to the Appeals Portal, where they can access additional details about the enforcement and the appeals process. Twitch provides Enforcement Notes within the CGs, which offer additional clarifications and examples to help explain how existing policies are applied in practice. Enforcement Notes do not introduce new rules, but instead provide further detail on how Twitch enforces its policies and address common areas of user confusion. For example, under the [Hateful Conduct policy](#), Enforcement Notes clarify how context is considered in enforcement decisions, including how certain terms may be treated differently depending on their use, and identify language that is prohibited under any circumstances.

### **Outcome 1c: Users can easily report offending content**

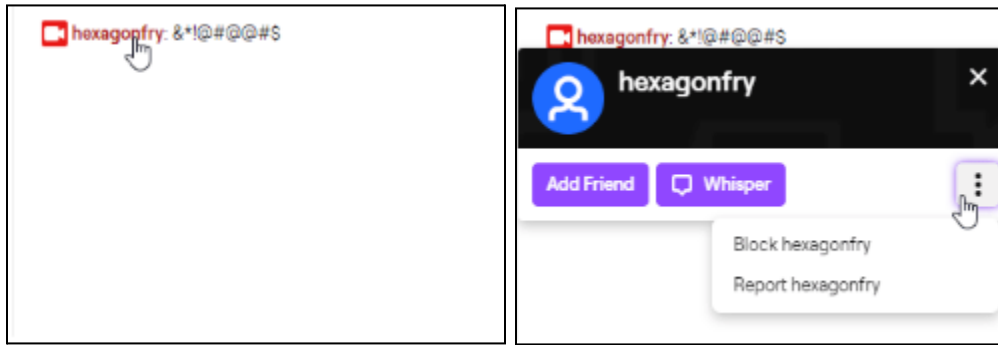
Twitch provides a dedicated, publicly available guide that explains [how to file a report](#), including step-by-step instructions for reporting across web and mobile surfaces, as well as best practices to help users submit clear and effective reports. The guide also explains what happens after a report is submitted, including how reports are reviewed and when users may receive follow-up notifications.

Users can report potential harmful misinformation actors from multiple entry points across the service, including live streams, clips, past broadcasts, and user profiles. Users can access reporting tools by clicking the three vertical dots icon associated with a stream or user, which opens the reporting interface. The reporting flow prompts users to select the most relevant category for the violation and provide additional context through a "Tell Us More" field to support review. Reports are reviewed by Twitch's moderation teams, and reporting can be submitted anonymously.

### Reporting a Channel

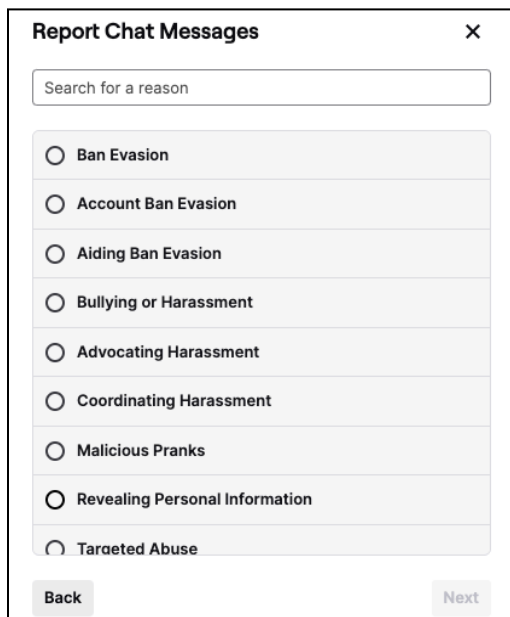


### Reporting a User in Chat



The interface prompts users to select the most relevant category for the violation, which in this case is “Misinformation.” Alternatively, a user can search for the appropriate reporting reason.

### Reporting Interface



Users can also submit a report to the specialised off-service investigations team through the team's email alias [OSIT@twitch.tv](mailto:OSIT@twitch.tv). These reports go directly to our Off-Service Investigation Team (OSIT), a global team responsible for investigating allegations of severe off-service offenses, which are handled with complete confidentiality.

### **Outcome 1d: Information about reported content available**

Twice-a-year, we publish a report outlining how we enforce our Community Guidelines, including our Harmful Misinformation Actor policy. This report is publicly available on our [website](#). We also provide a publicly-available [annual transparency report](#) under the EU Code of Practice on Disinformation. In Australia during 2025, Twitch received 31 user reports related to misinformation concerns, and all reports were reviewed and resolved within 24 hours.

### **Outcome 1e: Information about recommender engines**

Twitch provides publicly available information explaining how its recommendation systems operate, including the key signals used to surface content and the ways users can influence what they see on the service. These signals include factors such as a user's watch history, engagement, language, region, and device, which are used to recommend content that is relevant to the viewer. Users can learn more through Twitch's help resources, including [guidance](#) on how to customize content recommendations.

Users can control their experience in several ways. When browsing, viewers can sort content by 'Recommended for You' or by other channel attributes. Users can also indicate they are "not interested" in specific streamers or content categories, which informs recommendations. In addition, users may block channels, which removes that content from recommendation surfaces. These preferences can be reviewed and updated at any time through account settings.

Twitch uses [Content Classification Labels](#) (CCLs) to identify streams that include sensitive or explicit themes. Building on this, [Content Display Preferences](#) allow users to filter out streams labeled with tags such as Politics and Sensitive Social Issues. These settings apply across recommendations, and search results. Content from followed channels may still appear, ensuring users retain access to content they have explicitly chosen to engage with.

### **Objective 2: Disrupt advertising and monetisation incentives for disinformation.**

Actors that systematically provide harmful misinformation are prohibited from the service, and are therefore not eligible for monetization. Additionally, [Twitch's advertising policies](#) prohibit ads that contain deceptive, false, or misleading content as well as political advertising, including campaigns for or against a politician, political party or related to an election, and/or content related to issues of public debate.

Twitch also applies advertiser-friendly content guidelines that link monetization eligibility to content and account behaviour. Content that includes certain sensitive categories, such as sexual themes, gambling, or violent content, may receive limited or no advertising demand, depending on advertiser preferences. In addition, recent enforcement actions for brand safety-related Community Guidelines violations may result in temporary or extended removal

of ad eligibility from a channel. In some cases, accounts that are reinstated following suspension may have advertising disabled for a period of time or, where appropriate, on an ongoing basis. These measures mean that eligibility to earn advertising revenue on Twitch is influenced not only by compliance of individual advertisements, but also by the broader content environment and account behaviour, reducing incentives to create content that may undermine advertiser suitability or violate Twitch policies.

Twitch maintains an advertising review process designed to ensure ads comply with its policies before appearing on the service. The majority of ads on Twitch are placed through a managed process, where Twitch personnel work directly with advertisers to ensure submitted ads meet our standards. A smaller portion of ads are placed through Amazon's Demand-Side Platform (DSP), which allows advertisers to purchase placements programmatically. Ads placed through the DSP also undergo review before appearing on Twitch. These controls help prevent policy-violating ads from appearing on the service.

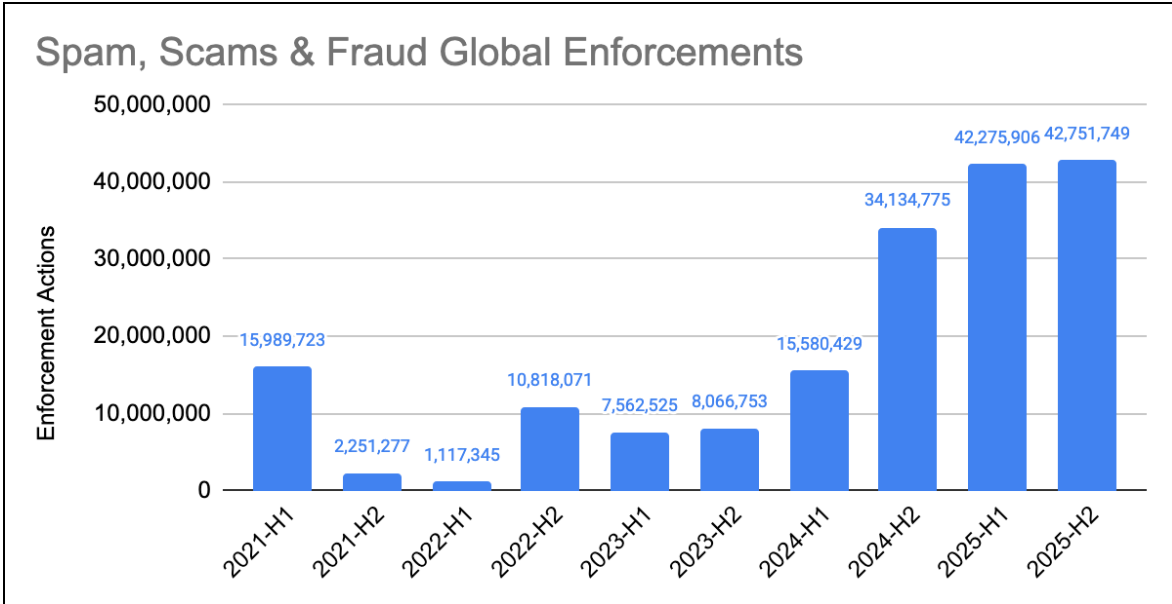
**Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms.**

Twitch's [policies state](#) that "Any content or activity that disrupts, interrupts, harms, or otherwise violates the integrity of Twitch services or another user's experience or devices is prohibited." This includes the creation of inauthentic and malicious bots, impersonation, engaging in viewership tampering (such as artificially inflating follow or live viewer stats), and selling or sharing user accounts, services, or features.

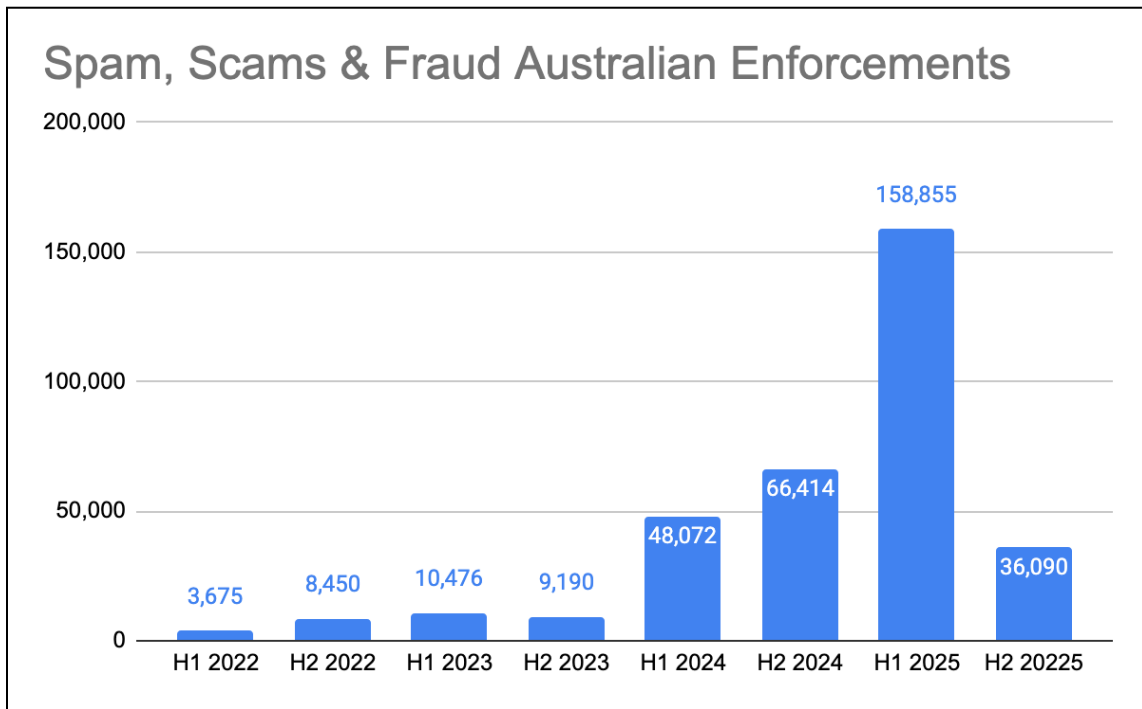
We use historical enforcement data to proactively identify patterns associated with bots and spammers. Depending on the level of confidence, we can take several actions against a suspected bot account, including requesting that the account verify a mobile phone, auto-reporting the account to be reviewed by our operations team, and adding client-side friction that increases the cost of automation.

Most cases of impersonation on Twitch are phishing attempts, where a fraudulent channel is trying to get a user to click on a malicious link. We scan the text on our channel pages for these malicious URLs and then report the channel for review by our operations team. We also actively monitor channels for viewership tampering, using a combination of handcrafted filters based on ASN and IP reputation, as well as a machine learning model based on past examples.

In H2 2025, we issued 42.7M account enforcements for spam, scams, and fraud globally; 36,090 of these were for accounts based in Australia. Spam can be both automated (published by bots or scripts) or coordinated (when an actor uses multiple accounts to spread deceptive content). Due to its automated and coordinated nature, spam is generally Twitch's largest category of enforcement and we often see significant fluctuations in enforcement between reporting periods. This is consistent with a general trend in the industry.



[\(Twitch Safety Center, H2 2025 Transparency Report\)](#)



These patterns are also reflected in the Australian data shown above. In Australia during H1 2025, most enforcement actions were taken against accounts associated with coordinated efforts to create large volumes of accounts for financial abuse. These actions were largely proactive, based on signals identified at account creation and early-stage activity. Spikes of this nature can occur when bad actors test platform systems at scale, and enforcement

actions may be concentrated within a short period before returning to baseline levels following mitigation efforts.

**Objective 4: Empower consumers to make better informed choices of digital content.**

Twitch mitigates the risk that users are exposed to harmful misinformation on the site through the measures discussed previously. We are also committed to providing users with information about how our recommendation systems work and options to customise their recommendations as discussed under Outcome 1e above.

Twitch has also invested in a media literacy campaign to empower users to think critically about what information they consume. Twitch collaborated with media literacy expert MediaWise to develop an array of educational materials that teach Twitch streamers and viewers how to better identify, and avoid spreading, misinformation and disinformation online. These materials are hosted on the [Twitch Safety Center](#).

Between January and December 2025, the media literacy materials hosted on the Twitch Safety Center received 1,886 global user visits, reflecting the worldwide number of users who accessed the educational page during that period. As part of broader media literacy efforts, Twitch has also promoted media literacy resources through creator education initiatives, including a Creator Camp livestream titled *Media Literacy with MediaWise*, which was featured on Twitch's front page

**Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research.**

Twitch remains open to supporting independent research and provides access to publicly available data through its API.

Twitch provides open access to its API, which may be used to retrieve most publicly available channel information, such as content classification labels, stream tags, and moderation settings. The API can be a useful tool for third-party researchers studying the service. Access to non-public data or permissions to make changes requires consent from the channel owner or an authorized user. The API is globally available, and developers must register for a client ID, accept the [Developer Services Agreement](#), and may be subject to additional review for research use.

**Objective 7: Signatories will publicise the measures they take to combat Disinformation.**

This report, including Outcomes 1b and 1e, provides details and links to publicly available information on the measures Twitch takes to address misinformation.

**Concluding remarks**

As a signatory to the ACPDM, Twitch takes a targeted approach to addressing misinformation on the service that balances freedom of expression with keeping our communities safe. We recognize that harmful misinformation, and its prevalence on our service, may evolve over time and we will continue to assess and adapt our approach to protect users and support the integrity of the service.



## **Appendix A: Covered Misinformation Topics Under Twitch’s Harmful Misinformation Actor Policy**

Under our Harmful Misinformation Actor Policy, we cover the following topic areas, and will continue to update this list as new trends emerge:

- Misinformation that targets protected groups, which is already prohibited under our Hateful Conduct & Harassment Policy
- Harmful health misinformation and wide-spread conspiracy theories related to dangerous treatments, COVID-19, and COVID-19 vaccine misinformation
  - Discussions of treatments that are known to be harmful without noting the dangers of such treatments
  - For COVID-19—and any other WHO-declared Public Health Emergency of International Concern (PHEIC)—misinformation that causes imminent physical harm or is part of a broad conspiracy
- Misinformation promoted by conspiracy networks tied to violence and/or promoting violence
- Civic misinformation that undermines the integrity of a civic or political process
  - Promotion of verifiably false claims related to the outcome of a fully vetted political process, including election rigging, ballot tampering, vote tallying, or election fraud
- In instances of public emergencies (e.g., wildfires, earthquakes, active shootings), we may also act on misinformation that may impact public safety