



**TikTok Australia**  
180 George Street  
Sydney  
NSW 2000

## **Australian Code of Practice on Disinformation and Misinformation**

TikTok

Annual Transparency Report

1 January 2025 – 31 December 2025

### **Summary**

At TikTok, over a billion people create and share content on topics that matter to them every day. Maintaining a safe, authentic space, where people can discover original content and engage with genuine voices, is our highest priority. In a global community, it is natural for people to hold a wide range of views. But when it comes to topics that impact people's safety, we seek to operate on a shared set of facts and reality. That is why our [Integrity and Authenticity \("I&A"\) policies](#), part of our broader [Community Guidelines](#), prohibit misinformation that could cause significant harm to individuals or society, no matter the intent of the person posting it.

Our approach to harmful misinformation is multi-layered: we remove content that violates our policies, reduce the reach of content that cannot yet be verified, partner with more than 20 IFCN-accredited fact-checking organisations across 130 markets. We also empower our community with media literacy resources and authoritative information. Critically, our enforcement outcomes are proportionate to harm, ranging from removal for the most serious violations, to reduced reach for content that is potentially harmful but unverified. This balance reflects our broader human rights commitment to protecting expression while preventing harm.

Further, as Artificial Intelligence (AI) continues to reshape content creation, TikTok recognises both the incredible creative opportunities AI enables and the potential for AI-generated content to confuse or mislead viewers. In 2025, we continued our investment in responsible AIGC practices, requiring creators to label realistic AI-generated content, extending labelling requirements to advertisers, and continuing our membership of the Coalition for Content Provenance and Authenticity ("C2PA") and the Content Authenticity Initiative (CAI).

### **Notable Highlights**

In 2025, TikTok implemented a number of global and Australia-specific initiatives to combat misinformation and disinformation and ensure a safe user experience. These included:

- Supporting users through the **Australian Federal Election** and the **Western Australian State Election** with localised Search Guides, directing users to authoritative sources of information for each respective election. For the Australian Federal Election, a dedicated Election Centre page was featured with key electoral information, media literacy videos, and resources provided by the Australian Electoral Commission and fact-checking partner AAP.
- **Restructuring our Community Guidelines** to make our standards clearer and more accessible to users across Australia and globally.
- **Responding rapidly to the Bondi Beach Terror Incident** in December, activating Search Guides within



hours to connect users with accurate information and mental health support resources during a fast-moving emergency.

- Sharing regular updates on our enforcement efforts through our quarterly [Community Guidelines Enforcement Reports](#) and an expanded [Transparency Centre](#), which, in 2025, added a dedicated [Regional Measures](#) page with a specific focus on Australia.

### Performance on Removal Efficiencies

Throughout the 2025 reporting period, we maintained high efficiency in removing content that violated our I&A policies in Australia, as illustrated in Figure 1A. We measure removal efficiency through three key metrics: proactive removals (prior to any user reporting), removals before content receives any views (“0vv”), and removals within 24 hours of posting. While Q1 2025 saw a temporary dip in our 0vv and 24-hour removal rates, this can be attributed to the volume of proactive enforcement actions taken ahead of the May 2025 Federal elections in Australia, a period during which our proactive removal rate remained strong at 99.5%, reflecting continued robustness of our internal detection mechanisms. Overall removal efficiency improved across subsequent quarters, as shown in Figure 1B, driven by our machine-based content recall and our operational and moderation capabilities.

**Fig. 1A** below shows the volume of content removed for violating our I&A policies in Australia by quarter in 2025, as well as our performance metrics on removal efficiency.

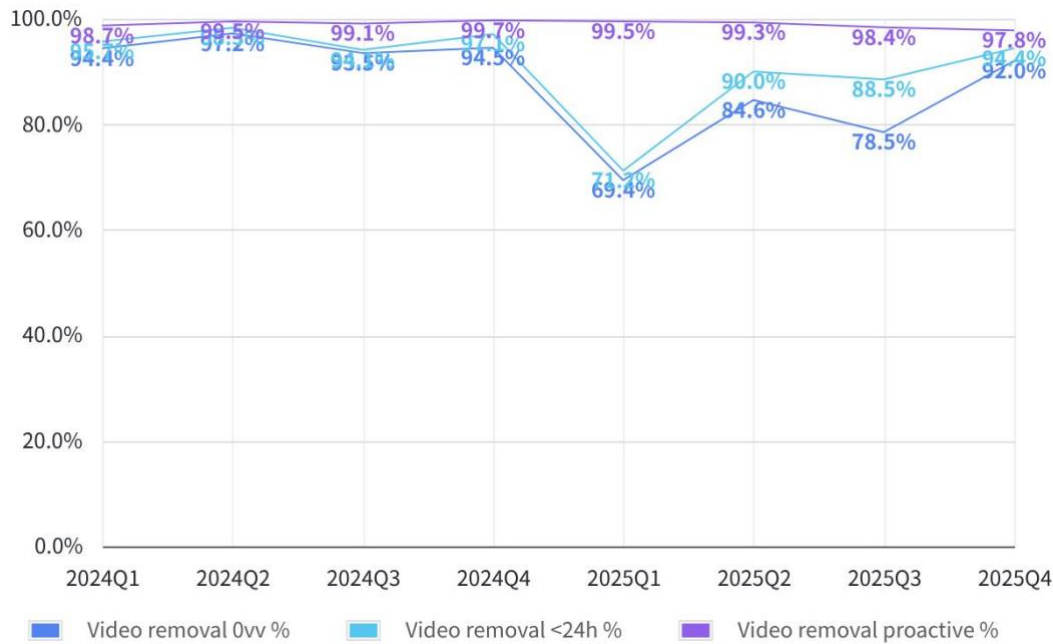
Quarter (2025)	Total videos removed	Proactive removal rate [1]	0vv removal [2]	Removal rate within 24 hours [3]
January - March	21,845	99.5%	69.4%	71.2%
April - June	12,687	99.3%	84.6%	90%
July - September	10,104	98.4%	78.5%	88.5%
October - December	18,965	97.8%	92%	94.4%

**Fig. 1A - Removal of I&A Violations in 2025 (Australia)**

In Q1 (January-March), we conducted additional sweeps targeting violative content on the platform. This led to the removal of historical content, which can be seen in the reduced removal rate within 24 hours. In Q4 (October-December), increased removal rates were driven by automated enforcement, leading to improvements in 0vv and 24 hour removal rates.

Removals form one part of a suite of tools we use to enforce our policies. For example, we may restrict the visibility of content rather than removing it completely.

**Fig. 1B** below shows that most of our removal efficiency metrics were consistently above 90% in 2025. There was a decrease in video removal rate at 0vv and video removal rate within 24 hours in Q1 2025, which can be attributed to the proactive enforcement actions taken prior to the May 2025 Federal Election in Australia. Despite the fluctuation in the 0vv removal rate, our proactive removal rate remained 99.5% in Q1 2025 which indicates high efficiency of our internal detection and enforcement mechanisms.



**Fig. 1B - Proportion of I&A violations in 2025 vs 2024 (Australia)**

**[CASE STUDY 1] Acting to reduce mis- and disinformation during 2025 Australian Federal Election**

The 2025 Australian Federal Election marked a significant milestone for TikTok in Australia as, for the first time, all major leaders and political parties were active on the platform, the Australian Electoral Commission (AEC) joined TikTok, and news outlets livestreamed election debates, reflecting TikTok's growing role in civic engagement among Australian voters.

In the weeks leading up to polling day on 3 May 2025, TikTok activated a coordinated, cross-functional election integrity operation spanning policy, trust and safety, content moderation, law enforcement outreach, and advertising integrity. Key measures included proactive impersonation sweeps, enhanced fact-checking support with shorter turnaround times, multi-layered misinformation monitoring combining human review and automated detection, and robust political advertising controls.

Ahead of the 2025 Federal Election, TikTok also convened its Safety Advisory Council for a dedicated session on election integrity, and hosted two expert roundtables bringing together academics, subject matter experts, and key opinion leaders to map the election information landscape and identify emerging risks.

User-facing election tools, developed in collaboration with the AEC and fact-checking partner AAP, reached Australians at significant scale. The Election Centre received more than 400,000 views in total, with the search guide being viewed more than 1.9 million times. More details on the feature are provided below in Objective 4.

TikTok also published enforcement statistics through its [Global Elections Integrity Hub](#) during the campaign, providing real-time public visibility into the platform's election-related content actions.

The operation concluded with zero major content integrity escalations during the election period, reflecting the effectiveness of TikTok's cross-functional preparation and real-time response capabilities.



## **[CASE STUDY 2] Acting to reduce mis- and disinformation during the Bondi Beach Terror Incident**

During the Bondi Beach Terror Incident on 14 December 2025, TikTok's Trust & Safety team rapidly activated a search guide to provide users with accurate information during a fast-moving emergency. The search guide linked to official government sources covering the incident and to a national mental health support provider, ensuring that users searching for information were directed to credible, authoritative resources at a critical moment. This marked one of the major instances of TikTok deploying a crisis search guide specifically in response to a domestic emergency event in Australia. This crisis guide was viewed more than 30,000 times on December 15. More details on the feature are provided below in Objective 4.

The following report details the measures, policies, and initiatives we undertook in 2025, reported against each of our seven commitments under the Australian Code of Practice on Disinformation and Misinformation.



## TikTok's Commitments under the Code

TikTok opts in to all Objectives and Outcomes under the Australian Code of Practice on Disinformation and Misinformation with respect to the TikTok platform.

<b>Objective 1: Provide safeguards against Harms that may arise from Disinformation and Misinformation</b>	
<ul style="list-style-type: none"> <li>• <u>Outcome 1a:</u> 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</li> <li>• <u>Outcome 1b:</u> 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</li> <li>• <u>Outcome 1c:</u> Users can report content or behaviours to Signatories that violate their policies under section 5.10 through publicly available and accessible reporting tools.</li> <li>• <u>Outcome 1d:</u> Users will be able to access general information about Signatories' actions in response to reports made under 5.11.</li> <li>• <u>Outcome 1e:</u> Users will be able to access general information about Signatories' use of recommender systems and have options relating to content suggested by recommender systems.</li> </ul>	<ul style="list-style-type: none"> <li>• Opt in</li> <li>• Opt in</li> <li>• Opt in</li> <li>• Opt in</li> <li>• Opt in</li> </ul>
<b>Objective 2: Disrupt advertising and monetisation incentives for Disinformation and Misinformation.</b>	
<ul style="list-style-type: none"> <li>• <u>Outcome 2:</u> Advertising and/or monetisation incentives for Disinformation and Misinformation are reduced.</li> </ul>	<ul style="list-style-type: none"> <li>• Opt in</li> </ul>
<b>Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms</b>	
<ul style="list-style-type: none"> <li>• <u>Outcome 3:</u> The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.</li> </ul>	<ul style="list-style-type: none"> <li>• Opt in</li> </ul>
<b>Objective 4: Empower consumers to make better informed choices of digital content.</b>	
<ul style="list-style-type: none"> <li>• <u>Outcome 4:</u> 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.</li> </ul>	<ul style="list-style-type: none"> <li>• Opt in</li> </ul>
<b>Objective 5: Improve public awareness of the source of Political Advertising carried on digital platforms.</b>	
<ul style="list-style-type: none"> <li>• <u>Outcome 5:</u> Users are better informed about the source of Political Advertising.</li> </ul>	<ul style="list-style-type: none"> <li>• Opt in</li> </ul>
<b>Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research.</b>	
<ul style="list-style-type: none"> <li>• <u>Outcome 6:</u> Signatories support the efforts of independent researchers to improve public understanding of Disinformation and Misinformation.</li> </ul>	<ul style="list-style-type: none"> <li>• Opt in</li> </ul>
<b>Objective 7: Signatories publicise the measures they take to combat Disinformation and Misinformation.</b>	
<ul style="list-style-type: none"> <li>• <u>Outcome 7:</u> The public can access information about the measures Signatories have taken to combat Disinformation and Misinformation.</li> </ul>	<ul style="list-style-type: none"> <li>• Opt in</li> </ul>

The following sections of the report outline the specific measures, policies and projects undertaken to promote authenticity and address misinformation on TikTok.



## Reporting against 2025 Commitments

**Objective 1: Provide safeguards against Harms that may arise from Disinformation and**

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

TikTok has continued to refine and expand its measures and policy frameworks to safeguard users from potential harms associated with misinformation. Our I&A policy framework has matured significantly, with 2025 marked by stronger enforcement consistency, deeper integration of machine-learning detection and expanded user facing transparency tools. While violations of I&A policies represented a slightly higher proportion of total Community Guidelines violations compared to 2024, our removal efficiency metrics continue to improve - reinforcing our commitment to timely and effective enforcement.

Our approach to addressing misinformation and disinformation, and I&A as a whole, is multi-faceted and centred on:

- **Protecting People from harmful content:** Our Community Guidelines prohibit inaccurate, misleading, or false content that may cause significant harm to individuals or society, regardless of the poster's intent. This includes not only physical harm, but also societal harm, like the undermining of trust in elections or public health initiatives.
- **Robust and adaptive policies:** These are regularly updated to reflect evolving enforcement challenges, such as the updates made to our global policies on Election Misinformation in 2025.
- **Advanced detection systems:** We leverage machine-based approaches to proactively detect and disrupt violative content at scale and ensure extensive coverage of misinformation threats.
- **I&A moderators:** For cases requiring a nuanced approach, we have dedicated teams and processes for I&A policies such as misinformation, including enhanced training, tooling and expertise.
- **Partnering with experts:** TikTok partners with expert, IFCN certified Fact-Checkers across the world to support consistent and accurate moderation, understand local context, and empower our community with authoritative information.
- **Empowering users with information:** We provide accessible and up-to-date resources to users, particularly around electoral events and the nature of harmful misinformation. For example, by maintaining current information through our Safety Centre landing pages and our [Countering Harmful Misinformation Guide](#)

**Fig. 2A** below shows the volume of content removed for violating our I&A policies in Australia by quarter in 2025, as well as our performance metrics on removal efficiency.

Quarter (2025)	Total videos removed	Proactive removal rate	Removal before content receives any views [0vv removal]	Removal rate within 24 hours
January - March	21,845	99.5%	69.4%	71.2%
April - June	12,687	99.3%	84.6%	90%
July - September	10,104	98.4%	78.5%	88.5%
October - December	18,965	97.8%	92%	94.4%

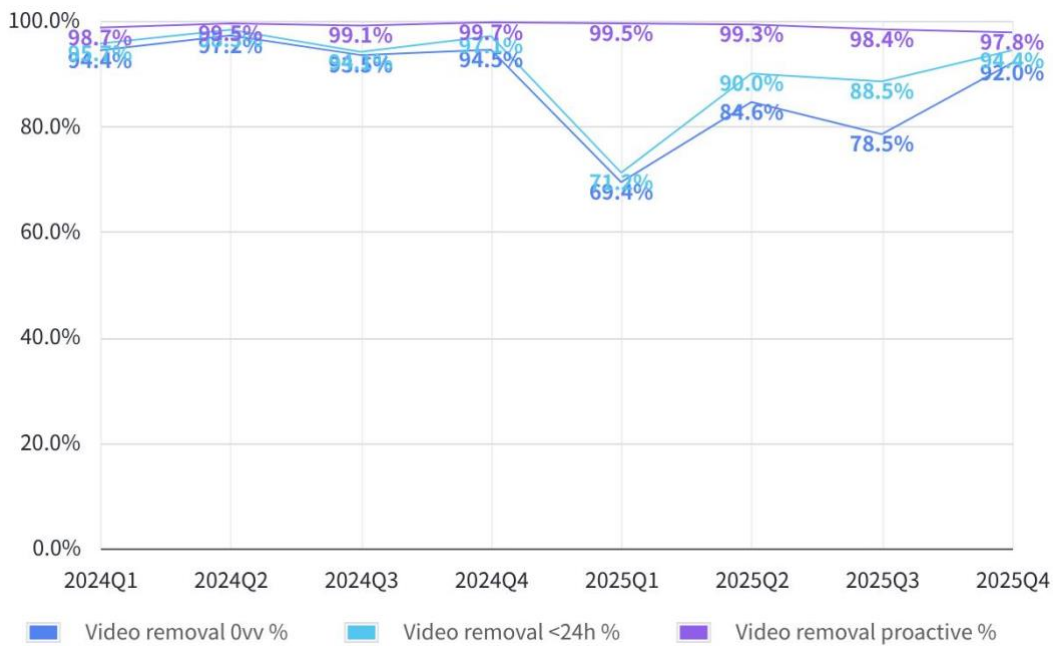
**Fig. 2A - Removal of I&A Violations in 2025 (Australia)**



We measure removal efficiency using three metrics: 1) proactive removal, 2) removal within 24 hours and 3) 0vv removal - removing a video before receiving any views.

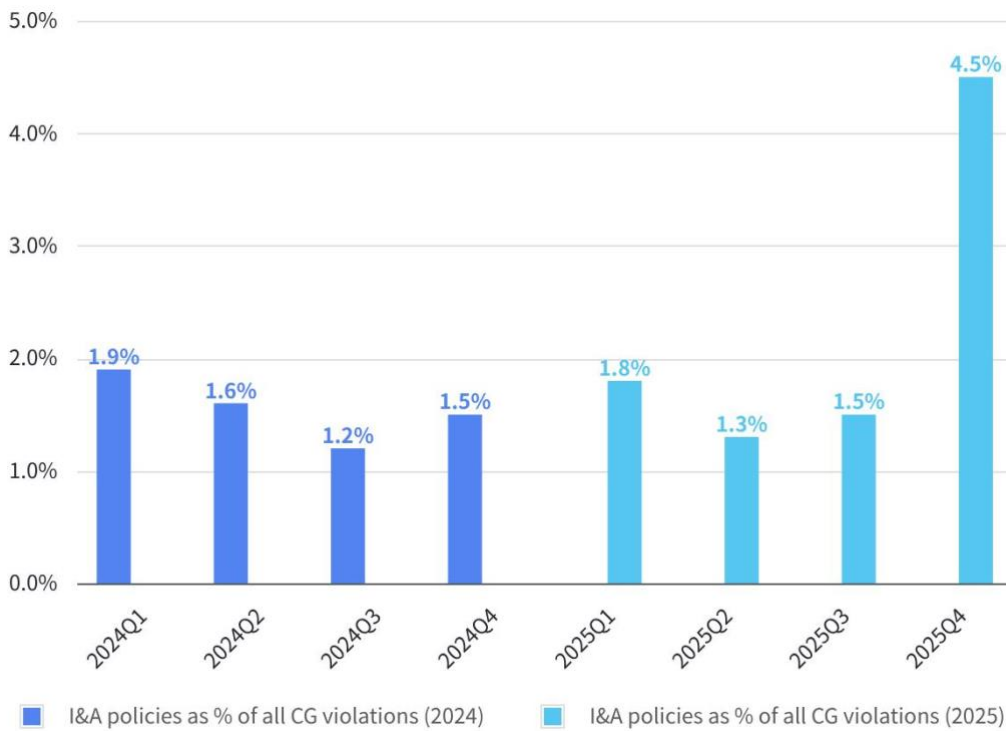
For 1) proactive removal, we calculate the percentage of videos that were removed without any user report record. For 2) removal within 24 hours, we calculate the percentage of videos that were removed within 24 hours of it being posted on our platform. For 3) 0vv removal, we calculate the percentage of videos that were removed without any views.

**Fig. 2B** below shows that most of our removal efficiency metrics were consistently above 90% in the year 2025. There was a fluctuation in video removal rate at 0vv and video removal rate within 24 hours in Q1 2025, which can be attributed to the proactive enforcement actions taken prior to the May 2025, Federal Election in Australia. Despite the fluctuation in the 0vv removal rate, our proactive removal rate has remained 99.5% in Q1 2025 which indicates high efficiency of our internal detection and enforcement mechanisms.



**Fig. 2B - Proportion of Integrity and Authenticity violations in 2025 vs 2024 (Australia)**

In 2025, I&A violations represented a higher share of total Community Guidelines violations in Australia – ranging from 1.8% to 4.5% per quarter, compared to 1.5% to 1.9% in 2024. This shift does not reflect an increase in related content on the platform. Rather, it is a result of an overall decline in enforcement volume: the absolute number of I&A removals was lower in 2025 than in 2024, but so too were total removals across all categories. Despite this change in composition, our proactive detection rate remained consistently strong, exceeding 97% each quarter.



**Fig. 2C - Proportion of Integrity and Authenticity violations in 2025 vs 2024 (Australia)**

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

TikTok's [Community Guidelines](#) are accessible to users both within the app and on our website, offering clear and detailed explanations of what constitutes misinformation, the types of harmful misinformation that are prohibited, and the eligibility criteria for content to appear in users' feeds. These Guidelines are shaped through extensive consultation with a broad range of stakeholders, including NGOs, regulators, academics, subject matter experts, and members of the community, to ensure they remain relevant and effective. We are committed to regularly reviewing and, where appropriate, updating these Guidelines, and we notify our community of any significant changes to maintain transparency and trust.

In 2025, we introduced major updates to our Community Guidelines to make our policies clearer to our users. As part of this, we organised our Community Guidelines by topic to make information easier to find for users. Under each category, users will find a "More information" section. That's where we define key terms, answer common questions, and share helpful examples. These updates to our policies included requirements for users to label AI and synthetic media content if it depicts real people or realistic scenes. These updates also clearly articulated to users that we "don't allow AIGC that misleads about matters of public importance or that harms individuals." Finally, the updates clearly re-articulated two enduring policy commitments TikTok has made publicly related to this Code, namely:

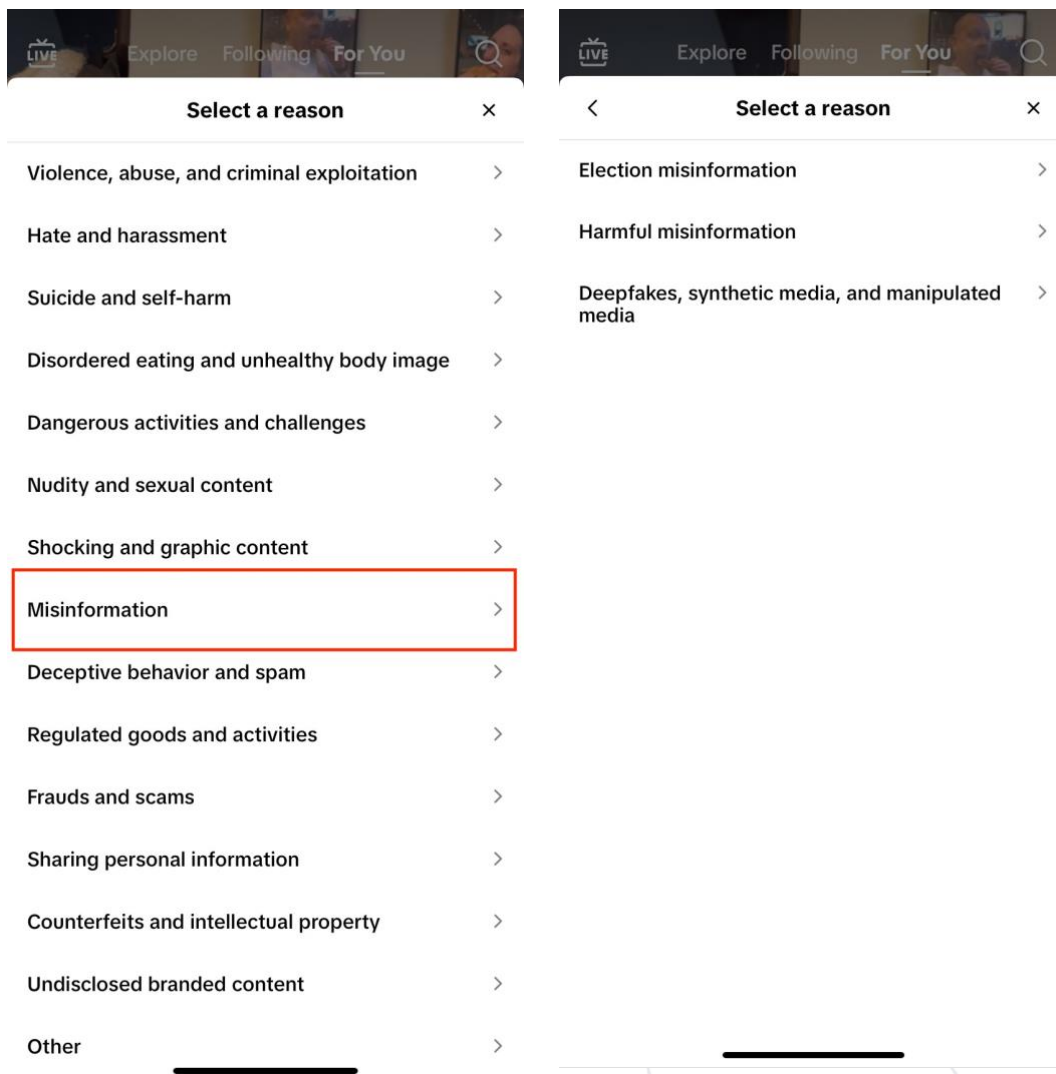
- **Misinformation:** We don't allow misinformation that could cause significant harm to individuals or society.
- **Civic and Election Integrity:** We don't allow content that could mislead voters or interfere with elections, including false claims about how to vote, who can vote, or the outcome of an election.



In the event a user violates TikTok’s Community Guidelines, we continue to provide notification to users of strikes they might incur with reference to the specific content in question. The user will then be afforded the opportunity to appeal the decision. We directly [advise](#) users that “When we take action on your content, we’ll let you know through an inbox notification or within the “[Account Status](#)” page in your in-app Safety Center. You can also submit an [appeal](#) if you think the moderation decision was a mistake. When you report harmful misinformation, you can track the report’s outcome in the “Report Records” page in your Safety Center.” This provides users with specific information about the nature and types of content that has violated TikTok’s Community Guidelines, as well as the means to appeal.

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

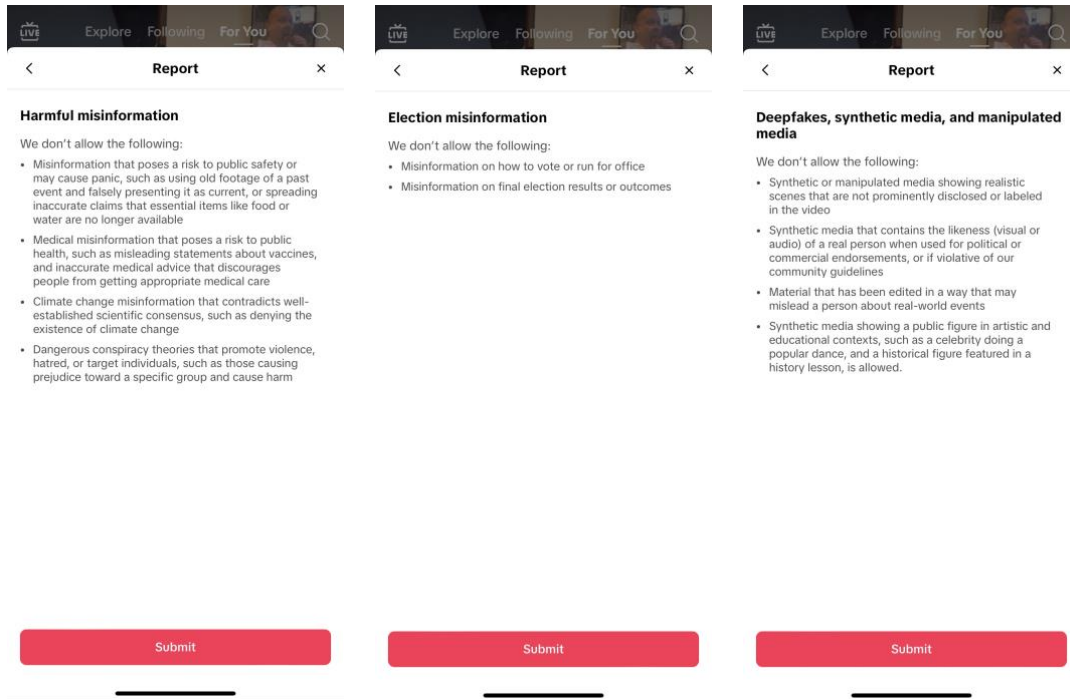
TikTok is designed so that users can easily report content they consider to be potentially violative of our Community Guidelines. This includes a dedicated category to report misinformation and a selection of sub-categories to choose from.



**Fig. 3A: User in-app reporting interface**



Before reporting any content for misinformation, users are informed about what is not allowed under each specific subcategory. This helps clarify our Community Guidelines, reduce ambiguity about permitted content, and enhance platform safety.



**Fig. 3B: User in-app reporting interface**

Users can keep track of their reports (including their status) and view their report history under Settings and Privacy > Help Centre > Safety Centre. When a user submits a report, we also provide them with the option to hide any further content being shown from the account in the feed.

Reporting misinformation is not limited to short-form videos. We enable users to report misinformation across other features of the platform, including [comments](#) on videos, [direct messages](#) they receive from other users, [accounts](#), [sounds](#), [hashtags](#) and [auto-suggestions](#) generated when they search for something on TikTok. Users can also report [LIVE videos](#) and [comments](#) on livestreams if they encounter content that violates our Community Guidelines. We also have reporting channels for non-TikTok users to report potentially harmful material. Our reporting form hosted on our website enables direct reports for our immediate review and action. Instructions for our publicly accessible reporting tools are available on our [website](#).

We publish information about response times to user reports routinely in our [Community Guidelines Enforcement Report](#).

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

Our Transparency Centre serves as a central hub to understand how TikTok moderates content, develops products, and protects user data. It provides users and the broader public with access to data and periodic reports, including:

- [Community Guidelines Enforcement Reports](#): Quarterly insights into our efforts to enforce guidelines and terms of service.
- [Information Requests Reports](#): Biannual data on user information requests from governments and law



enforcement, along with our responses.

- [Government Removal Requests Reports](#): Biannual data on requests from government agencies to restrict content and our actions in response.
- [Intellectual Property Removal Requests Reports](#): Biannual data on requests to remove copyrighted and trademarked content, along with our responses.
- [Covert Influence Operations](#): Monthly insights and data into our efforts to detect and disrupt covert influence operations that attempt to establish themselves on TikTok and undermine the integrity of our platform.
- [Global Elections Integrity Hub\\*](#): A centralised resource to provide users with authoritative election information, safety tips, and resources to navigate civic processes securely, while also highlighting our ongoing work to promote election integrity and counter misinformation.

*\*The Global Elections Hub was only launched in 2025, following the launch of the US hub in 2024.*

In 2025, we also added the following page to our transparency centre:

- [Regional Measures](#): With specific focus on Australia.

These pages and reports are published in multiple languages, are available for download in machine-readable formats from our Transparency Centre, and can be visualised in interactive charts and graphs.

Our [latest Community Guidelines Enforcement Report](#) for the period October 2025 - December 2025, published in March 2026, summarises our capabilities to proactively detect and remove violative material. We have provided Australia-specific, 2025 statistics below.

Quarter (2025)	Total videos removed	Proactive removal rate	Removal before content receives any views	Removal rate within 24 hours
January - March	926,625	99.1%	90.4%	95.1%
April - June	716,701	99.1%	90.2%	95.9%
July - September	607,935	98.7%	83.9%	92.38%
October - December	486,040	98.1%	77.6%	90.4%

*NOTE: The fluctuation in removal rates, between October and December 2025, before the content received any views, is due to actions we took to counter violative content related to Sensitive and Mature Themes. It was not directly related to Misinformation.*

**Fig. 4: Video Removals in 2025 (Australia)**

**Outcome 1e: Users can access general information about Signatories’ use of recommender systems and have options relating to content suggested by recommender systems.**

The content on TikTok is created by our community, and recommendations are based on users' interactions. The algorithm analyses signals like view counts, likes, and shares to generate a prediction score, which helps rank and recommend videos.

Our [support page](#) provides detailed information to users about how content is recommended across TikTok and how users can influence what they see on the platform.



In addition to the signals users provide through their interactions on TikTok, we have developed additional tools to help our community better control the content recommended to them. These include:

- **[Manage Topics](#)**: With Manage Topics, you can customise how often content related to over 10 popular topics is recommended in your feed - including Creative Arts, Travel, Nature and Sports.
- **Not interested**: Users can long-press on the video in their For You feed and select '[Not interested](#)' from the pop-up menu. This will let us know they are not interested in this type of content and will limit how much of that content we recommend.
- **Video keyword filters**: Users can [add keywords](#) – both words and/or hashtags that they would like to filter from their For You feed.
- **For You feed refresh**: To help discover new content, users can [refresh their For You feed](#), which provides an entirely new side of TikTok for them to explore.

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

### **Outcome 2: Advertising and/or monetisation incentives for Disinformation and Misinformation are reduced.**

We place considerable emphasis on proactive content moderation and the vast majority of the violative content we remove is taken down before it is reported to us by users or receives any views.

We [do not allow](#) visual or audio material that has been edited, spliced, or combined in a way that may mislead a viewer about real-world events. Statistics about the number of ads taken down for mis- and disinformation are not available at this stage, as this data is captured as part of broader policies that cover harmful misinformation in ads. A total of 184,205 ads were removed in Australia in 2025 under those broader advertising policies which cover (among other things) harmful misinformation risks. Ads removal volumes reported here include both misinformation and violations covered by the broader policies, and therefore reflects a higher figure.

- Prohibited and Restricted Industry: 55,832
- Prohibited and Restricted Content: 815
- Misleading and False Content: 127,558

TikTok is in the process of launching granular misinformation Advertising Policies, which we anticipate will be reflected in future reporting periods. We are continuously reviewing and improving our tools and processes to fight disinformation and will report on any further developments in the next report.

### **Transparency and Scrutiny of Advertising**

Advertisements must comply with, and are reviewed against, our [Advertising Policies](#) before being allowed on our platform. These policies specifically prohibit misleading, inauthentic and deceptive behaviours.

Like all users of our platform, participants in content monetisation programs must adhere to our Community Guidelines, including our I&A policies. Those policies make clear that we do not allow activities that may undermine the integrity or authenticity of our platform. They also make clear that we remove content or accounts, including those of creators, which contain misleading information that causes significant harm or deceptive behaviours. In certain scenarios, we may remove a creator's access to a creator monetisation feature.

We continue to engage with external stakeholders in order to increase the effectiveness of our scrutiny of ad placements. We offer pre-campaign solutions to advertisers so they can put additional safeguards in place before their content goes live to mitigate the risk of their advertising being displayed adjacent to certain types



of user-generated content.

We also work with third parties to offer post-campaign solutions that enable advertisers to assess the suitability of user content that ran immediately adjacent to their ad in the For You feed, against their chosen brand suitability parameters.

### **Our policies and approach**

Our I&A policies within our [Community Guidelines](#) are the first line of defence in combating harmful misinformation and deceptive behaviours on our platform. All users are required to comply with our Community Guidelines, which set out the circumstances where we will remove or otherwise limit the availability of content.

Paid ads are also subject to our [Advertising Policies](#) and are reviewed against these policies before being allowed on our platform. Our Advertising Policies specifically prohibit false, misleading, or unsubstantiated content that may cause significant harm to individuals or society, regardless of intent. They also prohibit other inauthentic and deceptive behaviours. We will make every reasonable effort to ensure that any content in breach of these policies does not appear on our platform. Accounts deemed in severe or repeated violation may be suspended or banned.

We also have other, existing policies that cover the majority of harmful misinformation in ads, due to these policies' expansive coverage. For example, our [Paid Political Advertising policy](#) prohibits ads featuring political content (irrespective of its veracity) across any of our monetisation features, including paid ads, creators being paid to make branded political content, and other promotional tools on the platform. This helps prevent misinformation which may reference, promote or oppose candidates or nominees for public office, political parties, or elected or appointed government officials. While our policies prohibit political ads and other forms of monetised political content, we allow [governments](#) to advertise in [limited circumstances](#), for example, to promote public health initiatives, if all other ad policies are adhered to.

Our Advertising Policies also require advertisers to meet a number of [requirements](#) regarding the landing page. For example, the landing page must be functioning and must contain complete and accurate information, including about the advertiser. Ads risk not being approved if the product or service advertised on the landing page does not match that included in the ad.

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

**Outcome 3: The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.**

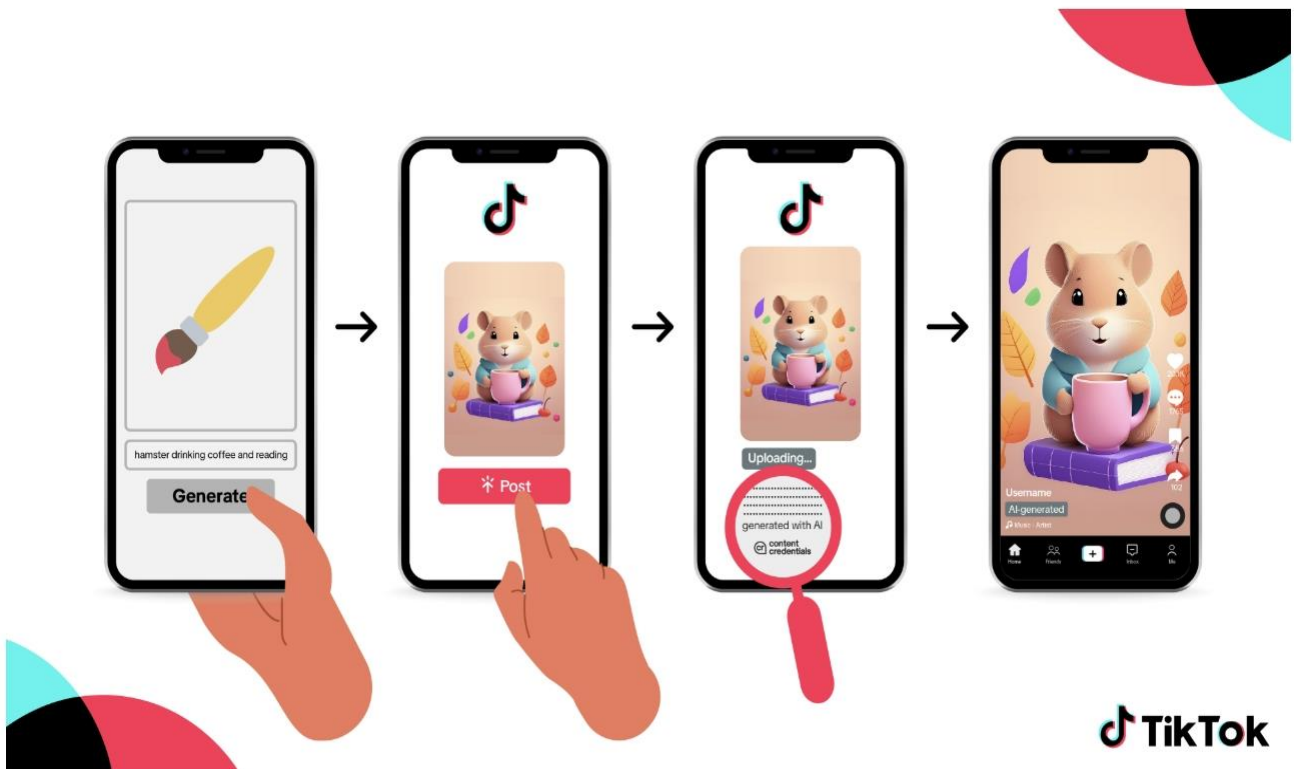
### **AI-Generated Content ("AIGC")**

Building on our efforts to improve transparency on AIGC, we empower creators to use the [AI-generated label](#) to disclose content that is completely AI-generated or significantly edited by AI. In March 2025, TikTok marked two years since becoming an inaugural supporter of the Partnership on AI's [Framework for Responsible Practices for Synthetic Media](#) by co-publishing a [case study](#) on how we applied the Framework to our AI policies. This builds on our efforts in 2024 to address inauthentic user behaviours, namely:

- In February 2024, we joined forces with 19 leading global tech companies to sign a new [pledge](#) to combat deceptive AI-generated election content. These commitments build on our continued work to advance AI transparency, combat misinformation and protect elections globally.



- In May 2024, we joined the [Coalition for Content Provenance and Authenticity \(C2PA\)](#) and the [Content Authenticity Initiative \(CAI\)](#) to help drive industry adoption of Content Credentials, a technology that helps platforms more easily label AIGC. During this time, we also:
  - a. Implemented the Coalition for Content Provenance and Authenticity (C2PA) [Content Credentials](#), which enables our systems to instantly recognise and automatically label AIGC.
  - b. Supported the Coalition’s working groups as a C2PA General Member.
  - c. Joined the Content Authenticity Initiative (CAI) to drive wider adoption of the technical standard.
- In November 2025, we [announced new measures](#) to strengthen transparency around AI-generated content, including: testing a new control within “Manage topics” to help people choose how much AIGC they see in their For You feeds; strengthening our labelling approach through creator labelling tools, detection models, C2PA Content Credentials and invisible watermarking; and launching a \$2 million AI literacy fund to support expert-led content about AI literacy and safety.

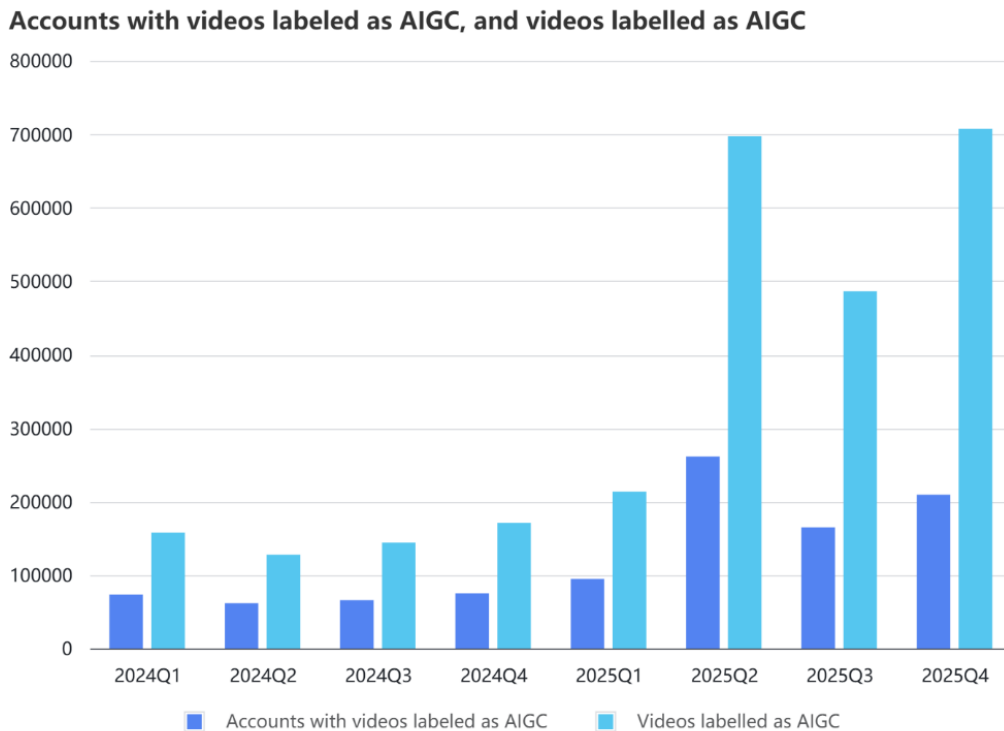


**Fig. 5: AI Generated Content Labels**

Similarly, in our [Advertising Policies](#), we require advertisers to use the AIGC label to disclose content that is completely AI-generated or significantly modified by AI.



The growth of AI labelling on TikTok in Australia is set out below:



This shows increased adoption of our AIGC labelling tools since the start of 2024. Users watching these videos are shown notices similar to those above.

We continue to take proactive measures to remove harmful AIGC as part of our broader approach to enforcing our Integrity and Authenticity policies.

### Covert Influence Operations (CIO)

TikTok remains committed to preventing, detecting and deterring inauthentic user behaviours on our platform. These efforts include removing inauthentic accounts, tackling fake account engagement and disrupting [Covert Influence Operations \(CIO\)](#). In 2025, we continued to publish the details of the CIO networks we identified and removed within our transparency reports, [available on our website](#). There were no identified networks that specifically targeted Australia.

To counter these emerging threats and stay ahead of evolving challenges, we have expert teams who detect, investigate, and disrupt covert influence operations. We use a standardised, data-driven analytical framework to determine whether accounts are violating our policies on covert influence operations. Our approach focuses on assessing accounts' behaviour and looking for evidence of linkages between them, their actions or their techniques to determine if they are coordinating together to misrepresent who they are and what they are doing. Specifically, we look for evidence that:

1. They are coordinating with each other. For example, they are operated by the same entity, share technical similarities like using the same devices, or are working together to spread the same narrative.
2. They are misleading our systems or users. For example, they are trying to conceal their actual location, or using fake personas to pose as someone they're not.
3. They are attempting to manipulate or corrupt public debate to impact the decision making, beliefs



and opinions of a community. For example, they are attempting to shape discourse around an election or a conflict.

Our experts work to counter covert influence operations by studying the many layers of techniques, tactics and procedures that deceptive actors use to try to manipulate platforms, drawing from a variety of disciplines including threat intelligence and data science. We use these learnings to develop local and global strategies to categorise and study deceptive behaviours on TikTok. We continuously evolve our detection systems, using insights from previous investigations to pinpoint networks and behaviours that violate our policy standards as early as possible. This includes detecting networks' attempts to re-establish their presence on our platform post-disruption.

### **Impersonation**

Ahead of the 2025 Australian Federal Election, TikTok conducted four rounds of proactive impersonation sweeps across Government, Politician, and Political Party Accounts (“GPPAs”). In addition, TikTok deployed an account creation prevention strategy targeting 473 name variations of high-profile politicians, political parties, and government agencies – a proactive measure that contributed to a measurable reduction in new impersonation account creation in the weeks immediately before election day.

### **Spam and Fake Engagement**

[Our I&A policies](#), which address fake engagement, do not allow the trade of services that attempt to artificially increase engagement or deceive TikTok’s recommendation system. We do not allow our users to facilitate the trade of services that artificially increase engagement, such as selling followers or likes, or to provide instructions on how to artificially increase engagement on TikTok.

If we become aware of accounts or content with inauthentically inflated metrics, we will remove the associated fake followers or likes. Content that tricks or manipulates others as a way to increase engagement metrics, such as “like-for-like” promises and false incentives for engaging with content, is ineligible for our For You Feed. More information about our approach to disrupting CIO is set out in the Appendix.

We remain vigilant in detecting external threats and protecting the platform from fake accounts and inauthentic engagements. These threats continue to target our systems, which can occasionally cause fluctuations in related metrics. We take action to both remove and prevent likes, followers, and follow requests when we deem the activity to come through automated or inauthentic mechanisms. When we remove accounts for spam, we also remove videos and comments created by those accounts under our spam policies.



Quarter (2025)	Fake likes removed	Fake likes prevented	Fake followers removed	Fake followers requests prevented	Fake accounts prevented	Videos removed from Fake accounts
January - March	11,491,846	12,447,209	663,980	15,516,907	933,296	<i>Due to data retention policies, data in Q1 and Q2 2024 is not currently available</i>
April - June	12,253,172	10,275,730	1,659,672	20,573,772	234,745	
July - September	31,429,039	7,304,360	12,934,648	18,501,362	1,206,954	555,555
October - December	79,649,643	11,084,510	49,747,938	18,136,186	688,195	427,584

**Fig. 6 - Removal of Spam & Fake Engagement in 2025 (Australia)**

We were able to remove significantly more fake likes, fake followers and videos from fake accounts in H2 2025, due to a new algorithm that rolled out in late Q3.

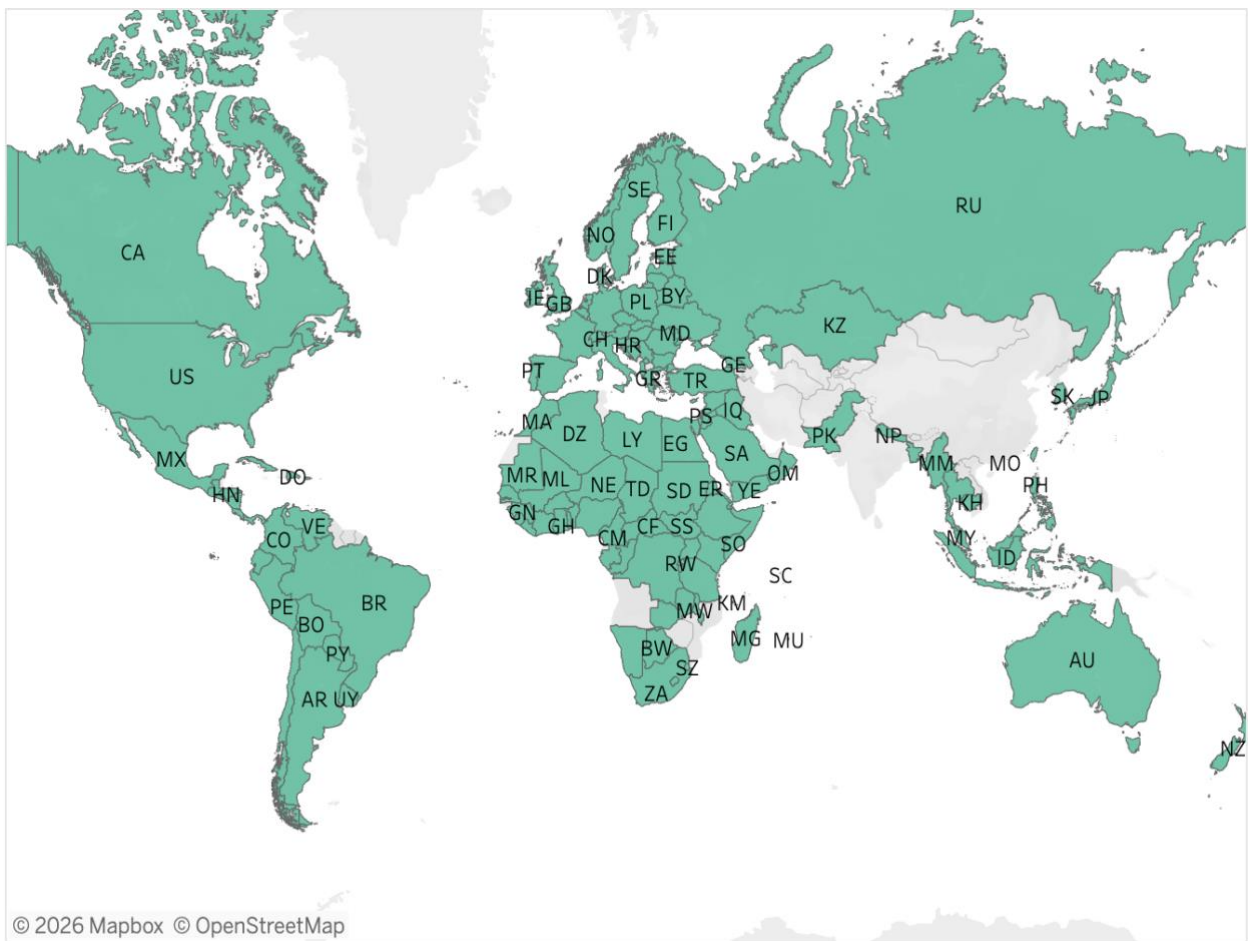


**Objective 4: Empower consumers to make better informed choices of digital content**

**Outcome 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.**

**Harmful Misinformation Guide**

In 2025, we continued to maintain and update our [Harmful Misinformation Guide](#), available via our online Safety Centre globally. Building on its introduction in 2024, the Guide has been refreshed to reflect our current policies, definitions, and fact-checking partnerships. It continues to provide users with accessible, easy-to-read information on harmful misinformation, including a visual map of our global fact-checking coverage, including in Australia.



**Fig. 7: Global Fact-Checking Coverage, Source: [Harmful Misinformation Guide](#)**

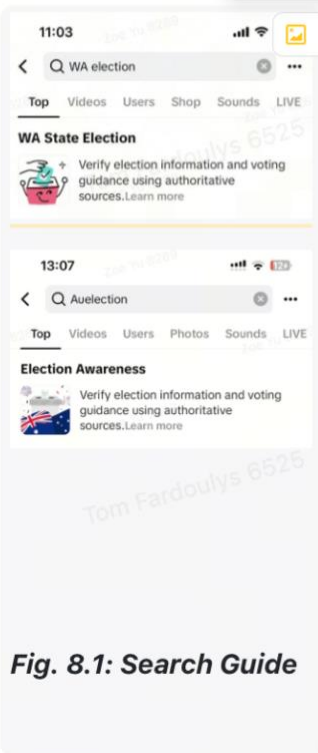
**Combating the Spread of Misinformation (in elections and high-risk events)**

We take our responsibility to protect the integrity of our platform with the utmost seriousness. We're proud to be a place that brings people together over creative and entertaining content, and we work hard to keep harmful misinformation and other violations of our policies off our platform. In the high-stakes environment of elections, preventing misinformation and disinformation is critical to safeguarding democratic processes and ensuring informed decision-making by voters. To protect civic integrity processes and prevent the spread of misinformation and disinformation, TikTok continued in 2025 to:

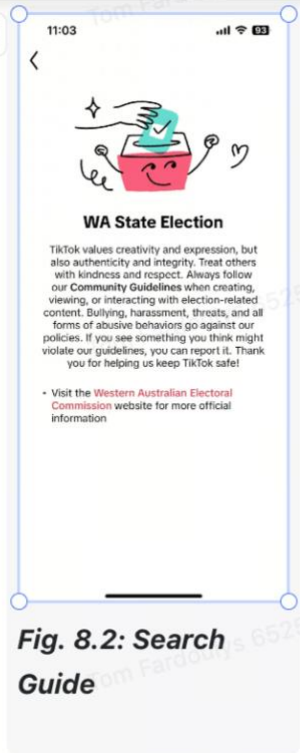
- Invest in training and development for our human moderation teams, while also advancing machine learning models to ensure comprehensive coverage of nuanced misinformation threats.
- Enhance the detection and removal of inauthentic visual and audio trends, including AI-generated content, to combat manipulated, edited, and deepfake content.
- Maintain a database of fact-checked claims to enable human moderators to accurately identify misinformation content.
- Roll-out localised election integrity campaigns in Australia in advance of the 2025 federal election and relevant state elections. This included support for user-facing interventions including Election Search Guide, video and live notice tags, and Election Centre H5 pages developed in collaboration with the Australian Electoral Commission and fact-checking partner AAP for the 2025 federal election and an Election Search Guide for the Western Australian state election. The Election Centre H5 page, developed in collaboration with the AEC and AAP for the federal election, received over 400,000 video views during its active period. The Election Search Guide received over 1.9 million views.

Search guides are developed by TikTok to provide users with authoritative information in the lead up to each federal/state election. The Search guides will appear when users search from a pre-determined list of keywords (please see Fig. 8.1 and 8.2 for a sample of a search guide).

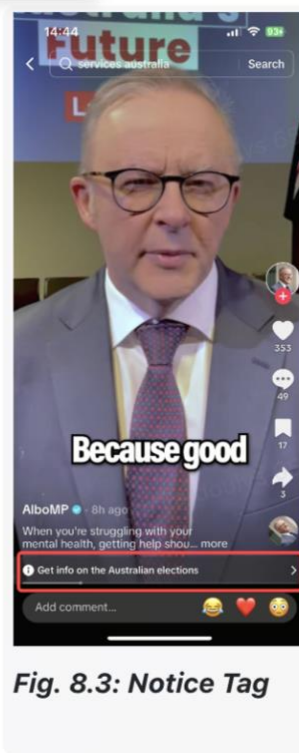
Users are also prompted to visit the Election Centre with dedicated notice tags appearing both on videos and LIVE streams related to elections, as well as through banners that appear on searches for related terms. (See Fig. 8.3 for a sample of a Video notice tag).



**Fig. 8.1: Search Guide**



**Fig. 8.2: Search Guide**

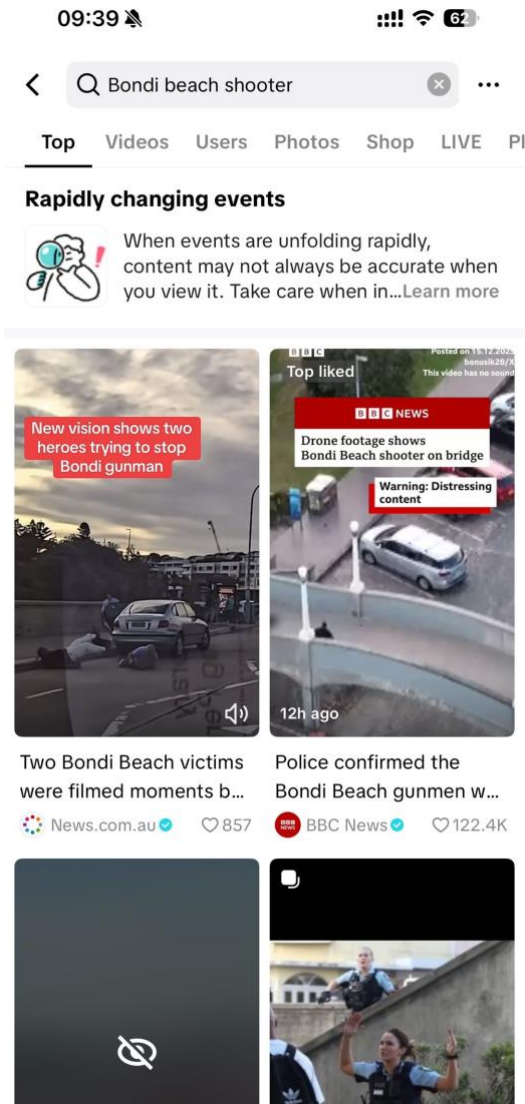


**Fig. 8.3: Notice Tag**



**Fig. 8.4: Election Centre**

During the Bondi Beach Terror Incident on 14 December 2025, we launched a search guide to provide users with accurate information during this rapidly changing event, by linking official government websites concerning the event and a national provider for mental health on the search guide. This crisis guide was viewed more than 30,000 times on 15 December.



### Rapidly changing events

When events are unfolding rapidly, content may not always be accurate when you view it. Take care when interacting with sensitive content, and remember to verify information with credible sources. Please report anything you think might violate our **Community Guidelines**

- If you have been impacted by this tragedy, you may find this resource on [the government's information page](#) useful.
- Check with **authoritative sources** for accurate and up-to-date information.

**Fig. 9: Bondi Terror Incident Search Guide**



## **Objective 5: Improve public awareness of the source of Political Advertising carried on digital platforms**

### **Outcome 5: Users are better informed about the source of Political Advertising.**

#### **Prohibiting Paid Political Ads**

TikTok does not allow [ads featuring political content](#) across any of our monetisation features, including paid ads, creators being paid to make branded political content, and other promotional tools on the platform. We also impose prohibitions at the account level for advertisers we identify as [politicians and political parties](#), who are not able to create advertising accounts or access advertising features.

Sharing political beliefs and engaging in political conversation is allowed as organic content; however, our policies prohibit users from paying to advertise or promote this content. In limited circumstances, we allow governments and official entities overseeing elections to advertise, for example, to promote public health or safety initiatives. In general, we allow public service announcements from government entities and non-profits, provided they are not politically partisan. Specifically, we do not allow:

- Referencing, promoting, or opposing candidates or nominees for public office, political parties, or elected or appointed government officials;
- Advocacy for or against past, current, or proposed referenda, ballot measures, and legislative, judicial, or regulatory outcomes or processes. This includes ads that promote or attack government policies or track records;
- Referencing, promoting, or selling merchandise that features prohibited individuals, entities, or content, including campaign slogans, symbols, or logos.

#### **Transparency Risk Controls**

Where accounts are designated as [Government, Politician, and Political Party Accounts \(GPPPA\)](#), those accounts are banned from placing ads on TikTok (with the exception of certain government entities in certain circumstances, as outlined above) and from monetisation features. We publish the details of our GPPPA policy on our [website](#), where we set out who we consider to be a GPPPA and the restrictions on those types of accounts.

We apply an internal label to accounts belonging to a [government, politician, or political party](#). Once an account has been labelled in this manner, a number of policies will be applied that help prevent misuse of certain features, e.g., access to advertising features and solicitation for campaign fundraising are not allowed.

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

### **Outcome 6: Signatories support the efforts of independent researchers to improve public understanding of Disinformation and Misinformation.**

We recognise the important role of researchers and subject matter experts in helping to identify misinformation trends and practices. We remain committed to increasing media literacy among our global community.

#### **Australian Associated Press Fact Check**

In 2025 we continued our fact-checking partnerships with the Australian Associated Press (AAP) to prevent



the spread of misleading information. In addition to retaining AAP to fact check information on our platform, AAP also produced and published up to 29 debunking articles a month on their independently-run website to combat misinformation and disinformation. Additionally, AAP proactively submitted leads they came across on TikTok to the organisation for review.

### **Community Partner Channel**

Our Global Community Partner Channel provides selected organisations an additional route for reporting content that they believe violates our Community Guidelines so that it can be reviewed by our teams. To date, more than 400 organisations who specialise in a range of safety issues use our Community Partner Channel. In Australia, 25 partners have been introduced to the program, including organisations focusing on combating antisemitism, Islamophobia, hate speech and racism.

The Community Partner Channel (CPC) Program empowers TikTok's trusted partners to flag potentially harmful content through a dedicated portal, enhancing TikTok's ability to keep the platform safe. Partnerships with these NGOs have helped us identify violative content, supported the evolution of our policies, and highlighted potential gaps or trends that have allowed T&S to make more informed decisions on content. The partners are experts in their fields and include representatives from animal welfare organisations, organisations combating antisemitism and Islamophobia, anti-bullying and anti-racism organisations and organisations representing the interests of groups such as LGBTQ+ Australians.

### **Transparency and Accountability Centres (TACs)**

TikTok remains committed to transparency and to ensuring the clarity of our practices for users, law enforcement agencies, governments and the general public. We continue active engagements with government and regulatory bodies, providing visits to our Transparency and Accountability Centres (TACs), to see up close how we moderate and recommend content, secure our platform, and protect people's privacy. Our TAC in Singapore complements our other TACs in Los Angeles, Washington DC and Dublin, and has hosted hundreds of guests both physically and virtually.

### **Consulting and Engagement**

TikTok undertakes regular consultation with independent researchers and experts from a range of institutions, including academics, NGOs and other experts, engaging in open dialogue and an exchange of information. The goal is to help key stakeholders understand how the platform works and to help TikTok inform policy around disinformation and misinformation, for example, as it relates to democratic processes (election integrity in Australia or referendums like the Voice to Parliament)

In addition, TikTok maintains a [Global Safety Advisory Council](#) comprised of regional experts, including 3 experts from Australia and New Zealand. The Council meets quarterly at a regional or global level to consult on emerging trends, threats and challenges in the digital space.

The Global Safety Advisory Council's membership is comprised of thought leaders from academia, law and government from across the region, who provide subject matter expertise and advise on TikTok's content moderation policies and practices to help shape regional and global guidelines. They support us in developing forward-looking policies that not only address the challenges we face today, but also identify existing and emerging issues which affect TikTok's platform and users, and develop strategies to tackle these challenges.

### **2025 Australian Federal Election**

Ahead of the 2025 Federal Election, TikTok convened its Safety Advisory Council for a dedicated session on election integrity, and hosted two expert roundtables bringing together academics, subject matter experts,



and key opinion leaders to map the election information landscape and identify emerging risks. These sessions surfaced a number of considerations that were fed back to cross-functional teams, including the challenges of moderating political humour, the need to address poor media literacy among users of all ages, and the risks posed by coordinated networks of accounts working to aggregate content in ways that shift its meaning at scale. TikTok also collaborated with AAP on a series of five videos for the Election Centre H5 page, designed to educate users and demonstrate TikTok's election integrity commitments to the broader public.

**Objective 7: Signatories publicise the measures they take to combat Disinformation and Misinformation. - (I&A)**

**Outcome 7: The public can access information about the measures Signatories have taken to combat Disinformation and Misinformation.**

In 2025 we continued our active engagements with government and regulatory bodies, providing visits to our TACs so that they could see up-close how we moderate and recommend content, secure our platform, and protect people's privacy. The TAC hosts guests such as government officials, law enforcement, delegates from organisations such as the OECD, academics, TikTok's Safety Advisory Council members, journalists, and creators.

In 2025, we continued to broaden the scope of information included in our [Community Guidelines Enforcement Reports](#) to bring greater transparency about our evolving moderation technologies and specialised safety initiatives by:

- Incorporating information on our work with **Large Language Models (LLMs)** to provide more precise and consistent moderation of human language in comments.
- Introducing new reporting metrics for **TikTok LIVE monetisation**, including granular data on warnings and demonetisation actions taken against creators.
- Enhancing **data visualisations** across the Transparency Centre to help our community more easily explore and interpret our safety and enforcement data.

We also refined our automated models to allow for a broader range of enforcement actions against comments, shifting from simple removal to options like restricting visibility, ensuring our moderation better supports self-expression while maintaining platform safety.

**Conclusion**

In 2025, TikTok continued to make significant and sustained investments in combatting misinformation and disinformation across our platform, building meaningfully on the progress made in 2024. Across each of our seven commitments under the Australian Code of Practice on Disinformation and Misinformation, we have strengthened our policies, improved our enforcement capabilities, and deepened our engagement with users, regulators, researchers, and civil society.

Locally, our enforcement performance in Australia remained consistently strong throughout 2025. Proactive removal rates for I&A violations exceeded 97% across all four quarters, and while there was some decrease in our removal-before-views metrics in Q1 2025, our enforcement performance continued to improve over the course of the year. Our response to the 2025 Australian Federal Election, including the deployment of user-facing election tools that reached Australians at significant scale, and our rapid activation of a crisis search guide after the Bondi Beach Terror Incident in December, demonstrated TikTok's capacity to act quickly and responsibly during both civic and emergency events.



We acknowledge that areas for continued improvement remain. We are in the process of launching granular advertising misinformation data for future reporting periods, and further refining our detection and enforcement systems as the information environment evolves, particularly as AI-generated content becomes an increasingly significant part of the content landscape.

As TikTok's community in Australia and globally continues to grow, we remain dedicated to fostering a space where creativity and authentic expression thrive alongside robust safeguards against harmful content. Earning and maintaining the trust of our users, our partners, and the broader public is not a one-time achievement; it is an ongoing commitment that drives everything we do. We look forward to continuing to work collaboratively with government, regulators, researchers, and civil society to meet that commitment in the year ahead.



## Appendix

### Policies

#### Approach to Disinformation and Misinformation

Our misinformation policies apply to content regardless of the poster's intent, as the content's harm is the same either way. Hence, they cover both "disinformation" (which is intentionally shared to mislead) and harmful misinformation that may not have been shared with the goal of deceiving people.

Like others in our industry, we do not prohibit people from sharing personal experiences, simply inaccurate myths, or misinformation that could cause reputational or commercial harm, in order to balance creative expression with preventing harm.

#### Policy on Misinformation

In a global community, it is natural for people to have different opinions, but we seek to operate on a shared set of facts and reality. **We do not allow inaccurate, misleading, or false content that may cause significant harm to individuals or society, regardless of intent.** Significant harm includes physical, psychological, or societal harm, and property damage. It does not extend to commercial and reputational harm, nor does it cover simply inaccurate information and myths. We rely on [independent fact-checking partners](#) and our database of previously fact-checked claims to help assess the accuracy of content.

Content is ineligible for the FYF if it contains general conspiracy theories or unverified information related to emergencies. To be cautious, content that warrants fact-checking is also temporarily ineligible for the FYF while it is undergoing review.

To help users manage their TikTok experience, we add warning labels to content related to unfolding or emergency events which have been assessed by our fact-checkers but cannot be verified as accurate, and we prompt people to [reconsider sharing](#) such content.

**Misinformation** includes inaccurate, misleading, or false content.

**Significant harm** includes severe forms of:

- Physical injury and illness, including death
- Psychological trauma
- Large-scale property damage
- Societal harm, including undermining fundamental social processes or institutions, such as democratic elections, and processes that maintain public health and public safety

**Conspiracy theories** are beliefs about unexplained events or involve rejecting generally accepted explanations for events and suggesting they were carried out by covert or powerful groups.

#### NOT allowed

- Misinformation that poses a risk to public safety or may induce panic about a crisis event or emergency, including using historical footage of a previous attack as if it were current, or incorrectly claiming a basic necessity (such as food or water) is no longer available in a particular location
- Medical misinformation, such as misleading statements about vaccines, inaccurate medical advice that discourages people from getting appropriate medical care for a life-threatening disease, and other misinformation that poses a risk to public health
- Climate change misinformation that undermines well-established scientific consensus, such as denying the existence of climate change or the factors that contribute to it



- Dangerous conspiracy theories that are violent or hateful, such as making a violent call to action, having links to previous violence, denying well-documented violent events, and causing prejudice towards a group with a protected attribute
- Specific conspiracy theories that name and attack individual people
- Material that has been edited, spliced, or combined (such as video and audio) in a way that may mislead a person about real-world events

#### **FYF ineligible**

- General conspiracy theories that are unfounded and claim that certain events or situations are carried out by covert or powerful groups, such as “the government” or a “secret society”
- Unverified information related to an emergency or unfolding event where the details are still emerging
- Potential high-harm misinformation while it is undergoing a fact-checking review

#### **Allowed**

- Statements of personal opinion (as long as they do not include harmful misinformation)
- Discussions about climate change, such as the benefits or disadvantages of particular policies or technologies, or personal views related to specific weather events (as long as they do not undermine scientific consensus)

#### **[Edited Media and AI-Generated Content \(AIGC\): What we say in our Community Guidelines](#)**

We welcome the creativity that new AI and other digital technologies may unlock. However, AI and other digital editing technologies can make it difficult to tell the difference between fact and fiction, which may mislead individuals or harm society. **We require users to label AIGC or edited media that shows realistic-appearing scenes or people.** This can be done using the [AIGC label](#), or by adding a clear caption, watermark, or sticker of their own.

Even when appropriately labeled, AIGC or edited media may still be harmful. **We do not allow content that shares or shows fake authoritative sources or crisis events, or falsely shows public figures in certain contexts.** This includes being bullied, making an endorsement, or being endorsed.

We are committed to protecting people's privacy. **We do not allow content that contains the likeness of young people, or the likeness of adult private figures used without their permission.**

#### **NOT allowed**

- Realistic-appearing people under the age of 18
- The likeness of adult private figures, if we become aware it was used without their permission
- Misleading AIGC or edited media that falsely shows:
- Content made to seem as if it comes from an authoritative source, such as a reputable news organisation
- A crisis event, such as a conflict or natural disaster

A public figure who is:

- Being degraded or harassed, or engaging in criminal or anti-social behaviour



- Taking a position on a political issue, commercial product, or a matter of public importance (such as an elections)
- being politically endorsed or condemned by an individual or group

**Allowed**

- The likeness of a deceased individual in certain educational settings, such as a World War I veteran in a museum exhibit
- The likeness of a public figure in certain artistic or humorous settings, such as a celebrity doing a popular TikTok dance or a spoof about a politician